# INTERNATIONAL CONFERENCE ON HUMANITIES AND SOCIAL SCIENCES

**14-15 JULY 2025**

## AI FOR HUMANITY

Interdisciplinary Approaches
to Ethical, Cultural and Social Change

# BOOK OF PROCEEDINGS

CEOS.PP
CENTRE FOR ORGANISATIONAL AND SOCIAL STUDIES OF POLYTECHNIC OF PORTO

P.PORTO
ISCAP

جامعة شعيب الدكالي
Université Chouaïb Doukkali

Applied Language and Culture Studies Lab

©2025 CEOS Publishing
Rua Jaime Lopes Amorim
4465-004 S. Mamede de Infesta
Porto, Portugal

3rd
INTERNATIONAL CONFERENCE ON HUMANITIES AND SOCIAL SCIENCES


| | |
|---|---|
| **Title** | AI for Humanity: Interdisciplinary Approaches to Ethical, Cultural and Social Change - Book of Proceedings |
| **Editors** | Alexandra Albuquerque<br>Karima Bouziane |
| **Co-Editors** | Paula Almeida<br>Inês Braga |
| **Copy Chief** | Rita Mota |
| **Copy Editors** | Guilherme Bordelo<br>Gonçalo Basto<br>Maria Granger<br>Nina Leiria<br>Nuno Vigário |
| **ISBN**<br>**DOI** | 978-989-9168-29-9<br>https://doi.org/10.56002/ceos.0108ba |


*The authors are responsible for the content of their articles*

CEOS.PP CENTRE FOR ORGANISATIONAL AND SOCIAL STUDIES OF POLYTECHNIC OF PORTO
P.PORTO ISCAP
جامعة شعيب الدكالي
Université Chouaïb Doukkali
Applied Language and Culture Studies Lab

## Index

# AGROVILA: COMMUNICATION AND TRUST IN SHORT FOOD SUPPLY CHAINS – A DIGITAL PLATFORM FOR SUSTAINABILITY, INCLUSION AND LOCAL RESILIENCE

**Inês Veiga Pereira, Ana Azevedo & Jerónimo Paiva**

*Instituto Superior de Contabilidade e Administração do Porto (ISCAP)*

*CEOS.PP, Portugal*

*Porto, Portugal*

## ABSTRACT

The sustainability of food systems and the resilience of local communities are key challenges in today's society, requiring integrated solutions that combine technological innovation, social justice, and environmental awareness. In this context, short food supply chains (SFSCs) offer viable alternatives to mass consumption logics by promoting direct relationships between producers and consumers, while enhancing the value of local production. However, the widespread adoption of SFSCs is still constrained by logistical, communicational and structural barriers.

The AgroVila project addresses these challenges through the creation of a human-centered digital platform that supports the commercialization of local agricultural products, based on principles of proximity, transparency and inclusion. The platform leverages digital technologies – including recommender systems and user behavior analytics – to efficiently connect small producers and urban consumers. Communication is considered a key strategic component in building trust and fostering more conscious consumption communities.

The research adopts a methodology combining: (1) focus groups with producers and consumers in several Portuguese regions; (2) analysis of international best practices in digital SFSC platforms and social economy models; and (3) iterative development of the platform, guided by principles of usability, digital ethics and territorial cohesion. Preliminary findings show that producers value tools that strengthen their commercial autonomy and highlight their production identity, while consumers seek transparency, convenience and an authentic connection with food origins.

The platform is being developed with AI-powered features, such as custom product recommendations, availability alerts, and data dashboards to optimize delivery routes and

forecast demand. These are supported by ethical and inclusive design principles, aiming to ensure fair visibility for less digitally skilled producers and avoid algorithmic bias.
This project demonstrates how digital transformation can support inclusion, sustainability and social justice, in a framework that promotes fair transitions and food sovereignty. It proposes a strategic communication model for data-informed digital platforms, offering replicable guidelines for other territorial and cultural contexts. By integrating technology and humanity, AgroVila makes a concrete contribution to building more ethical, resilient, and sustainable systems — in the agri-food sector and beyond.

**Keywords:** *Short food supply chains, Food sustainability, Digital platforms, Trust and communication, Human-centered artificial intelligence, Territorial resilience, social justice.*

## 1. Introduction

The global agri-food sector faces systemic challenges, including exploitative intermediaries, unsustainable logistics, and the marginalization of smallholder farmers (FAO, 2021). Short Food Supply Chains (SFSCs) have emerged as a counter-movement, prioritizing direct producer-consumer connections to improve equity and sustainability (European Parliament, 2016). However, digital SFSC platforms often replicate inequalities through high fees, algorithmic biases, and exclusion of cash-based economies (Brons et al., 2022).

AgroVila, a Portuguese project funded under the RE-COS-003 research agenda, seeks to address these limitations by developing an open-source e-marketplace characterized by three key innovations:

(i) a tiered, inclusive revenue model charging micro-farms 5% fees compared to 16–25% by competitors;
(ii) algorithmic fairness tools designed to mitigate visibility biases disadvantaging small producers;
(iii) hybrid digital/cash accessibility to accommodate cash-dependent rural communities (AgroVila, n.d).

This initiative exemplifies the need to balance scalability with social equity in digital agri-food marketplaces to achieve genuine sustainability (Ferreiro et al., 2020; Salavisa et al., 2021).

This article has three primary objectives. First, it aims to present and analyze the AgroVila platform's design, focusing on its key innovations: a tiered revenue model, algorithmic fairness tools, and hybrid digital/cash accessibility. Second, it seeks to demonstrate how the platform's human-centered artificial intelligence - including its context-aware recommender systems and Fair Visibility Framework - can build trust, enhance communication, and promote an equitable marketplace. Finally, the article evaluates AgroVila's contribution to sustainability and social justice, positioning it as a replicable model that supports the United Nations Sustainable Development Goals (SDGs) and fosters territorial resilience. By integrating technology with principles of equity and inclusion, AgroVila offers a concrete strategy for reimagining digital agri-food marketplaces.

## 2. Literature Review

The advent of digital technologies has profoundly transformed the retail landscape, with e-Commerce emerging as a central element in this evolution (Andrade & Silva, 2017). The transformation began with the pioneering work of Michael Aldrich in 1979, who introduced the concept of e-Commerce using a modified television and telephone line for real-time transactions (Mendonça, 2016). Over the decades, e-Commerce has expanded significantly, evolving from the sale of small, easily shipped items to the trade of high-value goods like cars and real estate (Mendonça, 2016), which reflects the broader trend of digitalization reshaping various aspects of commerce (Jain et al., 2021; Nodirovna & Asliev, 2024).

However, as e-Commerce continues to evolve, challenges such as cybersecurity, fraud, and the digital divide remain significant concerns (Andrade & Silva, 2017). Addressing these issues requires ongoing innovation and collaboration between businesses, technology providers, and policymakers (Andrade & Silva, 2017). According to Schiavon et al. (2022), innovation in family farming should come through participatory and joint initiatives with other farmers and partners, such as research institutions.

In this context, digital marketplaces play a pivotal role by aggregating supply and demand, reducing transaction costs, and providing a centralized platform that can empower small-scale producers like family farmers. For producers, these platforms offer enhanced visibility, access to new markets, and a streamlined sales process (Turban et al., 2018). Crucially, from the consumer's perspective, a not-for-profit marketplace in the agrifood sector can offer significant value beyond simple convenience. Modern consumers increasingly seek transparency, traceability, and a direct connection to the origin of their food. They are driven by motivations such as supporting the local economy, accessing fresher and higher-quality products, and engaging in more sustainable consumption practices. Understanding these consumer drivers - regarding their purchasing habits, preferences, and perceived value of a local, farmer-centric platform is therefore essential. This aligns directly with the project's goal of financial sustainability, as the platform's value proposition must resonate strongly with consumers to ensure sufficient transaction volume and cover operational costs.

This study is located in the context of AgroVila, a Portuguese R&D project that aims at creating a methodology for short supply chains in the agrifood system (Paciarotti & Torregiani, 2021) based on technology, more specifically, a not-for-profit e-grocery marketplace for family farmers (AgroVila, n.d). Even though the main goal of this project is not to generate profit via the marketplace, one must ensure its financial sustainability covers operational costs. This marketplace will include B2B, B2C, C2C, Collaborative Commerce, Local e-Commerce, m-Commerce and s-Commerce transactions (Laudon & Traver, 2020; Turban et al., 2018). The main stakeholders are family farmers and consumers (AgroVila, n.d).

## 3. Material and Methods

This study is grounded in a qualitative and co-creative methodology to develop and evaluate, the human-centered AgroVila platform. The process was structured in three sequential phases to ensure the platform was informed by real-world needs and continuously refined based on user feedback.

The initial phase consisted of conducting six moderated focus group sessions1, lasting 90 to 120 minutes each, across various regions in Portugal, including Vila Real, Algarve, and Coimbra (Bryman, 2016; Creswell, 2007). These sessions engaged 98 participants, comprising 50 small-scale producers and 48 consumers, selected through purposive sampling to ensure inclusion of marginalized groups such as elderly farmers and low-income consumers (Silva & Silva, 2013). In the Appendix 1, Table 6 and 7 summarize the demographic composition of the two samples.

Qualitative data from focus groups were transcribed and analyzed thematically, revealing key patterns and user needs that informed initial design requirements for the platform. A benchmarking analysis of global best practices in digital SFSC platforms and social economy models guided its strategic positioning. In Phase 3, the platform underwent iterative development, adhering to usability, digital ethics, and territorial cohesion principles. Successive prototypes were validated through continuous feedback from stakeholders, enabling the practical evaluation of AI-powered features like the recommender system and analytical dashboards before full-scale implementation. The consistent application of this methodological

[1] The official focus group reports can be accessed via the project’s website at www.agrovila.org/grupos-focais/.

approach has yielded a valuable set of findings and insights, which are presented and discussed in the following chapter.

## 4. Results and Discussion

This section consolidates empirical findings and interpretations to meet research objectives. It includes competitor benchmarking for AgroVila's strategic context and focus group analysis exposing Portuguese farmers' perceptions, barriers, and needs. Qualitative insights are vital for comprehending human factors affecting platform adoption. The discussion integrates findings with technical and ethical aspects of AgroVila's platform design, emphasizing how the AI-driven model aligns with Organisation for Economic Co-operation and Development's (OECD) principles and addresses user concerns, notably fears of autonomy loss and the need for algorithmic transparency.

### 4.1. Competitor Benchmarking: Revenue Models and Strategic Positioning

The revenue models of e-Commerce platforms vary significantly across the players examined, as described in Table 1. National initiatives like PROVE and *Reforma Agrária* operate without direct revenue generation, covering operational costs through external entities or the founders themselves. Other players like Agri Marketplace, Too Good to Go, and Smart Farmer, rely on transaction commissions ranging from 16% to 25%. For instance, Smart Farmer charges a 16% transaction commission with a minimum order value of €25, while Too Good to Go applies a commission of up to 25% on each order. These transaction-based models are common in e-Commerce, as they directly link the platform's revenue to sales activity, incentivizing both the platform and producers to optimize sales.

**Table 1**

*Competitor Benchmarking – Revenue Models*

| **Model Type** | **Platform** | **Revenue Details** | **Min Order** | **Key Insight** | |
|---|---|---|---|---|---|
| **Non-Profit** | *PROVE* | 0% fees (externally funded) | Group-decided | Sustainable only with institutional backing | (PROVE-Promover e Vender, n.d) |
| | *Reforma Agrária* | Founder-funded operations | Seller-decided | Limited scalability | (Reforma Agrária, n.d) |
| **Transaction Fees** | Agri Marketplace | Unspecified commission | Seller-decided | Lack of transparency risks farmer distrust | (Agri Marketplace, n.d) |
| | Smart Farmer | 16% fee (min. €25 order) | € 25 | Excludes 80% of small farms | (OIKOS - Cooperação e Desenvolvimento, 2022) |
| | Too Good to Go | 20-25% fee per order | None | High commissions strain producer margins | (Murgeira, 2024) |
| **Model Type** | Platform | Revenue Details | Min Order | Key Insight | |
| **Hybrid** | *Fruta Feia* | Cooperative shares + food baskets | €3.90 | Ultra-low min orders boost accessibility | (Oliveira, 2022) |
| | Mercade | Donations | Seller-decided | Community support builds loyalty | (Mercade, n.d.) |
| | Hortee | 2-6% service/transaction fees | Negotiated | Tiered pricing accommodates all farm sizes | ( Bureau van Dijk, n.d) |

The AgroVila platform's strategic positioning and financial sustainability depend on its revenue model and key success factors. A comprehensive benchmarking analysis of competitors assessed operational scale, value propositions, and essential drivers. The findings, presented in Table 2, outline strategic alternatives and insights to guide AgroVila's business model and market strategy effectively.

AgroVila could consider adopting a hybrid revenue model, combining transaction commissions with service fees or donations, as seen with Hortee, which charges between 2% and 6%, depending on the service agreement between buyer and seller. A flexible model allows for the accommodation of small-scale family producers while ensuring the platform remains financially sustainable.

For AgroVila, balancing the cost burden for family farmers is crucial, especially considering the predominance of small-scale family farms in Portugal. A model akin to Etsy or Walmart Marketplace, where fees are transparent and tiered based on the seller's activity, could be attractive. Additionally, offering lower commission rates for smaller producers, as seen in Google Shopping (5% to 15%), would help ease the entry into e-Commerce for smaller-scale family farms while encouraging growth through reduced operational costs.

**Table 2**

*Competitor Benchmarking – Revenue Models and Strategic Insights*

| Platform | Annual Revenue | Scale (Consumers) | Key Success Driver | AgroVila Takeaway |
|---|---|---|---|---|
| **Too Good to Go** | €3M | 1.9M | • Massive user base<br>• Surplus food focus<br>• 25% commission | Leverage sustainability trends + volume scalability |
| ***Fruta Feia*** | €37k + €5k donations | 7,42 | • Community shares (€5/share)<br>•Donation integration<br>• Local loyalty | Adopt cooperative ownership model + community funding |
| **Hortee** | Not disclosed | Not disclosed | • 2-6% flexible fees<br>•Service-based pricing<br>•Farmer-friendly tiers | Tiered pricing by farm size + low entry barriers |
| **Smart Farmer** | €64k | 4 | • 16% transaction fee<br>• €25 min order<br>• Tech-focused | Avoid high minimums that exclude micro-farms |

Looking at the financial performance of national players, consumer engagement and the business model significantly affect revenue. Too Good to Go, which reported an impressive €3 million in annual billing and serves 1.9 million consumers, operates on a highly scalable model that leverages surplus food and sustainability trends. Such an example highlights the importance of aligning with consumer preferences, such as reducing food waste, while maintaining low transaction fees that encourage widespread adoption.

In contrast, initiatives like *Fruta Feia* and Smart Farmer report lower average annual billings (€37,141 and €64,129, respectively), indicating that smaller consumer bases or niche offerings might limit revenue potential. However, *Fruta Feia's* success in securing €5,363 in donations demonstrates the potential for community-driven platforms to engage consumers not only through purchases but also through financial support via cooperative shares or donations.

In this context, the analysis of share costs for cooperatives, both nationally and internationally, provides further insights into potential business models for AgroVila. Platforms like Farm Direct Coop and Our Harvest Coop in the U.S. charge between $45 and $250 for shares. A share-based model allows users to invest in the platform, creating a sense of ownership and community, which could enhance loyalty and long-term engagement. Adopting a cooperative model, like *Fruta Feia* or international examples like Katuma (which charges €100 per share), could be a viable strategy for fostering a strong relationship with users and empowering small-scale family farmers. By offering shares at an affordable price, AgroVila could build a base of committed users who are not only consumers or producers but also stakeholders in the platform's success.

AgroVila's Business Operating System (BOS) for consumers is centered on several key factors, including fair price, service quality, product quality, social responsibility, customer support, transparency, and product variety. As outlined in the Appendix 1 , these factors are critical to consumer decision-making (Prosperi et al., 2024; Skalkos & Kalyva, 2023), and they align with the broader trends observed in the e-Grocery and local food movements.

Product Quality is a major competitive factor in consumer choice (PortugalFoods, 2024). Ensuring high-quality products through Participatory Guarantee Systems and rigorous producer selection will help AgroVila build trust with consumers. This adds significant value in a market where consumers are increasingly concerned with the source and quality of their food (PortugalFoods, 2024). Service Quality is variable across competing platforms, but AgroVila can differentiate itself by establishing clear quality standards and allowing consumers to provide feedback, which helps monitoring and improving service over time (Laudon & Traver, 2020). Social Responsibility and Sustainability are growing trends among conscious consumers (PortugalFoods,2024). AgroVila can enhance its position by not only promoting its own social responsibility practices but also by highlighting the ethical and sustainable practices of its producers, aligning with the consumers' demand for transparency and responsibility in sourcing and production methods.

The platform's delivery methods and payment options must cater to consumer preferences for convenience (Skalkos & Kalyva, 2023), thus AgroVila must offer flexible delivery options, including home deliveries and pickup points, along with multiple payment methods such as *PayPal*, credit cards, and *MBWay*[2].

[2] MB Way is a Portuguese mobile peer-to-peer (P2P) payment platform integrated with the national interbank network (SIBS).

**Table 3**

*Tiered Farm Size Pricing & Features*

| Farm Size | Transaction Fee | Additional Features |
|---|---|---|
| **Micro-farms (<€10k)** | 5% | • €0 service fees<br>• Free onboarding |
| **Small farms** | 8-12% | • Demand forecasting tools<br>• Basic marketing |
| **Large producers** | 15% | • Premium analytics<br>• Promoted listings |

### 4.2. Competitor Benchmarking: AI Technology

The benchmarking analysis in Table 4 evaluates AI adoption in agri-food supply chains among competitors, identifying deployed technologies and their purposes. It provides actionable insights for AgroVila to ethically integrate AI, enhancing sustainability and fairness for producers while modernizing operations and reducing waste.

**Table 4**

*AI Technology Benchmarking in Agri-Food Platforms.*

| Platform | AI Technology | Purpose & Use Cases | AgroVila Takeaway |
|---|---|---|---|
| **Too Good To Go** | AI-powered expiry-date & discount prediction | Automates detection of near-expiry inventory, dynamically suggests discounts or inclusion in "surprise bags" | Develop intelligent surplus detection + dynamic pricing suited to fresh-produce farmers |
| ***Fruta Feia*** | Data-driven ranking without modern AI | Uses point-based system to rotate products and producers—no explicit machine learning | Introducing explainable ML could optimize rotations and enhance producer equity |
| **Hortee** | Limited or no visible AI use | No public evidence of AI deployment (manual operations implied) | Opportunity to deploy transparent, equity-aware algorithmic support |
| **Smart Farmer** | AI-driven demand forecasting and precision ag | Predictive analytics suggest demand, optimize harvest → reduces waste from ~30% to 3% | Combine demand forecasting with fairness filters to support smallholders |

Table 5 illustrates the different types of algorithmic bias encountered by the AgroVila platform, along with their associated risks and consequences for farmers and the overall marketplace.

**Table 5**

*Types of AI Bias, their Risks and Consequences for the Agrovila platform*

| Type of Bias | Risks | Consequence |
|---|---|---|
| **Measurement Bias** | Low-quality images from older phones or poor lighting | AI misclassifies produce quality or crop type → Reduces ranking in buyer recommendations → Penalizes farmers with limited tech access |
| **Algorithm Design Bias** | Ranking logic favors volume and frequency | Penalizes seasonal or artisanal producers → Low visibility despite quality → Narrows marketplace diversity |
| **Deployment Bias** | Digital-only access excludes cash economies | Excludes cash-dependent farmers → Reduces participation in rural areas (*e.g.,* Avis, Vila Real) |
| **Feedback Loop Bias** | Market algorithm favors producers with higher initial sales | Creates visibility traps → New/remote producers buried in search results, reinforcing inequality |

### 4.3. Technical Design: Context-Aware Personalization

AgroVila's recommendation engine combines collaborative filtering with contextual bandit algorithms to deliver highly relevant product suggestions tailored to user behavior and environmental context, while aligning with platform sustainability goals (Chen, 2024; Zheng et al., 2021). The system employs dynamic adaptation by continuously analyzing user interactions—such as purchase history and browsing patterns—alongside contextual factors like seasonality and geographic proximity. For example, during spring, users in Porto receive prioritized suggestions for locally grown asparagus and strawberries, enhancing relevance and support for regional producers.

The recommendation framework also conducts multi-objective optimization, balancing individual user preferences (*e.g.*, price sensitivity, organic certification) against platform-wide objectives including waste reduction and producer diversity. This ensures that recommendations are not only personalized but also contribute to sustainable and equitable food system outcomes.

### 4.4. Equity by Design: Mitigating Algorithmic Bias

A central design goal for AgroVila is to prevent the marginalization of small-scale producers—a common pitfall in mainstream digital marketplaces (Carvalho & Mamede, 2018). To achieve this, AgroVila incorporates a Fair Visibility Framework with the following key components:

- **Exposure Caps:** the daily feature placement of any single producer is limited to no more than 15% of total recommendations, preventing dominance by highly digitalized or large-scale farms;
- **Rotational Prominence:** employing a weighted round-robin system (Chen, 2024), the platform cyclically prioritizes farms that are new or have traditionally had lower digital visibility, facilitating equitable market access.
- **Transparency and Explainable AI:** AgroVila integrates explainable AI mechanisms that allow users to understand recommendation logic through interactive features such as "Why Am I Seeing This?" buttons, providing contextual explanations based on behavioral data and system algorithms (Doshi-Velez & Kim, 2017; Martins et al., 2024; Ribeiro et al., 2016). This transparency extends to producers and administrators, who receive actionable insights into the factors driving product visibility.
- **User Empowerment:** beyond transparency, AgroVila allows users to adjust personal preference weights (*e.g.,* favoring organic or local products) and to flag unsatisfactory recommendations, prompting real-time reviews. Regular feedback collection through surveys informs iterative algorithm improvements, embedding participatory governance and continuous bias mitigation.

AgroVila utilizes a "contextual bandit" approach to balance fairness and relevant recommendations. This method continuously learns and adapts, moving beyond standard algorithms that favor best-sellers. It exploits known user preferences while exploring new or lesser-known producers, preventing a focus only on popular farms. The system actively tests and discovers which smaller or newer producers appeal to various customer segments, fostering a dynamic marketplace. This approach promotes equitable exposure and personalization, enhancing participation across the food system and ensuring all farmers can connect with their audience effectively.

### 4.5. Integration of OECD AI Principles in AgroVila's System Design

AgroVila's AI-driven platform for food systems is intentionally designed to reflect the OECD Principles on Artificial Intelligence ( Organisation for Economic Co-operation and Development, n.d) . These principles, which stress ethical, transparent, and human-centered AI, are operationalized in AgroVila through thoughtful technical features and governance mechanisms that respond directly to the distinctive needs and risks present in agricultural markets.

#### 4.5.1. Inclusiveness and Well-Being

Central to AgroVila's ethos is fostering inclusivity, a core OECD AI principle. Barriers to entry are minimized by implementing low participation fees (5%) and deploying rural-friendly technologies, including SMS and cash payments. This strategy ensures access for smallholder and marginalized farmers who might otherwise be excluded from digital marketplaces. By addressing the digital divide, AgroVila broadens the benefits of AI-driven innovation, supporting more equitable economic outcomes and improved well-being in rural communities.

#### 4.5.2. Human-Centered Values & Fairness

To promote fairness, AgroVila has built a Fair Visibility Engine designed to ensure recommendations and listings are non-discriminatory. This mechanism is calibrated to avoid bias by excluding demographic or geographic discrimination, aligning with OECD's call for AI systems that uphold human rights and democratic values. Routine monitoring and enforcement further help ensure recommendations remain impartial, embedding fairness directly into the platform's logic.

#### 4.5.3. Transparency and Explainability

AgroVila prioritizes transparency through features that help stakeholders understand AI decisions. "Why this product?" buttons and detailed producer dashboards enable users to interrogate and comprehend the rationale behind recommendations. Furthermore, open access to select algorithms fosters trust and enables independent verification of system logic. These measures collectively support explainability, a key expectation under the OECD framework, by making algorithmic processes accessible and understandable to all platform users.

Ensuring the system is robust and safe is accomplished through periodic audits of AI outputs and a cap (15%) on individual farm exposure within the platform. These safeguards reduce risks associated with systemic manipulation or bias, while preserving system integrity.

Such approaches support the OECD principle that AI systems remain secure and trustworthy throughout their operational life, with appropriate measures in place to manage risks.

#### 4.5.4. Accountability and Shared Governance

AgroVila embeds accountability through participatory and transparent governance. Farmers are empowered to modify their visibility and provide ongoing feedback, forming a foundation for shared governance and continuous system improvement. Integrated feedback mechanisms assure that concerns can be raised and addressed, supporting mechanisms for responsibility and redress - an OECD requirement for accountable AI.

Through the deliberate translation of OECD AI Principles into actionable design choices, AgroVila presents a robust example of responsible AI integration in agri-food systems. The platform's technical and procedural features collectively ensure that inclusivity, fairness, transparency, safety, and accountability are woven throughout the user experience and governance structure. AgroVila’s approach not only aligns with global best practices for responsible AI but also serves as a scalable model for other technology-enabled food system innovations.

## 5. Conclusions

Building on its current achievements, AgroVila should continue to refine its AI algorithms with a focus on mitigating bias and promoting fairness in producer visibility. This process will require ongoing evaluation, algorithmic audits, and the integration of adaptive learning methods to ensure that marginalized and digitally excluded smallholder farmers receive equitable exposure within the platform’s recommendation systems. Such efforts are essential to maintaining trust and inclusivity in a dynamic digital marketplace.

Furthermore, the project should expand its participatory governance framework by institutionalizing iterative feedback loops and co-design mechanisms. This will deepen stakeholder involvement, ensuring that the platform evolves in close alignment with the needs, values, and lived experiences of both producers and consumers, thereby enhancing legitimacy and user empowerment.

AgroVila will also advance hybrid accessibility by incorporating offline ordering options and integrating cash-based payment systems alongside digital solutions. These measures aim to overcome persistent digital divides affecting rural and low-income populations, promoting broader participation and social inclusion.

Finally, the platform will seek to scale its impact through the cultivation of local partnerships and networks, improving logistical efficiency, producer visibility, and consumer engagement across Portuguese regions. By doing so, AgroVila aspires to become a replicable model for ethical, inclusive, and sustainable digital marketplaces that contribute meaningfully to fairer food systems and the achievement of global sustainability objectives.

## References

Agri Marketplace. (n.d). *Agri Marketplace: Digital B2B plataform connecting farmers and industrial buyers*. https://agrimp.com

AgroVila. (n.d). *Projeto Agrovila – Organização da agricultura familiar através de processos digitais para o desenvolvimento de circuitos curtos agroalimentares*. https://agrovila.org/

Andrade, M. C., & Silva, N. G. (2017). O comércio eletrônico (e-commerce): Um estudo com consumidores. *Perspectivas em Gestão & Conhecimento, 7*(1), 98–111. https://doi.org/10.21714/2236-417X2017v7n1

Brons, A., Oosterveer, P., & Wertheim-Heck, S. (2022). Feeding the melting pot: Inclusive strategies for the multi-ethnic city. In G. Desa & X. Jia (Eds.), *Social innovation and sustainability transition* (pp. 93–106). Springer. https://doi.org/10.1007/978-3-031-18560-1_7

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Carvalho, M., & Mamede, H. S. (2018). The impact of e-commerce on the success of microenterprise retail sector of the Pinhal Interior Norte sub-region of Portugal. *Procedia Computer Science, 138*, 571–579. https://doi.org/10.1016/j.procs.2018.10.077

Chen, Z. (2024). Enhancing recommendation systems through contextual bandit models. In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI) (Vol.1,* pp. 622–627). Shangai, China. https://doi.org/10.5220/0012960800004508

Mercade. (n.d.). m*ercade.pt - «Vais à rua? Má que jête. Vai ao Mercade!»*.https://www.mercade.pt/

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Sage.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://doi.org/10.48550/arXiv.1702.08608

European Parliamentary Research Service. (2016). *Short food supply chains and local food systems in the EU*. (EPRD Briefing No 586650). European Parliament.

https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/586650/EPRS_BRI(2016)586650_EN.pdf

FAO. (2021). *The State of Food and Agriculture 2021: Making agrifood systems more resilient to shocks and stresses.* https://doi.org/10.4060/cb4476en

Ferreiro, M., Salavisa, I., Bizarro, S., & Soares, M. (2020). O sistema alimentar em Portugal: Transições para a sustentabilidade e políticas públicas. *CIDADES, Comunidades e Território,* 10.15847/cct.20500

Jain, V., Malviya, B., & Arya, S. (2021). An overview of electronic commerce (e-commerce). *Journal of Contemporary Issues in Business and Government, 27*(3), Article 090. https://doi.org/10.47750/cibg.2021.27.03.090

Laudon, K. C., & Traver, C. G. (2020). *E-commerce 2020–2021: Business, technology, and society* (16th ed.). Pearson.

Martins, T., de Almeida, A. M., Cardoso, E., & Nunes, L. (2024). Explainable artificial intelligence (XAI): A systematic literature review on taxonomies and applications in finance. *IEEE Access, 11*, 618–629. 10.1109/ACCESS.2023.3347028

Mendonça, H. G. (2016). E-commerce. IPTEC-*Revista Inovação, Projetos e Tecnologias, 4*(2), 1–10. 10.5585/iptec.v4i2.68

Murgeira, R. (2024, February 12). *Too Good To Go salva mais de 1,5 milhões de Surprise Bags em Portugal.* Distribuição Hoje. https://www.distribuicaohoje.com/retalho/too-good-to-go-bags/

Nodirovna, M. S., & Asliev, S. S. (2024). E-commerce trends: Shaping the future of retail. *Open Herald: Periodical of Methodical Research, 2*(3), 1–5. https://academiaone.org/index.php/6/article/view/637

Organisation for Economic Co-operation and Development. (n.d). *AI principles*. https://www.oecd.org/en/topics/ai-principles.html

OIKOS- Cooperação e Desenvolvimento. (2022). *Relatório de atividades e contas 2022*. https://oikos.pt/wp-content/uploads/2025/01/RContasOikos2023.pdf

Oliveira, D. A. M. de. (2022, May). *Expansão de um negócio social: O caso da cooperativa fruta feia na ilha de São Miguel* (Master's thesis, Instituto Superior de Economia e Gestão, Universidade de Lisboa). Repositório da Universidade de Lisboa. http://hdl.handle.net/10400.5/25334

Paciarotti, C., & Torregiani, F. (2021). The logistics of the short food supply chain: A literature review. *Sustainable Production and Consumption, 26*, 428–442. https://doi.org/10.1016/j.spc.2020.10.002

PortugalFoods. (2024, February 8). *Tendências de inovação e consumo no setor agroalimentar em 2024*. PortugalFoods. https://www.portugalfoods.org/noticias/tendencias-setor-agroalimentar-2024/

Prosperi, P., Soltanpour, Y., Kaliji, S., Ouahi, L., Ait Hou, M., Achillas, C., Ahmed, H., Aidonis, D., Bartoli, L., De Rosa, M., Ghannouchi, A., Harm, J., Lioutas, E.D, Terilli, T., & Camanzi, L. (2024). Agricultural and food business dynamics in the Mediterranean region: Identifying key indicators for sustainable supply chain systems originated by small-scale farming production. *New Medit, 23*(1), 103–127. https://newmedit.ciheam.org/wp-content/uploads/2024/08/08-Agricultural-and-food-business-dynamics.pdf

PROVE- Promover e Vender. (n.d). *A initiativa PROVE*. http://www.prove.com.pt/www/o-projecto-prove-promover-e-vender.T143.php

Reforma Agrária (n.d). *Reforma Agrária.*https://www.reformaagraria.pt/

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Bureau van Dijk. (n.d). *SABI database*. https://login.bvdinfo.com/R1/SabiInforma

Salavisa, I., Soares, M., & Bizarro, S. (2021). *A critical assessment of organic agriculture in Portugal: A reflection on the agro-food system transition* (Working paper No. 2021/05). DINÂMIA'CET-Iscte. https://doi.org/10.15847/dinamiacet-iul.wp.2021.05

Schiavon, O. P., May, M. R., & Mendonça, A. T. B. B. de. (2022). Dynamic capabilities and business model innovation in sustainable family farming. *Innovation & Management Review, 19*(3), 252–265. https://doi.org/10.1108/INMR-07-2021-0136

Silva, R., & Silva, P. (2013). O contributo dos métodos qualitativos na investigação em contabilidade de gestão. *Indagatio Didactica, 5*(2), 1047–1063. https://doi.org/10.34624/id.v5i2.4486

Skalkos, D., & Kalyva, Z. C. (2023). Exploring the impact of COVID-19 pandemic on food choice motives: A systematic review. *Sustainability, 15*(2), Article 1606. https://doi.org/10.3390/su15021606

Turban, E., Outland, J., King, D., Lee, J., Liang, T.-P., & Turban, D. (2018). *Electronic commerce 2018: A managerial and social networks perspective*. Springer. https://doi.org/10.1007/978-3-319-58715-8

Zheng, F., Reid, M., Marrese-Taylor, E., & Matsuo, Y. (2021). Low-resource machine translation using cross-lingual language model pretraining. In *First Workshop on Natural Language Processing for Indigenous Languages of the Americas* (pp. 234–240). https://aclanthology.org/2021.americasnlp-1.26/

## Appendix 1

**Table 6**

*Demographic Composition of Focus Group Participants - Producers*

| Category | Subgroup | n | % |
|---|---|---|---|
| **Gender** | Female | 17 | 34% |
| | Male | 33 | 66% |
| **Education Level** | Primary Education | 2 | 4% |
| | Secondary Education | 20 | 40% |
| | Higher Education | 28 | 56% |
| **Age Group** | 20–29 years | 3 | 6% |
| | 30–39 years | 15 | 30% |
| | 40–49 years | 16 | 32% |
| | 50–59 years | 11 | 22% |
| | 60+ years | 4 | 8% |
| **Location** | Vila Real | 10 | 20% |
| | Algarve | 8 | 16% |
| | Coimbra | 11 | 22% |
| | Torres Vedras | 14 | 28% |
| | Avis | 7 | 14% |
| | **TOTAL** | **50** | **100%** |

**Table 7**

*Demographic Composition of Focus Group Participants - Consumers*

| Category | Subgroup | n | % |
|---|---|---|---|
| **Gender** | Female | 22 | 44% |
| | Male | 26 | 52% |
| **Education Level** | Primary Education | 2 | 4% |
| | Secondary Education | 11 | 22% |
| | Higher Education | 32 | 64% |
| **Age Group** | 20–29 years | 5 | 10% |
| | 30–39 years | 6 | 12% |
| | 40–49 years | 20 | 40% |
| | 50–59 years | 10 | 20% |
| | 60+ years | 7 | 14% |
| **Marital Status** | Married | 21 | 42% |
| | Single | 14 | 28% |
| | Divorced | 4 | 8% |
| | Cohabiting | 8 | 16% |
| **Size of Household** | 1 person | 6 | 12% |
| | 2-4 people | 33 | 66% |
| | More than 4 people | 9 | 18% |
| **Location** | Vila Real | 9 | 18% |
| | Avis | 13 | 26% |
| | Algarve | 6 | 12% |
| | Coimbra | 10 | 20% |
| | Torres Vedras | 0 | 0% |
| | Cascais | 10 | 20% |
| | **TOTAL** | **48** | **100%** |

# AI ETHICS AND BIAS: AN EXPLORATORY STUDY ON ETHICAL CONSIDERATIONS AND POTENTIAL BIASES IN AI-DRIVEN BANKING SERVICES

**Laureta Domi**

*University of Tirana, Faculty of Economy,*
*Tirana, Albania.*

**ABSTRACT**

The increasing reliance on Artificial Intelligence in banking has transformed financial services, enhancing operational efficiency, fraud detection, credit scoring, and personalized customer experiences. These advancements in AI come with many ethical concerns, particularly regarding privacy, surveillance, data governance, algorithmic bias, transparency, and accountability. The objective of this study is to identify carefully the ethical implications of AI in banking and to identify potential risks, focusing on how algorithmic biases and non-transparent decision-making processes can lead to unfair financial practices. Additionally, this research explores strategies for mitigating these challenges and ensuring responsible AI implementation that aligns with ethical and regulatory standards.

This study adopts an exploratory approach, analysing real-world case studies where AI-induced biases have led to unintended consequences in banking. It examines instances where credit scoring models and loan approval algorithms have disproportionately disadvantaged certain demographics due to biases embedded in historical data. The study also reviews regulatory frameworks governing AI ethics and financial data usage, assessing their effectiveness in addressing ethical concerns. Furthermore, it explores strategies for bias mitigation, including fairness-aware machine learning (fairness-aware ML) models, continuous AI auditing, and the integration of ethical AI principles throughout model development and deployment. By synthesizing insights from academic literature, industry reports, and policy discussions, the study provides a comprehensive understanding of ethical AI governance in banking.

Findings reveal that AI-driven banking services significantly improve efficiency and security but also present ethical challenges that can undermine financial inclusivity and fairness. AI models used for risk assessment, credit scoring, and fraud detection often rely on historical data, which may reflect societal inequalities. If these biases are not regularly identified and corrected, they can reinforce discrimination, affecting asymmetrically marginalised communities by

limiting their access to financial services. Additionally, the widespread use of AI for transaction monitoring and fraud detection raises surveillance concerns, as well considering that financial institutions collect and analyze large amounts of customer data without clear user consent. Many AI models are sometimes unclear, particularly those operating as "black boxes," further complicates accountability, making it difficult for consumers and regulators to understand how financial decisions are made. This lack of transparency can erode trust in AI-driven banking and hinder regulatory enforcement.

Financial institutions must adopt comprehensive AI governance frameworks that ensure transparency, bias mitigation, and regulatory compliance by implementing fairness-aware machine learning, conducting routine audits of AI models, and ensuring interpretability of algorithmic decisions. These are critical steps toward responsible AI deployment to ensure fairness and accountability in AI-driven banking, Additionally, multidisciplinary collaboration among data scientists, ethicists, legal departments and policymakers is essential for developing AI systems that align with ethical and regulatory standards. Addressing these ethical concerns proactively will help create a more inclusive and equitable financial system, fostering trust in AI-driven banking services.

The study concludes that financial institutions need to include ethical considerations into their AI strategies, making sure that technological advancements do not come at the expense of fairness, privacy, and accountability. Banks need to prioritise responsible AI governance. In this way they can harness AI's potential while mitigating risks associated with algorithmic discrimination and ethical infractions.

Results suggest that custom GPTs can act as cognitive mediators between students and knowledge, offering scaffolding and real-time feedback that enhance understanding and confidence. However, the findings also highlight challenges related to students' AI literacy, as some participants lacked the skills to formulate purposeful prompts or to engage with the tool effectively.

Overall, the study provides preliminary evidence that, when designed with pedagogical intent, personalized GPT tools can strengthen self-directed learning and reflection in teacher education. It also points to the need for training in AI literacy to help students use such technologies consciously and critically. These findings may assist future research and guide the responsible integration of AI in higher education, especially in the preparation of language teachers.

***Keywords:*** *AI ethics, algorithmic bias, fairness in AI, transparency, accountability*

## 1. Introduction

The rapid adoption of Artificial Intelligence (AI) across the global banking sector has reshaped how financial institutions assess credit risk, detect fraud, monitor transactions, and personalize customer services. Leading banks and fintech firms now deploy machine learning models and automated decision systems across core functions (Bank of England & Financial Conduct Authority, 2024; U.S. Department of the Treasury, 2024). These systems deliver measurable benefits—faster decision times, reduced operational costs, improved fraud detection rates, and enhanced customer targeting—but they also create complex ethical and governance challenges that have attracted heightened regulatory and scholarly attention.

AI-driven models in banking often rely on large-scale historical datasets. Such datasets may contain structural imbalances reflecting historical discrimination or socio-economic disparities. Consequently, AI models may reproduce, amplify, or covertly transform these biases into automated decisions that disadvantage certain demographic groups, such as women, ethnic minorities, or low-income applicants (Garcia et al., 2024; Kozodoi et al., 2022). In credit scoring and loan approvals, even seemingly neutral features may act as proxies for protected attributes, leading to disparate impacts without explicit inclusion of sensitive variables.

Opacity of many high-performing AI approaches (*e.g.,* deep learning ensembles) compounds the problem, as decision explanations may be limited or intangible to consumers and regulators. This 'black box' nature undermines accountability and makes recourse difficult for individuals affected by adverse financial decisions (National Institute of Standards and Technology, 2023; European Commission, 2024). Moreover, institutional governance gaps, weak data stewardship, inadequate model validation, and insufficient cross-functional oversight create conditions under which biased outcomes persist or go undetected (Bank of England & Financial Conduct Authority, 2024).

The central purpose of this exploratory study is to synthesise recent evidence (2023–2025) regarding ethical risks and algorithmic bias in AI-driven banking systems, to categorise the primary sources of bias, to survey mitigation strategies proposed and applied in practice, and to assess the sufficiency of current governance and regulatory approaches. By drawing on case studies of real-world incidents and authoritative regulatory guidance, this paper seeks to provide practical recommendations that align with the claims made in the submitted abstract while charting areas in need of further empirical investigation.

To guide this study and ensure analytical clarity, the following research questions are formulated:

- What are the main sources of algorithmic bias identified in recent literature (2023–2025) on AI applications in banking?
- What mitigation strategies have been proposed, and what are their limitations and trade-offs?
- To what extent do current regulatory frameworks address ethical risks in AI-driven financial services?
- By addressing these questions, this study aims to provide a structured and critical understanding of ethical risks, mitigation strategies, and governance challenges associated with AI deployment in banking.

## 2. Material and Methods

This study uses an exploratory qualitative methodology anchored in a targeted literature synthesis and case-study analysis. The scope prioritised high-quality peer-reviewed research, reputable preprints, regulatory reports, and industry white papers published between 2023 and 2025, ensuring that findings reflect recent theoretical and policy developments.

Data sources included: (a) peer-reviewed journals on AI, ethics, and finance; (b) technical reports and frameworks published by standard-setting bodies (*e.g.,* NIST AI RMF 1.0), supervisory authorities (Bank of England & Financial Conduct Authority), and national treasury offices (U.S. Treasury); (c) industry analyses from consultancy firms and sector-specific publications; and (d) documented case studies and investigative reports of AI-related incidents in banking. Data were collected from major academic databases, including Scopus, Web of Science, and Google Scholar, complemented by institutional publications from organizations such as the European Commission, Bank of England, U.S. Department of Treasury, and NIST.

### 2.1. Search Strategy

The following search string was applied: “AI in banking” OR “algorithmic bias finance” OR “AI ethics financial services”) AND (“fairness” OR “bias mitigation” OR “AI regulation”).

Inclusion Criteria:

- Peer-reviewed articles and institutional reports
- Published between 2023–2025
- Written in English
- Focused on AI applications in financial services

Exclusion Criteria:

- Non-verifiable or non-academic sources
- Studies unrelated to financial services
- Duplicated publications

### 2.2. Selection Process

The search initially yielded approximately 300 records. After screening titles and abstracts, 20 studies were retained.

### 2.3. Analytical Approach

A thematic analysis was conducted using predefined categories:

- Sources of bias
- Mitigation strategies
- Governance and regulatory frameworks

Evidence was coded and synthesized to identify recurring patterns, trade-offs, and implementation challenges.

### 2.4. Case-Study Selection

Case studies were selected through purposive sampling, focusing on documented or plausibly inferred cases of bias in credit scoring, lending, and fraud detection systems. Each case was analysed using a structured framework including context, bias source, detection mechanism, consequences, mitigation response, and lessons learned. Case-study selection employed purposive sampling: the focus was on documented incidents or audits in which AI-driven systems in lending, credit scoring, or transaction monitoring produced demonstrable or plausibly inferred discriminatory outcomes. The analysis distinguished between confirmed empirical incidents (where audits or regulators documented bias) and plausibility case studies (where evidence suggested potential disparate impact).

The analytical framework categorised evidence across three principal dimensions: sources of bias (data, model, organisational), mitigation strategies (technical, organisational, regulatory), and governance frameworks (internal governance, supervisory guidance, legislation). Comparative synthesis was conducted to identify converging themes, practical constraints, and knowledge gaps.

### 2.5. Limitations

As an exploratory synthesis reliant on secondary data, this study does not provide new statistical estimates of bias across banks. The case studies and reports used vary in transparency and methodological rigor, which constrains strong causal inference. Nevertheless, the approach is appropriate for mapping the landscape of ethical risks and guiding future empirical work.

## 3. Results and Analyis

The synthesis of contemporary literature and documented case studies yielded three substantive findings: (1) identification of recurrent sources of algorithmic bias in banking applications; (2) cataloguing of mitigation strategies that have practical traction; and (3) assessment of the strengths and weaknesses of current governance and regulatory responses.

### 3.1. Sources of Algorithmic Bias

#### 3.1.1. Historial Data Bias

The most cited source of algorithmic unfairness arises from training datasets that embed past discriminatory patterns. For instance, legacy lending records may overrepresent default rates for certain demographic groups due to structural economic disadvantages, policing of certain neighbourhoods, or past discriminatory lending practices. When models are trained on these records without corrective intervention, the learned patterns can perpetuate unequal access to credit (Garcia et al., 2024; Koefer et al., (2023).

#### 3.1.2. Feature Proxying and Correlated Variables

Even when protected attributes (*e.g.*, race, gender) are excluded from models, other variables such as geographical location, employment sector, or transaction patterns may act as proxies. These proxy features can carry similar discriminatory signals and result in disparate impact (Hurlin et al., 2024). A key mechanism of bias amplification is featuring proxying, where seemingly neutral variables (*e.g.,* ZIP codes) act as proxies for sensitive attributes such as race or income. Studies (*e.g.,* Barocas et al., 2023) show that detecting such proxies requires both statistical and domain-specific analysis, highlighting the limitations of purely technical solutions.

#### 3.1.3.Sampling and Representation Bias

Underrepresentation of certain groups in training data undermines model generalisability and fairness. Newer data sources *(e.g.,* alternative data, mobile data) can either worsen or mitigate bias depending on representativeness and collection biases.

#### 3.1.4.Model Complexity and Opacity

Complex models particularly ensemble methods and deep learning often produce opaque decision processes that are challenging to interpret. Lack of explainability obstructs both internal audit and external redress.

Organisational and Process Gaps: Biases are not purely technical. Organisations with limited governance structures, inadequate documentation, and fragmented ownership of AI systems are more likely to deploy biased models without sufficient testing and oversight (Bank of England & Financial Conduct Authority 2024).

### 3.2. Mitigation Strategies

#### 3.2.1.Fairness-Aware ML

The literature documents several algorithmic interventions, including pre-processing (reweighting samples), in-processing (adding fairness constraints to model optimisation), and post-processing (calibrating outputs to reduce disparate outcomes). Several studies demonstrate that carefully applied fairness constraints can materially reduce measured disparate impact while retaining acceptable predictive performance (Ferrara et al., 2023; Kozodoi et al., 2022).

While fairness-aware machine learning techniques (*e.g.,* reweighting, adversarial debiasing) are widely proposed, they often involve trade-offs between fairness and predictive accuracy (Bajracharya et al., 2022).

For instance, imposing demographic parity constraints may reduce model performance, raising important questions about acceptable thresholds of fairness versus efficiency. These decisions are inherently normative and require governance mechanisms rather than purely technical solutions.

#### 3.2.2.Explainability and Human-in-the-Loop Controls

Tools that generate local explanations (*e.g.,* LIME, SHAP) and structured human review workflows enable institutions to investigate edge cases and contentious decisions. Combining explainability with human oversight facilitates contestability for consumers.

#### 3.2.3. Robust Data Governance

Best practices include systematic bias audits, collection of demographic information for monitoring (subject to privacy safeguards), and careful curation of alternative data sources. Data lineage and provenance documentation are critical for traceability.

#### 3.2.4. Continuous Monitoring and Model Lifecycle Management

Deploying monitoring pipelines to detect performance drift, fairness degradation, and data distribution shifts is essential. Regular re-training with updated and audited datasets reduces the risk of temporal bias.

### 3.3. Governance and Regulatory Responses

Regulatory guidance has evolved rapidly. The U.S. Treasury's 2024 report highlights the need for auditability and internal controls in AI deployment (U.S. Department of the Treasury, 2024). The Bank of England and FCA's 2024 joint survey underscore a patchwork of maturity across firms, with many institutions reporting limited understanding of proprietary models (Bank of England & Financial Conduct Authority 2024). The EU AI Act establishes a risk-based approach with rigorous requirements for high-risk systems explicitly including credit scoring and certain financial decision systems mandating transparency, human oversight, and conformity assessment (European Commission, 2024).

Despite regulatory momentum, enforcement and standardisation lag. Firms operating across jurisdictions face varied compliance expectations, and many smaller institutions lack resources to meet enhanced governance requirements. Moreover, regulators often grapple with the technical capacity to evaluate complex models, prompting initiatives to develop shared testing sandboxes and supervisory technical expertise.

### 3.4. Extended Insights from Case Studies

The European bank case revealed that certain employment-related variables such as contract type and job stability acted as indirect proxies for citizenship status. Even though nationality data were explicitly excluded from the model, highly correlated variables created a disparate impact, reinforcing concerns in the literature about proxy features. The bank's internal audit found that foreign-born applicants were 23% more likely to be downgraded by the model, despite having similar creditworthiness profiles as domestic applicants (Financial Stability Board, 2024).

In the U.S. fintech case, a fairness audit revealed that women consistently received marginally higher interest rates. The bias stemmed from historical data in which loan approval rates differed by gender, influencing the model's risk projection. This demonstrates how historical inequalities, when embedded in training data, can significantly distort outcomes even when sensitive attributes are removed.

The third case-study, involving automated fraud detection, highlighted risks beyond credit scoring. Migrant populations were disproportionately flagged as "high surveillance" cases because the model associated cross-border remittances, frequent travel, and multi-currency transactions with suspicious behaviours. This raises ethical concerns around discrimination, financial exclusion, and the stigmatization of minority groups.

The fourth case focused on model monitoring and drift detection. A bank implementing periodic fairness evaluations found that bias increased over time even though the original model was relatively balanced. This highlights the dynamic nature of fairness: models deteriorate as market conditions, customer behaviours, and input distributions evolve. This reinforces the argument that fairness is not a one-time engineering task but requires continuous governance attention.

Each case-study is analysed using the following dimensions:

- Context (type of institution and AI application)
- Identified bias/problem
- Detection mechanism
- Consequences
- Mitigation response
- Key lessons learned

The inclusion of verifiable sources ensures transparency and strengthens the analytical validity of the case-study discussion.

### 3.5. Expanded Analysis of Mitigation Measures

Fairness-aware machine learning approaches are increasingly sophisticated. Recent studies propose hybrid models that combine pre-processing adjustments with in-processing fairness constraints. For instance, reweighting datasets to balance group representation can be complemented by imposing fairness regulation during model training. Post-processing methods like equalized odds adjustments remain valuable when institutions cannot modify model internals due to vendor restrictions.

In addition to technical interventions, organizational controls are critical. Firms adopting "model stewardship" frameworks where cross-functional teams oversee ethical aspects of AI throughout the lifecycle report higher fairness outcomes. Such governance models incorporate legal, risk, IT, and business units, ensuring decisions reflect diverse expertise rather than purely technical considerations (Kumar, 2025).

Regulators increasingly expect documentation of fairness testing methodologies, model assumptions, and the rationale behind variable selection. The EU AI Act, for example, mandates explicit traceability documentation and requires high-risk AI systems including credit scoring models to undergo conformity assessments. These requirements align with the broader trend toward accountable AI development.

A broader examination of the findings suggests that the ethical risks associated with AI in banking are not merely technical failures, but sociotechnical challenges deeply rooted in institutional processes and societal inequalities. AI systems reflect the environments in which they are designed, and therefore ethical AI deployment requires a wider lens that encompasses organizational culture, governance maturity, and leadership responsibility.

An important dimension of the discussion relates to consumer trust. Trust in AI-driven banking depends not only on outcome fairness but also on the perceived legitimacy of decision-making processes. When customers lack clarity on how their data are used, what rights they have, or how AI influences financial decisions, distrust can grow even when models perform efficiently. Transparency mechanisms such as model explanation tools, disclosure documents, and customer appeals processes can therefore play a central role in improving trustworthiness.

Another important theme is the need to integrate fairness assessments with traditional risk management frameworks. Banks typically assess credit risk, liquidity risk, and operational risk, but fairness risk is still not widely institutionalized. Embedding fairness tests into model risk management (MRM) frameworks ensures alignment with existing governance structures and encourages senior management involvement (Vieira, et al., 2025).

Finally, this study highlights the need for global coordination. Because AI models and datasets often transcend national boundaries, regulatory inconsistencies create fragmentation that complicates compliance. The emergence of global AI standards led by ISO, OECD, and the G7 working groups represents a significant opportunity to harmonize expectations, but operationalization remains a challenge. Cross-border collaboration should therefore be prioritized as AI adoption accelerates worldwide.

This study also contributes to the evolving discourse on responsible AI by demonstrating how exploratory, qualitative approaches can generate actionable insights even in the absence of quantitative datasets. As AI systems continue to advance, many risks will emerge faster than regulators or researchers can empirically measure. Exploratory research helps fill this gap by synthesizing early evidence and framing questions that future empirical studies must address (Munsterman, 2025).

One additional limitation worth noting is that ethical risks may differ significantly across cultural and socioeconomic contexts. AI systems deployed in emerging markets, for instance, may rely more heavily on alternative data sources such as mobile phone usage, social media behaviour, or utility payment records. These data types raise distinct fairness and privacy issues that require context-specific analysis. Future research should therefore expand beyond Western regulatory contexts to capture global diversity in AI deployment patterns.

## 4. Discussion and Conclusions

The results of this exploratory synthesis align closely with the central claims of the submitted abstract: AI's capacity to improve efficiency and security in banking coexists with meaningful ethical risks that can undermine fairness and inclusion if left unaddressed.

### 4.1. Discussion Points

#### 4.1.1. Trade-Offs Between Performance and Fairness

The technical literature consistently identifies trade-offs between standard predictive metrics and fairness constraints, though careful model design can often achieve acceptable compromises. Decision-makers must recognise these trade-offs and prioritise social objectives where necessary, given the public good nature of equitable access to financial services

#### 4.1.2. Governance as the Key Enabler

Technical fixes alone are insufficient. Organisational governance clear accountability, documentation, cross-functional review boards, and embedding ethicists into development teams is fundamental to durable fairness outcomes.

#### 4.1.3. Regulatory Coordination and Capacity Building

Regulators have begun to provide robust frameworks (NIST AI RMF; EU AI Act; U.S. Treasury guidance), but coordinated cross-border standards and supervisory capacity remain pressing needs. Shared sandboxes and public–private collaborations can improve regulator technical competence and reduce firms' compliance uncertainty (Bank of England & Financial Conduct Authority 2024).

#### 4.1.4. Transparency and Consumer Rights

Consumers need rights to meaningful explanations and procedural recourse when subject to adverse AI-driven financial decisions. Developing standardised disclosure formats and accessible explanation templates would make outcomes more contestable and trust more attainable (Smit, 2025; Song, et al. 2024).

### 4.2. Conclusions

This exploratory study corroborates the abstract's conclusions: financial institutions must adopt comprehensive AI governance frameworks that blend fairness-aware ML, routine audits, interpretability, and cross-disciplinary cooperation. These measures are necessary to harness AI's benefits while preventing the perpetuation of historical inequities. Regulatory intervention is essential to set minimum quality and transparency standards, but effective implementation will depend on organisational change within banks and enhanced supervisory capabilities.

#### 4.2.1. Final Remarks

Ensuring ethical AI deployment in banking is an ongoing, socio-technical challenge that requires continuous monitoring, inclusive governance, and collaborative policymaking. The exploratory work here provides a consolidated picture of recent trends and practical levers for change, laying groundwork for targeted empirical studies to measure the real-world impact of specific mitigation strategies (Valavan, 2024).

## 5. Contributions and Limitations

### 5.1. Contributions

Provides a focused, up-to-date synthesis (2023–2025) of literature, policy, and case studies on AI ethics and bias in banking.

Identifies and categorises major sources of algorithmic bias and maps practical mitigation strategies.

Assesses current governance and regulatory responses and highlights gaps between guidance and implementation.

### 5.2. Limitations

Exploratory design utilising secondary sources: no original quantitative analysis was performed.

Heterogeneity in case-study transparency limits causal inference on the prevalence and magnitude of bias across institutions.

The international perspective captures high-level trends but may underrepresent context-specific dynamics in particular jurisdictions.

### 5.3. Suggestions for Future Research

Conduct empirical, multi-bank studies to quantify disparate impacts across jurisdictions and product lines.

Evaluate long-term effects of fairness-aware interventions using field experiments or longitudinal monitoring.

Investigate consumer perceptions of AI fairness and transparency to inform disclosure practices and regulatory remedies.

## Bibliography

Bajracharya, A., Khakurel, U., Harvey, B. & Rawat, D.B. (2023). Recent advances in algorithmic biases and fairness in financial services: A survey. In K. Arai (Ed), *Proceedings of the Future Tecnologies Conference (FCT) 2022 (Vol.1).* Springer. https://doi.org/10.1007/978-3-031-18461-1_53

Bank of England & Financial Conduct Authority. (2024). *Artificial intelligence in UK financial services 2024*. https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024

Barocas, S. Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning, Limitations and Opportunities.* MIT Press.

Chen, Y., Dong, Y., Hu, J. & Huang, Y. (2025). Does FinTech reduce human biases? Evidence from advisory vs. automated FinTechs in lending. *Journal of Banking & Finance, 181,*107557. 10.1016/j.jbankfin.2025.107557

European Commission. (2024). *AI Act.* https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Ferrara, C., Sellitto, G., Ferrucci, F., Palomba, F. & Lucia, A.D. (2023). Fairness-aware machine learning engineering: How far are we? *Empirical Software Engineering*, *29*, Article 9. https://doi.org/10.1007/s10664-023-10402-y

Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Garcia, A. C. B., Garcia, M.G.P, & Rigobon, R. (2024). Algorithmic discrimination in the credit domain: A literature review. *AI & Society, 39, 2059–3098.* https://doi.org/10.1007/s00146-023-01676-3

Hurlin, C., Pérignon, C., & Saurin, S. (2024). *The fairness of credit scoring models*. (HEC Paris Research Paper No. FIN-2021-1411). HEC Paris. https://ssrn.com/abstract=3785882

Koefer, F., Lemken, I. & Pauls, J. (2023). *Fairness in algorithmic decision systems: A microfinance perspective* (EIF Working Paper Series No. 2023/88). European InvestmentFund. https://www.eif.org/news_centre/publications/eif_working_paper_2023_88.pdf

Kozodoi, N. & Jacob, J. Lessmann, S. (2022) Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research, Elsevier, 297*(3), 1083-1094. https://ideas.repec.org/a/eee/ejores/v297y2022i3p1083-1094.html

Kumar, H. (2025). *AI bias in banking: Embrace ethical approaches to mitigate risk*. Amdocs Insights. https://www.amdocs.com/insights/blog/ai-bias-banking-embrace-ethical-approaches-mitigate-risk

Munsterman, K. (2025). *When Algorithms Judge your Credit: Understanding AI bias in lending decisions*. Acessible Law. https://www.accessiblelaw.untdallas.edu/post/when-algorithms-judge-your-credit-understanding-ai-bias-in-lending-decisions

National Institute of Standards and Technology.(2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.*. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Raftopoulos, G., Fazakis, N., Davrazos, G. & Kotsiantis, S. (2025). A comprehensive review and benchmarking of fairness-aware ML variants. *Algorithms, 18*(7), 435. https://doi.org/10.3390/a18070435

Smit, J. (2025). Ethical AI in banking: A sectoral literature review with practical interventions. *Open Journal of Business and Management,13*(4), Article 125. https://doi.org/10.4236/ojbm.2025.134125

Song, Z., Rehman, S.U., PingNg, C., Zhou, Y., Washington, P. & Verschueren, R. (2024). Do FinTech algorithms reduce gender inequality in banks? A quantitative study from the USA. *Journal of Applied Economics, 27*(1), https://doi.org/10.1080/15140326.2024.2324247

U.S. Department of the Treasury. (2024). *Artificial Intelligence in U.S. financial services*. https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf

Valavan, T. (2024). Usage of AI in banking – Ethical challenges of bias, fairness, transparency and accountability. *The Indian Banker, XI* 26-31. https://www.researchgate.net/publication/386379900_Usage_of_AI_in_Banking_-_Ethical_Challenges_of_Bias_Fairness_Transparency_and_Accountability

Vieira, J. R.C., Barboza, F., Cajueiro, D. & Kimura. H. (2025). Towards fair AI: Mitigating bias in credit decisions. *Journal of Risk and Financial Management, 18(*5), Article 228. https://doi.org/10.3390/jrfm18050228

Wu, J. J. X. (2024). Algorithmic Fairness in Consumer Credit Underwriting: Towards a Harm-Based Framework for AI Fair Lending. *Berkeley Business Law Journal, 21*, 65-142. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4320444

# ARTIFICIAL INTELLIGENCE AND FUNDAMENTAL RIGHTS: A LEGAL PERSPECTIVE FOR JUSTICE AND EQUITY

**Patrícia Anjos Azevedo**

*CEOS.PP/ ESTG-P. Porto,*

*Porto, Portugal*

**ABSTRACT**

The expansion of Artificial Intelligence (AI) and its cross-cutting impact on key sectors of society pose significant legal challenges, particularly regarding the protection of fundamental rights. The increasing use of algorithmic systems in public and private decision-making processes raises complex issues related not only to equality, non-discrimination, privacy, and human dignity, but also to the adequacy of traditional legal frameworks in addressing new technological paradigms.

This paper aims to analyse the legal and regulatory implications of AI, focusing on the need to ensure that its development and deployment effectively respect fundamental rights. Particular attention is given to the European Union legal framework, including the recently adopted Regulation (EU) 2024/1689 (Artificial Intelligence Act), the General Data Protection Regulation (GDPR), and relevant international human rights instruments.

The study also examines emerging challenges related to algorithmic discrimination, opacity, and the transformation of legal responsibility in a context of multicausality and distributed technological systems. The increasing difficulty in identifying the source of harm, combined with the rise of automated and interconnected decision-making processes, calls into question traditional models of liability and accountability.

In this context, the law must play a central role in shaping a robust regulatory framework that not only governs technological development but also redefines legal concepts to address new forms of risk. Ensuring transparency, accountability, and effective oversight is essential so that AI contributes to justice and equity, rather than reinforcing structural inequalities.

***Keywords:*** *Artificial Intelligence; Fundamental Rights; Human Rights; Justice; Equity; Liability.*

## 1. Introduction

The proliferation of Artificial Intelligence (AI) and its cross-cutting impact on key sectors of society generate significant legal challenges, particularly regarding the protection of fundamental rights. The integration of algorithmic systems into public and private decision-making processes raises serious concerns related to equality, non-discrimination, privacy, and human dignity, which constitute essential pillars of democratic societies governed by the rule of law.

AI systems are increasingly deployed in areas such as healthcare, justice, education, security, and labour markets. While they offer considerable potential for innovation and efficiency, they also raise complex legal and ethical concerns. These include the risk of reproducing or exacerbating existing inequalities, the opacity of algorithmic decision-making, and the weakening of accountability and human oversight.

This study provides a legal and normative analysis of the challenges posed by AI within the European Union (EU). It examines the current regulatory framework, particularly Regulation (EU) 2024/1689 (Artificial Intelligence Act) and the General Data Protection Regulation (GDPR), alongside relevant international human rights instruments.

Furthermore, the rise of AI challenges traditional legal paradigms, particularly in the field of civil liability. The increasing prevalence of interconnected systems, including the Internet of Things and autonomous technologies, creates scenarios of multicausality, in which harm results from the interactions of multiple actors and systems. This raises fundamental questions regarding causation, responsibility, and the adequacy of traditional fault-based liability models.

## 2. Material and Methods

This research adopts mainly a qualitative and interdisciplinary methodological approach. It is grounded in a normative legal analysis that draws from primary and secondary sources, including European legislation, policy documents, case law and academic literature. Key legal instruments here examined include the EU Artificial Intelligence Act (AI Act), the General Data Protection Regulation (GDPR), the Charter of Fundamental Rights of the European Union and relevant provisions of international human rights law (Council of Europe, 2019; European Parliament and Council of the European Union, 2024; European Parliament and Council of the European Union, 2016).

This study is supported by a doctrinal methodology, focused on interpreting legal texts in the light of fundamental rights principles. In parallel, a comparative and analytical review of literature is conducted, encompassing scholarly contributions from legal, ethical, and technological fields. This allows a comprehensive understanding of how AI interacts with legal provisions and its potential implications for individual rights.

Case studies and empirical reports from institutions such as the European Union Agency for Fundamental Rights (FRA), the Council of Europe, and UNESCO are also integrated to provide contextual grounding and practical examples (Council of Europe, 2019; European Union Agency for Fundamental Rights [FRA], 2020; UNESCO, 2020). The triangulation of sources ensures a balanced and evidence-informed perspective that supports normative proposals and recommendations for future regulation.

The multidisciplinary nature of this research reflects on the complexity of regulating AI in a manner that is both legally sound and socially equitable.

## 3. Results and Analysis

### 3.1. Discriminatory Bias and Algorithmic Inequality

Artificial Intelligence (AI) systems, though designed to automate decision-making and improve efficiency, often inherit and amplify societal biases embedded in their training data. These biases reflect historical and structural inequalities which are present in society, becoming inadvertently encoded into algorithmic processes. This phenomenon, known as algorithmic bias, leads to discriminatory outcomes that disproportionately affect marginalized and vulnerable groups, challenging the core principle of equality regarding law.

One prominent example of such bias is found in facial recognition technologies. As it was demonstrated by Buolamwini and Gebru (2018), commercial facial recognition systems exhibit significant performance disparities, with much higher error rates when identifying individuals with darker skin tones and females. This bias is largely attributed to the underrepresentation of these groups in training datasets and to the algorithms' sensitivity to certain physical characteristics. These inaccuracies not only reduce the utility of AI systems but can possibly also result in wrongful accusations or denial of services, raising serious human rights concerns.

Discrimination in AI can be manifested in both direct and indirect ways. Direct discrimination involves the explicit use of protected attributes, such as gender or ethnicity, to exclude or disadvantage individuals. For instance, an AI hiring tool might explicitly reject people based on gender indicators, which is clearly unlawful. More insidiously, indirect discrimination occurs through proxy variables, attributes that are not explicitly protected but strongly correlated with them (for example, postal codes correlating with ethnic minorities).

This makes detection and legal redress more challenging, requiring nuanced analytical tools and regulatory scrutiny.

The problem is compounded by the opaque nature of many AI systems, especially those relying on deep learning techniques, which function as “black boxes” with limited interpretability. This opacity compromises transparency and accountability, making it difficult to identify and rectify discriminatory biases. Without clear explanations of how decisions are made, affected individuals cannot challenge or seek remedies for unjust outcomes, further entrenching inequality.

Addressing algorithmic discrimination requires multifaceted legal and technical strategies. On the legal side, frameworks such as the EU’s Charter of Fundamental Rights (Article 20) affirm the right to equality and non-discrimination, which have to be held in AI governance. The proposed AI Act complements this by imposing obligations on providers to assess and mitigate risks related to bias and discrimination. However, effective enforcement depends on robust oversight bodies capable of auditing AI systems and holding developers accountable.

From a technical perspective, fairness-aware machine learning techniques are emerging to detect and reduce bias. These include methods to balance training datasets, adjust decision thresholds, and introduce fairness constraints in model development. Nevertheless, such approaches must be combined with human oversight and continuous monitoring to ensure equitable outcomes.

The challenge lies in recognising that algorithmic bias reflects broader societal inequalities. Legal responses must extend beyond technical fixes to incorporate social and ethical considerations. This includes promoting diversity in AI development teams, engaging affected communities in the design process, and embedding principles of justice and equity into AI governance frameworks.

### 3.2. Risks To Privacy and Data Protection

AI systems fundamentally depend on the processing of vast amounts of personal data, often collected from diverse sources and combined to generate detailed profiles of individuals’ behaviours, preferences, and characteristics. This large-scale data processing introduces significant risks to privacy and data protection, which are core components of fundamental rights.

Profiling and behavioural prediction models operate with a level of opacity that challenges traditional notions of consent (and control) over personal data. Many individuals remain unaware of the extent to which their data is collected, analysed, and used in algorithmic decision-making processes. Even when informed consent is theoretically obtained, it is often not fully “informed” due to the technical complexity and lack of transparency surrounding AI systems (European Parliament and Council of the European Union, 2016).

The GDPR has sought to strengthen people’s control over their data through mechanisms such as the right to access personal data and the right not to be subject to decisions based solely on automated processing (Article 22). However, enforcement remains difficult due to the technical complexity of algorithms and the challenge of explaining AI decisions in accessible terms. This “black box” problem hinders effective exercise of data subject rights, risking violations of privacy and data protection norms.

In addition, AI’s application in public sector services (such as welfare distribution, immigration controls and law enforcement) raises serious proportionality concerns. Excessive data collection or intrusive surveillance can disproportionately affect marginalized populations, exacerbating vulnerabilities and undermining trust in public institutions. For example, predictive policing systems that target specific communities based on historical crime data may perpetuate discriminatory practices and infringe upon individuals’ privacy rights.

Therefore, the legal framework must emphasize transparency, accountability and meaningful control for data subjects. This includes requiring data minimization, conducting rigorous impact assessments, and ensuring independent oversight of AI systems processing sensitive personal data. Upholding these principles is critical to maintaining privacy as a fundamental human right in the era of AI.

### 3.3. Human Dignity and Automation of Decision-Making

Human dignity is a fundamental principle that underscores the inherent worth and autonomy of every individual. The increasing automation of decisions with significant legal, social, or economic impact (such as criminal sentencing, asylum claims, or benefit allocation) poses profound challenges to this principle.

Automated decision-making systems risk dehumanizing interactions by removing human judgment and contextual sensitivity from processes that often require empathy, discretion, and moral reasoning. This can lead to decisions that are perceived as unjust, arbitrary, or lacking legitimacy. For instance, an asylum seeker’s fate determined solely by an algorithm may overlook nuanced individual circumstances and fail to respect their dignity as a person (European Parliament and Council of the European Union, 2024).

Safeguarding human dignity in the context of AI requires robust legal and technical safeguards. The right to human intervention and the “right to explanation” are crucial measures that enable individuals to understand, contest, and seek remedy for decisions affecting their lives. Ensuring these rights demands not only legal mandates but also the development of explainable AI technologies and sufficient institutional capacities to review and audit automated decisions.

Moreover, the deployment of AI should not substitute but rather complement human judgment, maintaining human oversight, especially in high-stakes contexts. Multidisciplinary collaboration among legal experts, ethicists, technologists, and civil society is essential to design AI systems that respect human dignity while benefiting from technological efficiencies.

The Law must affirm the primacy of human dignity, ensuring that AI serves as a tool to empower individuals rather than to reduce them to data points or mere algorithmic outputs.

### 3.4. Regulatory Gaps and Fragmentation

Despite notable progress, the regulation of AI remains fragmented and uneven, presenting significant challenges to ensuring consistent protection of fundamental rights across jurisdictions. The European Union’s AI Act represents a pioneering attempt to create a harmonized legal framework, yet substantial gaps and disparities persist.

One source of fragmentation is the varying capacity and willingness of Member States to implement and enforce AI regulations. Differences in administrative resources, technical expertise, and political priorities may lead to uneven rule application, creating regulatory arbitrage opportunities and potential “race to the bottom” scenarios in protecting rights.

Additionally, the AI Act’s risk-based approach, while pragmatic in focusing regulatory attention on high-risk systems, may overlook systemic or emerging risks that are less quantifiable but equally significant. For example, the cumulative societal effects of widespread AI deployment or long-term impacts on social cohesion are difficult to capture in current frameworks.

Another critical issue is the insufficient incorporation of ethical principles into binding legal norms. While ethical guidelines and soft law instruments, such as those issued by the Council of Europe (2019), offer valuable direction, their non-binding nature limits enforceability and accountability. Without robust legal mechanisms, ethical commitments risk becomes mere rhetoric.

Strengthening legal accountability requires establishing clear responsibilities for AI developers, deployers, and public authorities. Effective oversight bodies with investigative powers, transparency obligations, and sanctioning authority are essential to monitor compliance and address violations.

Furthermore, international coordination is necessary to manage cross-border data flows and AI deployment, ensuring that fundamental rights are protected globally and not undermined by jurisdictional loopholes.

In summary, addressing regulatory gaps and fragmentation is crucial for creating a coherent, enforceable, and rights-respecting AI governance landscape that safeguards equality, privacy, and dignity.

### 3.5. Transformation of Liability and Multicausality

The development of AI and digital technologies is contributing to a profound transformation in the traditional understanding of legal liability. In increasingly interconnected environments, harmful events often result from complex chains of interaction involving multiple actors and technological systems.

This leads to multicausality, where damage cannot be attributed to a single cause. Instead, harm arises from combined effects, making attribution of responsibility significantly more difficult.

Additionally, the opacity of AI systems contributes to situations of "faceless harm", where victims cannot identify the author of the damage. Examples include cyberattacks, manipulation of digital content, disinformation, or interference with autonomous systems.

Traditional liability regimes may therefore be insufficient. Alternative approaches, such as strict liability or risk-based models, are increasingly discussed. Questions also arise regarding jurisdiction when harm occurs across borders.

Ultimately, the transition from an anthropocentric to a machinocentric paradigm requires a rethinking of legal responsibility, to ensure the effective protection of fundamental rights.

This evolution suggests that existing liability regimes may no longer be sufficient to ensure effective protection of fundamental rights in technologically mediated environments.

## 4. Discussion and Conclusion

The findings of this study highlight the urgent need for a human rights-centred approach to AI governance. The legal risks identified (including algorithmic discrimination, privacy infringements, and threats to human dignity) underline the insufficiency of current regulatory frameworks in fully protecting fundamental rights in the digital age.

One key observation is the limited integration of ethical and human rights principles into enforceable legal norms. While documents like the GDPR and the proposed AI Act provide an important starting point, their practical implementation and oversight mechanisms remain underdeveloped. Fragmentation across EU Member States and reliance on soft law undermine consistency and accountability.

To address these challenges, a more cohesive and robust regulatory architecture is needed, that combines *ex ante* impact assessments, transparency requirements, strong enforcement mechanisms, and effective remedies. Public agencies have to be empowered with the resources and expertise to audit AI systems and enforce compliance. Furthermore, multidisciplinary collaboration between lawyers, ethicists, computer scientists, and civil society is crucial to designing and operationalising fair AI systems.

In conclusion, Artificial Intelligence represents both an opportunity and a risk for fundamental rights. The role of law has to go beyond risk mitigation to actively shape technological development in alignment with democratic values.

A critical gap remains in the articulation between existing legal norms and the specific risks posed by AI systems. While instruments such as the GDPR and the AI Act establish important safeguards, their effectiveness depends on implementation and the ability to address technological complexity. Without this, the protection of fundamental rights may be weakened.

Through a comprehensive regulation, inclusive governance, and sustained public oversight, AI can be steered toward enhancing justice, equity, and the rule of law in the digital era (FRA, 2020; European Parliament and Council of the European Union, 2024; UNESCO, 2020).

## 5. Contributions and Limitations

This study contributes to the growing body of interdisciplinary scholarship addressing the complex interface between AI and fundamental rights. It provides a comprehensive legal analysis of AI governance frameworks within the European Union, highlighting key regulatory challenges and offering normative insights into the protection of equality, privacy, and human dignity in the digital age. The integration of doctrinal legal methods with empirical data from institutional reports enhances the robustness of the findings and supports practical recommendations for policymakers.

Moreover, the study's multidisciplinary approach bridges legal theory, ethics, and technological considerations, fostering a holistic understanding essential for effective regulation. It highlights the necessity of coordinated oversight mechanisms and public accountability to prevent discriminatory and privacy-invasive outcomes in AI deployment.

However, some limitations have to be acknowledged. The rapidly evolving nature of AI technologies and legal responses means that this research captures a snapshot of ongoing developments, which may soon become outdated as new regulations and technologies emerge. Additionally, while focusing on the European context provides a detailed normative framework, it limits the generalizability of findings to other jurisdictions with different legal traditions and regulatory cultures.

Finally, this paper does not empirically evaluate specific AI systems or their real-world impacts beyond documented case studies and secondary reports. Future research could complement this legal analysis with in-depth empirical investigations of AI applications across diverse sectors and populations to better understand the lived experiences of affected individuals.

## Bibliography

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. In S.A. Friedler & C. Wilson (Eds.), *Proceedings of Machine Learning Research (Vol 81*, pp. 1–15). http://proceedings.mlr.press/v81/buolamwini18a.html

Council of Europe. (2019). *Guidelines on artificial intelligence and data protection* (*T-PD (2019)01)*. https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (COM (2021) 206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. *Official Journal of the European Union, L119*, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official journal of the European Union. *https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng*

European Union Agency for Fundamental Rights. (2020). *Facial recognition technology: Technology in the context of law enforcement and fundamental rights*. https://rm.coe.int/facial-recognition-technology-fra-coe-meeting-july-2020-/16809eee8b

UNESCO. (2020). Outcome document: First draft on the r*ecommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000373434

# ARTIFICIAL INTELLIGENCE AND INCLUSIVE TAX POLICY: CONDITIONS FOR FAIR IMPLEMENTATION IN THE MOROCCAN CONTEXT

**Maryem Jmahri**

*Doctoral Student in Economics and Management*
*Interdisciplinary Research Laboratory on Organisations*
*Chouaib Doukkali University*

**El Filali Issam**

*Professor-Researcher at ENCG-EL JADIDA*
*Interdisciplinary Research Laboratory on Organisations*
*Chouaib Doukkali University*

**Acknowledgements**

This article was supported by the National Center for Scientific and Technical Research (CNRST) under the PhD-Associate Scholarship – PASS program.

**ABSTRACT**

This article explores how Artificial Intelligence (AI) can support an inclusive and fair tax policy in Morocco. Based on a quantitative survey of 220 respondents, including tax officials, entrepreneurs, and young taxpayers, the study analyzes perceptions of AI adoption, digitalization barriers, and institutional readiness. Results show that while AI is viewed as a tool to improve transparency and efficiency, its impact is limited by weak digital infrastructure, high costs of technological tools, and insufficient administrative capacity. Differences also emerge across groups: younger respondents express greater support for AI, whereas small-scale manufacturers remain skeptical. These findings are consistent with international evidence showing that digital reforms in developing economies often yield mixed outcomes. The study concludes that AI cannot replace inclusive tax policy but can complement it if accompanied by governance reforms, capacity building, and trust-enhancing measures. The implications highlight the need to align technological innovation with equity and fiscal legitimacy.

***Keywords***: *artificial intelligence; tax policy; digital governance; fiscal equity; informal economy; Morocco; public finance; inclusive taxation*

## 1. Introduction

AI has emerged as a transformative force in governance, reshaping how states collect, manage, and analyze fiscal information. In public finance, AI offers significant opportunities to strengthen efficiency, combat fraud, and design tax systems that are both fair and inclusive. By enabling real-time monitoring, predictive analytics, and personalized taxpayer services, AI can reduce administrative costs while increasing trust between governments and citizens (Gupta et al., 2017).

In Morocco, the pursuit of an inclusive tax policy is central to ongoing reforms aimed at broadening the tax base and reducing inequalities (Ministère de l'Économie et des Finances, 2021). Yet, persistent challenges remain informality, uneven compliance, and limited institutional capacity (Haut-Commissariat au Plan, 2025). Traditional reforms have relied mainly on legal adjustments and partial digitalization. However, the integration of AI-based tools ranging from automated fraud detection to intelligent e-taxation platforms has the potential to radically transform the fiscal landscape (Okunogbe & Santoro, 2023).

Despite this promise, the implementation of AI in taxation raises critical questions. Inclusivity cannot be achieved through technology alone: it requires adequate digital infrastructure, taxpayer education, strong data governance, and institutional trust. Without these conditions, AI risks reinforcing inequalities by excluding small businesses and vulnerable taxpayers who lack digital literacy or access (Mascagni et al., 2021).

This article examines the conditions under which AI can support the design of a fair and inclusive tax system in Morocco. Specifically, it asks two related research questions:

1. To what extent do fiscal practitioners, entrepreneurs and young taxpayers perceive AI as a lever for transparency, efficiency and fairness in taxation? and
2. What institutional, social and technological conditions are required for AI-based reforms to contribute to a more inclusive tax system in the Moroccan context?

To address these questions, the study combines a systematic review of the international literature with empirical evidence from a structured questionnaire administered to tax officials, entrepreneurs and young taxpayers. The contribution is twofold. Empirically, it provides one of the first quantitative analyses of perceptions of AI in tax policy in Morocco, a middle-income country with a large informal sector and ongoing reforms. Conceptually, it links debates on AI and digital governance to the literature on inclusive taxation and fiscal legitimacy, showing that technological innovation can only support equity if it is embedded in broader governance reforms.

## 2. Literature Review

### 2.1. Artificial Intelligence and Smart Governance

AI has been increasingly applied in public governance to improve decision-making, optimize resource allocation, and enhance citizen services. AI-powered systems allow governments to process large volumes of data, detect irregularities, and design more transparent and efficient policies (Mikhaylov et al., 2018). In the fiscal domain, AI can be leveraged for predictive analytics, enabling tax administrations to better identify risks of non-compliance and reduce administrative burdens (Gupta et al., 2017). International experiences suggest that the success of AI in governance depends not only on technological innovation but also on institutional readiness, regulatory frameworks, and citizen trust (OECD, 2019).

### 2.2. Inclusive Tax Policy in Developing Economies

Inclusive taxation refers to the design of fiscal systems that ensure fairness, broaden the tax base, and reduce inequality. In developing economies, achieving inclusivity is particularly challenging due to the large size of the informal sector, weak administrative capacity, and low tax morale (Joshi et al., 2013). Research highlights that inclusivity requires not only efficient enforcement but also legitimacy and fairness in tax collection (Alm, 2019). Without addressing equity concerns, tax reforms often fail to gain public acceptance and can exacerbate inequalities, as evidenced by several cases in which reforms did not adequately account for structural vulnerabilities (Perry et al., 2007).

In this article, we define an inclusive tax policy as one that ensures broad and equitable participation in the tax system, minimizes the exclusion of vulnerable groups, and reduces barriers to compliance for small firms and informal workers. Inclusiveness, therefore, goes beyond revenue mobilization to encompass accessibility of procedures, affordability of compliance costs, and the ability of different categories of taxpayers to interact with the administration on fair terms. Fairness is understood in both horizontal and vertical dimensions: taxpayers with similar capacity to pay should bear a similar burden, while those with greater ability should contribute more, considering redistributive objectives and the specific constraints of small and medium-sized enterprises. In digital environments, fairness and inclusiveness also imply that technological innovations do not systematically disadvantage taxpayers with limited digital skills or access, and that AI-based systems are perceived as transparent, accountable and responsive to diverse taxpayer needs.

### 2.3. AI Applications in Taxation

Recent advances in AI have reshaped tax administration through digital platforms, e-invoicing, and real-time data collection. Studies from Latin America and Sub-Saharan Africa show that electronic invoicing improves compliance and reduces opportunities for fraud (Bellon et al., 2022). Machine learning algorithms are also increasingly used to detect anomalies in tax declarations, reducing human error and corruption risks (Mascagni et al., 2021). Despite these advances, scholars caution that digital taxation may unintentionally exclude vulnerable groups if technological adoption is uneven or if digital literacy remains low, which can undermine inclusivity and further marginalize small taxpayers (Junquera-Varela et al., 2022).

### 2.4. Conditions for Fair Implementation of AI in Fiscal Policy

The literature underlines several critical conditions for AI to foster inclusive tax policies. First, robust digital infrastructure is necessary to ensure access to online tax systems and the reliability of electronic platforms (Nose & Mengistu, 2023). Second, strong data governance frameworks are essential to protect taxpayer privacy, guarantee cybersecurity, and strengthen confidence in digital reforms (OECD, 2019). Third, citizen engagement and trust play a pivotal role, as public acceptance of AI-based tax reforms depends on perceptions of fairness, transparency as well as the extent to which reforms reduce compliance costs (Alm & Torgler, 2011). Finally, building capacity within both the tax administration and among taxpayers is indispensable, as officials require adequate training to manage AI systems, while citizens need sufficient digital literacy to interact effectively with them (Gupta et al., 2017).

### 2.5. The Moroccan Context and Research Gap

Morocco has embarked on ambitious fiscal reforms, including the digitalization of tax services and the introduction of professional income tax regimes (Ministère de l'Économie et des Finances, 2021). The informal economy, however, still accounts for more than 30% of

Domestic Product (GDP) (Haut-Commissariat au Plan, 2025), limiting the effectiveness of these reforms. While Morocco has made progress in adopting digital platforms for tax declaration and payment, there is limited research on how AI could reinforce inclusivity and fairness in this process. Most existing studies focus on legal reforms or descriptive analyses of informality, leaving a significant gap regarding the conditions under which AI can transform Morocco's fiscal system into a more inclusive model. This article seeks to address that gap by examining the institutional, social, and technological prerequisites for AI-driven tax reform in Morocco.

## 3. Methodology

This research adopts a quantitative approach to analyze the conditions for implementing AI in Morocco's tax policy in an inclusive and fair manner. Data were collected through a structured questionnaire administered between January and March 2025 to a sample of 220 respondents, including tax officials, entrepreneurs, and young taxpayers. The questionnaire was designed based on the international literature on digital taxation, tax morale, and AI in public finance, and adapted to the Moroccan context through consultations with local experts. A small pilot test with 10 respondents was conducted to verify the clarity of the questions and response scales, leading to minor wording adjustments before full deployment.

The questionnaire comprised four main modules:

1. perceptions of AI adoption in tax policy (*e.g.* expected effects on transparency, efficiency and fairness);
2. barriers to digitalization (infrastructure, costs, and skills);
3. institutional readiness (training, coordination, and data governance); and
4. attitudes toward tax compliance and trust in the tax administration.

Most items used five-point Likert scales, complemented by multiple-choice questions on access to digital tools and open-ended questions on perceived risks and opportunities.

The sampling strategy combined purposive and snowball methods to ensure coverage of key categories most affected by tax reforms. Tax officials were contacted through administrative networks, while entrepreneurs and young taxpayers were reached via professional associations, business support structures and university networks. Respondents represented diverse sectors such as services, trade, and small-scale manufacturing, as well as different age groups, allowing for both individual-level and subgroup-level analysis. While the sample is not statistically representative of the entire Moroccan taxpayer population, it captures a range of actors who are directly concerned by ongoing and prospective digital reforms.

Data analysis proceeded in two stages. First, descriptive statistics (percentages and frequency distributions) were employed to capture the main trends and barriers identified by respondents. Second, cross-tabulations and subgroup comparisons were carried out, particularly by age group and sector of activity, to highlight variations in perceptions of AI adoption. These comparisons are exploratory and non-inferential: they aim to identify patterns and contrasts rather than to test formal statistical hypotheses. Differences across groups are therefore interpreted as indicative trends in perceptions, not as statistically validated effects. Finally, the findings were compared with international experiences documented in the literature, particularly from Sub-Saharan Africa and Latin America, in order to situate the Moroccan case within broader global debates on digital governance and tax reform. This methodological design provides both empirical rigor and contextual relevance, ensuring that the analysis addresses the research problem in a comprehensive and convincing manner.

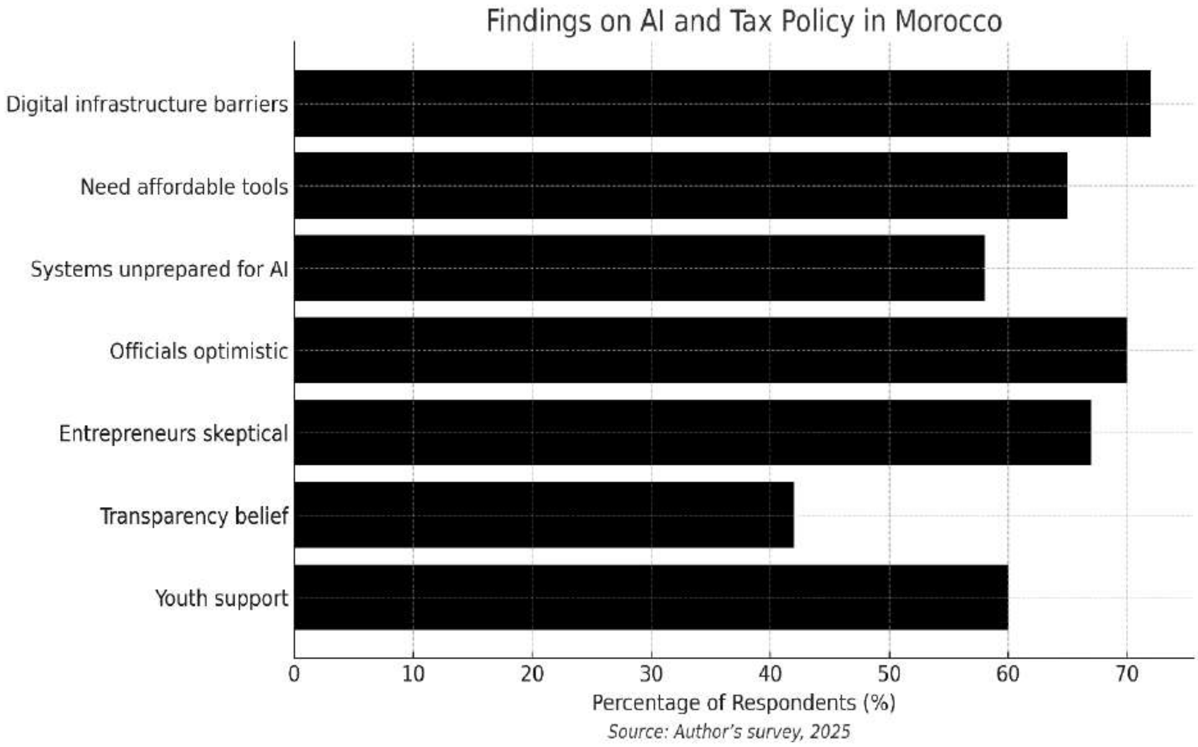
## 4. Results

### 4.1. Adoption Barriers of AI in Moroccan Tax Policy

Survey results reveal that the main obstacle to the integration of AI in Morocco's tax administration is the weakness of the digital infrastructure. Nearly 72% of respondents identified insufficient connectivity, lack of integrated databases, and outdated systems as the primary barriers to successful adoption. These findings echo international studies emphasizing the crucial role of infrastructure in enabling digital fiscal reforms (OECD, 2022). Moreover, 65% of small entrepreneurs reported the high cost of digital tools as a significant constraint, reflecting the need for affordable and inclusive technological solutions.

**Figure 1**

*Findings on AI and Tax Policy in Morocco.*




As illustrated in Figure 1, weaknesses in digital infrastructure clearly emerge as the most frequently cited barrier, ahead of the cost of digital tools and limited administrative capacity.

Figure 1 visually summarizes these adoption constraints and highlights the extent to which infrastructure and affordability shape perceptions of AI in tax policy.

### 4.2. Institutional Preparedness and Human Resources

The data also indicate that 58% of surveyed officials believe that current administrative systems are unprepared to integrate AI-driven tools. A majority emphasized the lack of technical training and expertise among tax officials, which undermines the credibility and efficiency of reforms. Previous studies highlight that without substantial investment in capacity building and digital literacy, AI projects in the public sector tend to face high failure rates (Mikhaylov et al., 2018). This underlines the urgency of aligning technological change with institutional development.

### 4.3. Perceptions of Fairness and Compliance

Perceptions regarding the fairness of AI-based taxation are divided. While 70% of tax officials expressed optimism that AI could enhance transparency, 67% of entrepreneurs were more skeptical, fearing that automation could disproportionately affect small businesses. Only 42% of respondents believed that AI would strengthen trust between taxpayers and the state, confirming earlier research that technology alone cannot overcome deep-rooted issues of tax morale and institutional trust (OECD, 2019).

### 4.4. Opportunities for Inclusive Reform

Despite these challenges, the survey highlights promising avenues for reform. Nearly 60% of young respondents expressed support for digitalization, viewing AI as a tool for modernization and simplification of procedures. This generational openness suggests that inclusive tax policy, if combined with awareness campaigns and progressive incentives, could leverage digital innovations to strengthen compliance. Such findings align with broader debates on digital governance and fiscal inclusion in developing economies (UNCTAD, 2021).

**Table 1**

*Key Findings from the Survey on AI and Tax Policy in Morocco*

| Dimension | Percentage of Respondents | Key Insights |
|---|---|---|
| Weak digital infrastructure | 72% | Core barrier to AI adoption; outdated databases and poor connectivity |
| High cost of digital tools | 65% | Small entrepreneurs lack affordable access to AI-enabled platforms |
| Unprepared administrative systems | 58% | Limited training and expertise among tax officials |
| Officials optimistic | 70% | Believe AI will enhance transparency |
| Entrepreneurs skeptical | 67% | Fear disproportionate burden on small firms |
| AI improves trust | 42% | Limited confidence in AI to increase tax morale |
| Youth support for AI | 60% | Younger respondents view AI as an opportunity for modernization |

The main quantitative results are summarized in Table 1, which brings together the key percentages reported across the different dimensions of the survey. Table 1 shows that weak digital infrastructure, high costs of digital tools and limited administrative preparedness jointly constrain the deployment of AI, while at the same time highlighting the coexistence of optimism among officials and skepticism among entrepreneurs regarding the fairness and transparency of AI-based reforms.

**Table 2**

*Support for AI by Age and Sector.*

| Group | Support (%) | Key Observations |
|---|---|---|
| Youth (<35 years) | 60% | Stronger openness to innovation |
| Older taxpayers | 38% | More cautious, perceive risks |
| Service sector firms | 55% | Expect simplification of processes |
| Small manufacturers | 41% | Associate AI with compliance risks |

As shown in Table 2, support for AI is significantly higher among younger respondents (under 35), who tend to view digital tools as an opportunity for modernization and simplification of procedures. By contrast, older taxpayers and small manufacturers are more cautious and associate AI with higher compliance risks. These differences are descriptive rather than statistically tested, but they suggest that both age and sector shape how taxpayers perceive the promise and risks of AI-based tax reforms.

### 4.5. International Comparisons

These results resonate with experiences in other developing economies. In Sub-Saharan Africa, presumptive taxes and partial digital reforms have struggled to integrate informal actors (Fjeldstad & Heggstad, 2012). In Latin America, while digitalization initiatives improved monitoring, informality persists due to structural inequalities (Perry et al., 2007). The Moroccan case illustrates similar tensions: modernization initiatives are promising but remain constrained by infrastructure gaps, capacity deficits, and persistent distrust.

### 4.6. Critical Insights

Taken together, the findings highlight that AI is not a panacea for fiscal inclusion. While it can streamline processes and improve transparency, structural, institutional, and behavioral barriers remain significant. Without addressing affordability, trust, and administrative preparedness, AI risks reproducing existing inequalities instead of resolving them. These results provide critical insights for policymakers designing inclusive tax reforms.

## 5. Discussion

The results of this study highlight the complex interplay between technological innovation, institutional preparedness, and social perceptions in shaping the implementation of

AI-based tax reforms in Morocco. Consistent with international findings, digital infrastructure deficiencies and the high cost of technological tools emerge as central barriers to adoption. This confirms that reforms cannot be reduced to the introduction of new technologies but must be accompanied by investments in infrastructure and affordable access (OECD, 2022).

Institutionally, the limited readiness of tax administrations reflects a structural weakness that is also common across developing economies (Mikhaylov et al., 2018). The lack of technical expertise and training among tax officials not only slows down adoption but also reduces the legitimacy of reforms in the eyes of taxpayers. Strengthening human capital through targeted capacity-building initiatives appears essential for ensuring that AI contributes to genuine modernization rather than reinforcing bureaucratic inefficiencies.

From a behavioral perspective, the results underline the persistent problem of trust. The skepticism expressed by entrepreneurs, especially in small-scale manufacturing, reveals the risk that AI reforms may exacerbate existing perceptions of inequality. This finding resonates with Alm and Torgler (2011), who argue that compliance depends not only on enforcement but also on perceptions of fairness and the quality of governance. An important dimension of trust that emerges from the literature, but is only indirectly captured in our survey, concerns data security and privacy in digital tax systems. AI-enabled tax administration relies on large-scale collection and processing of sensitive information, which may reinforce distrust if taxpayers fear misuse or cyber vulnerabilities. Experiences from other developing and emerging economies show that digitalization can strengthen compliance only when taxpayers perceive that their data are adequately protected and that digital systems are transparent and accountable (OECD, 2021). In the Moroccan context, this implies that AI reforms must be accompanied not only by investments in infrastructure and training, but also by robust data governance frameworks, clear communication on privacy safeguards, and mechanisms for redress in the event of errors or perceived unfair treatment. Without such safeguards, the risk is that small taxpayers and less digitally literate groups will either disengage from formal channels or perceive AI-based systems as instruments of control rather than tools for fairness.

Comparisons with Sub-Saharan Africa and Latin America indicate that Morocco faces similar challenges: digitalization can improve transparency and monitoring, but without addressing structural inequalities and building trust, informal actors remain excluded from the fiscal net (Fjeldstad & Heggstad, 2012). This suggests that AI has the potential to support inclusive tax policy, but only under conditions where reform is embedded within broader governance reforms.

Overall, the findings point to a critical insight: AI is not a substitute for inclusive tax policy but a complement to it. To avoid reproducing inequalities, policymakers must combine technological adoption with targeted incentives, institutional strengthening, and efforts to rebuild taxpayer trust. This integrated perspective is particularly relevant in the Moroccan context, where the informal sector represents a significant share of the economy and where fiscal legitimacy remains fragile.

It is important to note, however, that our survey does not fully cover informal workers and micro-entrepreneurs operating entirely outside the formal tax net. Their views are only indirectly reflected through some small entrepreneurs at the margins of formality. This constitutes a significant limitation, given the central role of the informal sector in debates on inclusiveness, and points to the need for dedicated qualitative and quantitative research focusing specifically on informal actors.

## 6. Conclusion

This article has examined the conditions for implementing AI in Morocco's tax policy with a particular focus on inclusiveness and fairness. Based on a quantitative survey combined with a review of the literature, the study shows that while AI offers opportunities to enhance transparency and efficiency, its successful integration is constrained by digital infrastructure deficits, limited institutional readiness, and persistent skepticism among taxpayers. These barriers indicate that AI is not a panacea but must be embedded within broader governance reforms that address issues of trust, capacity, and equity.

The findings make two key contributions to the literature. First, they demonstrate that the effectiveness of AI in fiscal reforms depends less on the technology itself and more on the socio-institutional context in which it is deployed. Second, they highlight the importance of generational and sectoral differences in shaping attitudes toward digital reforms, thereby offering insights into how inclusive policies can be designed.

Nevertheless, the study is not without limitations. The reliance on a cross-sectional survey limits the ability to capture the evolution of perceptions over time or to disentangle the causal impact of AI reforms from broader changes in the policy environment. Moreover, the sample, while diverse in terms of sectors and age groups, does not include fully informal workers and micro-entrepreneurs who remain outside the formal tax net, even though they are central to debates on inclusive taxation in Morocco. Finally, the survey captures perceptions of trust and fairness in general terms but does not explicitly measure concerns related to data security and privacy, which are increasingly salient in digital tax systems.

Future research could build on these findings by employing longitudinal designs, expanding the sample to include informal sector actors and vulnerable groups, and integrating survey modules that explicitly address issues of data protection, cybersecurity and algorithmic fairness. Comparative studies with other developing countries undergoing similar fiscal digitalization reforms would also help to identify which combinations of institutional arrangements, technological choices and trust-building strategies are most conducive to inclusive AI-based tax policy.

In practical terms, the results suggest that Moroccan policymakers should prioritize investment in digital infrastructure, provide affordable access to AI-enabled tools for small enterprises, and strengthen the technical capacity of tax officials. Equally important is the need to rebuild taxpayer trust through transparency, fairness, and inclusive dialogue. By addressing these conditions, Morocco can ensure that AI becomes not only a tool of modernization but also a driver of fiscal equity and sustainable development.

## References

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys, 33*(2), 353–388. https://doi.org/10.1111/joes.12272

Alm, J., & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics, 101*(4), 635–651. https://doi.org/10.1007/s10551-011-0761-9

Bellon, M., Dabla-Norris, E., Khalid, S., & Lima, F. (2022). Digitalization to improve tax compliance: Evidence from VAT e-invoicing in Peru. *Journal of Public Economics, 210*, 104661. https://doi.org/10.1016/j.jpubeco.2022.104661

Fjeldstad, O.-H., & Heggstad, K. K. (2012). *Building taxpayer culture in Mozambique, Tanzania and Zambia: Achievements, challenges and policy recommendations*. Chr. Michelsen Institute. https://www.cmi.no/publications/4499-building-taxpayer-culture-in-mozambique

Gupta, S., Keen, M., Shah, A., & Verdier, G. (Eds.). (2017). *Digital revolutions in public finance*. International Monetary Fund. https://doi.org/10.5089/9781484315224.071

Haut-Commissariat au Plan. (2025). *Rapport des résultats de l'Enquête Nationale sur le Secteur Informel 2023/2024*. HCP. https://www.hcp.ma/Rapport-des-resultats-de-l-Enquete-Nationale-sur-le-Secteur-Informel-2023-2024-Mai-2025_a4110.html

Joshi, A., Prichard, W., & Heady, C. (2013). Taxing the informal economy: Challenges, possibilities and remaining questions. */IDS Working Paper, 2013*(429), 1–37. https://doi.org/10.1111/j.2040-0209.2013.00429.x

Junquera-Varela, R. F., Lucas-Mas, C. Ó., Krsul, I., Calderon Yksic, V. O., & Arce Rodriguez, P. (2022). *Digital transformation of tax and customs administrations*. World Bank. https://openknowledge.worldbank.org/entities/publication/8f69a6d3-4633-5dd4-9ddb-d1c21f315f23

Mascagni, G., Mengistu, A. T., & Woldeyes, F. B. (2021). Can ICTs increase tax compliance? Evidence on taxpayer responses to technological innovation in Ethiopia. *Journal of Economic Behavior & Organization, 189*, 172–193. https://doi.org/10.1016/j.jebo.2021.06.007

Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 20170357. https://doi.org/10.1098/rsta.2017.0357

Ministère de l’Économie et des Finances. (2021). *La réforme fiscale*. https://www.finances.gov.ma/fr/Reformes/Pages/Reforme-fiscale.aspx

Nose, M., & Mengistu, A. (2023). *Exploring the adoption of selected digital technologies in tax administration: A cross-country perspective* (IMF Notes No. 2023/008). International Monetary Fund. https://doi.org/10.5089/9798400258183.068

OECD. (2019). *Tax morale: What drives people and businesses to pay tax?* OECD Publishing. https://doi.org/10.1787/f3d8ea10-en

OECD. (2021). *Supporting the digitalisation of developing country tax administrations*. OECD Publishing. https://doi.org/10.1787/8600724a-en

OECD. (2022). *Tax administration 2022: Comparative information on OECD and other advanced and emerging economies*. OECD Publishing. https://doi.org/10.1787/1e797131-en

Okunogbe, O., & Santoro, F. (2023). The promise and limitations of information technology for tax mobilization. *The World Bank Research Observer, 38*(2), 295–324. https://doi.org/10.1093/wbro/lkac008

Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., & Saavedra-Chanduvi, J. (2007). *Informality: Exit and exclusion*. World Bank. https://ideas.repec.org/b/wbk/wbpubs/6730.html

UNCTAD. (2021). *Digital economy report 2021: Cross-border data flows and development*. United Nations. https://unctad.org/publication/digital-economy-report-2021

# ARTIFICIAL INTELLIGENCE AS A NORMATIVE ACTOR: TOWARDS A SILENT PRIVATIZATION OF LAW

**Youmna TOUBI & Amina KHALLOUFI**

*Chouaib Doukkali University,*
*FSJES, El Jadida, Morocco*

**ABSTRACT**

As Artificial Intelligence (AI) systems increasingly permeate decision-making processes in sectors such as finance, taxation, public administration, and justice, they are generating real-world effects that shape behaviors, allocate rights, and exclude individuals—all without formal legal validation or democratic control. This paper argues that AI is emerging as a normative actor, capable of silently reshaping the legal order by imposing behavioral standards through opaque, privately controlled systems. Such a phenomenon represents what can be described as a silent privatization of law, raising critical concerns regarding legal certainty, state sovereignty, and the protection of fundamental rights. This study adopts a critical legal and interdisciplinary methodology, combining a review of regulatory frameworks (including the European Union Artificial Intelligence Act (EU AI Act), the General Data Protection Regulation (GDPR), and international ethical instruments) with an in-depth case study of Morocco. It examines how AI is currently being deployed in Moroccan public and private sectors, particularly in financial regulation and anti-corruption strategies, highlighting regulatory gaps and asymmetries in contexts where digital governance frameworks remain underdeveloped. A comparative perspective is integrated through reference to European and international models, allowing for a broader reflection on global governance challenges. Findings suggest that the normative power of AI is growing in ways that often bypass legal accountability and democratic oversight. In countries like Morocco, this trend is intensified by technological dependency and limited institutional capacity to regulate algorithmic systems. The paper concludes by calling for a reaffirmation of legal sovereignty in the face of technological normativity, advocating for stronger regulation, transparent algorithmic governance, and the enforcement of rights to explanation and recourse. These are essential steps to ensure that legal certainty and the rule of law are not eroded in the age of intelligent systems.

***Keywords:*** *Algorithmic Governance, Privatization of Law, Legal Certainty, Digital Sovereignty, Fundamental Rights.*

## 1. Introduction

Artificial Intelligence (AI) has transitioned from a mere technological instrument to a substantive normative force that progressively influences human conduct and societal structure. Algorithms now govern credit scoring, financial transactions, and even judicial reasoning, yielding effects analogous to those of legal norms (Zuboff, 2019). But these algorithmic standards are made and controlled by private groups that are not subject to normal democratic oversight. This creates a parallel normative order where technological design takes the place of legal deliberation. Lessig (2006) famously said, "code is law": the way software systems are built controls how people act just as well as laws do. The legitimacy of these "code-based" norms is dubious, as they embody corporate interests and market logic rather than the collective will (Balkin, 2018).

This phenomenon represents a covert privatization of law—covert in that it functions imperceptibly under the pretext of efficiency, and privatization in that normative authority transitions from the State to corporations. The black box problem, which is the lack of transparency in AI, makes legal certainty and due process less reliable (Burrell, 2016). When algorithmic systems determine who is credited or investigated, individuals forfeit their capacity to anticipate legal repercussions and contest automated results (Yeung, 2018). This loss of predictability and accountability puts the rule of law in danger.

The AI Act (European Parliament & Council of the European Union, 2024), the GDPR (2016), and ethical frameworks like the UNESCO (2021) recommendations are examples of the European Union's regulatory response. They show an effort to bring back normative transparency. These tools try to find a balance between innovation and legal certainty. This balance is still fragile in new places like Morocco, where digital transformation is moving faster than legal adaptation. Even though there are programs like Digital Morocco 2020, the lack of a complete AI framework makes it hard to balance technological ambition with normative governance (Ministry of Digital Transition and Administrative Reform, 2024). Current legislation, particularly Law 09-08 concerning personal data and Law 43-05 regarding anti-money laundering, provides limited protections while failing to tackle systemic algorithmic risks.

This article argues that AI acts as a new normative actor, redistributing power between states, markets, and technologies. It examines how this silent privatization of normativity challenges legal sovereignty and security, focusing on the Moroccan case in comparison with European regulatory approaches. The central question is therefore twofold: *To what extent does*

*AI contribute to the privatization of law, and how can states preserve legal certainty and fundamental rights amid this transformation?*

**2. Literature Review**

Scholarly research is progressively acknowledging artificial intelligence (AI) as a normative phenomenon rather than merely a technological tool. Lawrence Lessig's early work (2006) laid the groundwork for the idea that "code is law," which means that digital architectures can control behavior just as well as laws can. Balkin (2018) and Zuboff (2019) have demonstrated that algorithmic infrastructures represent novel forms of private governance, wherein corporations perform regulatory roles typically associated with states. This privatization of normativity transpires via opaque, data-driven systems that delineate rights, obligations, and resource access devoid of democratic accountability.

The lack of transparency in algorithmic decision-making has been thoroughly investigated in socio-legal research. Burrell (2016) characterized the "black box problem" as a structural impediment to transparency, whereas Yeung (2018) contended that algorithmic regulation alters legal reasoning from normative interpretation to probabilistic prediction. These developments undermine the fundamental tenets of the rule of law, especially the principles of foreseeability and accountability that are essential for legal certainty (Hart, 2012). Hildebrandt (2015) posited that smart technologies alter the relationship between individuals and norms, resulting in what she terms "computational legal subjectivity." Together, these contributions show that AI does not just help people make decisions; it also changes the structure of legal power.

The European Union has become a leader in trying to bring AI under regulatory oversight within the legal system at the policy level. The AI Act (Regulation (EU) 2024/1689), the GDPR (2016), and soft-law tools like UNESCO (2021) try to make technology systems more open, fair, and responsible. These frameworks show a move toward rights-based governance of AI, which strikes a balance between progress and basic rights. However, researchers point out that these models assume that institutions have strong abilities that aren't always there (Hildebrandt, 2015).

In developing economies, the literature emphasizes an increasing disparity between technological adoption and normative regulation. Moroccan studies (Ministry of Digital Transition and Administrative Reform, 2024; CNDP, 2023) indicate that although initiatives like Digital Morocco 2020 have enhanced digital infrastructure, the lack of comprehensive AI regulation leaves users vulnerable to unclear algorithmic practices. Laws No. 09-08 on data protection and No. 43-05 on anti-money laundering are some examples of legal protections that provide partly coverage of the risks of algorithmic governance. Comparative scholars advocate for a framework of algorithmic constitutionalism to guarantee that AI systems function within well-defined and enforceable legal parameters (Balkin, 2018; Hildebrandt, 2015).

In conclusion, the literature converges on a significant revelation: AI has emerged as a normative entity that contests the state's monopoly on legislation. Discussions in theory, law, and policy are now centered on how to balance algorithmic innovation with legal clarity, openness, and human rights, particularly in situations where governance frameworks are still being established.

## 3. Methodology

The research adopts a comparative and interdisciplinary approach, combining doctrinal legal analysis with insights from the sociology of law and critical legal studies. It examines key regulatory instruments (the EU AI Act, the GDPR, the OECD and the UNESCO frameworks) and relevant jurisprudence to trace how normative authority is redistributed through AI systems. The study then conducts an in-depth case study of Morocco, analyzing national policies, legislative instruments, and institutional initiatives in digital governance. This dual perspective allows the identification of both regulatory asymmetries and structural vulnerabilities in emerging legal systems. By confronting European rights-based models with Morocco’s evolving framework, the study aims to propose elements of algorithmic constitutionalism—a legal paradigm ensuring that AI operates within transparent, accountable, and rights-respecting boundaries.

## 4. Results

The examined jurisprudential and doctrinal corpus indicates converging trends within European and Moroccan legal systems concerning the conflict between algorithmic decision-making and the rule of law.

### 4.1. Algorithmic Decision-Making and Legal Norms

In Europe, courts have gradually seen data protection and proportionality as ways to limit automated systems. The CJEU in Nowak v. Data Protection Commissioner (2017) acknowledged the broad definition of personal data as including any information associated with an identifiable individual, thereby bolstering human-centered governance. The Court struck down the EU–US Privacy Shield in Schrems II (2020), saying that data transfer systems and algorithmic processes must not undermine the essence of fundamental rights or take away people's right to go to court. These rulings establish the judicial foundation for future AI accountability: algorithmic efficiency cannot override legality.

This position is strengthened by parallel jurisprudence from the European Court of Human Rights (ECtHR). The Court ruled in Szabó and Vissy v. Hungary (2016) that indiscriminate surveillance violates Article 8 of the ECHR unless it is done with judicial safeguards and oversight. In the same way, Digital Rights Ireland (2014) and Tele2 Sverige (2016) struck down laws that required the storage of large amounts of data because they didn't pass tests for proportionality and foreseeability. These decisions create a legal framework where openness and human oversight are necessary for legality.

### 4.2. Doctrinal Outcomes: Algorithmic Normativity and Legal Certainty

Doctrinal reflection substantiates that AI engenders a novel form of normativity, frequently circumventing conventional legislative procedures. Lessig (2006) said, "code is law." Algorithms organize behavior by design, not by order. This understanding forms the basis of algorithmic constitutionalism (Celeste, 2019), which integrates constitutional principles—such as transparency, accountability, and rights—into the architecture of digital systems.

Nevertheless, academics caution that algorithmic governance undermines legal certainty. Pasquale (2015) shows that predictive algorithms turn law's ex ante rationality into ex post optimization by replacing clear rules with probabilistic correlations. These kinds of systems make results that are hard to predict and can't be checked, which makes it hard for people to know what will happen legally.

National institutions are starting to recognize these risks. The Conseil d’état (2022) said that public algorithms must always be able to be checked and explained. The CNIL (2023) made the same point about administrative AI, saying that it must be clear and fair. These standards are similar to the ECtHR's Kruslin v. France (1990) and Rotaru v. Romania (2000), in which the Court said that any interference with rights must be predictable and available.

### 4.3. Comparative Findings: The Developing Jurisprudence of Morocco

In Morocco, the jurisprudence is still in its early stages, but it is in line with trends around the world. Article 6 of the 2011 Constitution establishes legality and equality as the fundamental principles of governance. In Decision No. 314/2018 of the *Cour de Cassation*, the *Principle of Sécurité Juridique* was reaffirmed, stressing the necessity for stable and comprehensible norms. In the same way, Decision No. 502/2021 from the *Tribunal de*

*Commerce* de Casablanca said that digital contracts and algorithmic audit trails were valid as long as they were clear and agreed upon.

Nonetheless, doctrinal analysis reveals enduring regulatory asymmetries. Law No. 09-08 (data protection) and Law No. 43-05 (AML) are the current laws, but they do not yet cover algorithmic transparency or accountability. Judges in Morocco are seeing more and more evidence that was made by algorithms, but they don't have the technical knowledge to tell if it's reliable. Without changes to the way things are done in institutions, the courts may rely on technological experts, which would mean that norms are being created in private.

These results show that the process of legal adaptation is changing but not yet finished. European courts are working on a doctrine of algorithmic accountability, but Morocco is still in the stage of normative approximation, trying to find a balance between digital modernization and legal sovereignty.

## 5. Discussion

The examined jurisprudential and doctrinal corpus indicates converging trends within European and Moroccan legal systems concerning the tension between algorithmic decision-making and the rule of law.

### 5.1. Algorithmic Decision-Making and Legal Principles

At the European level, courts have increasingly construed data protection and proportionality as means to limit automated systems. The CJEU in Nowak v. Data Protection Commissioner (2017) acknowledged the broad definition of personal data as including any information associated with an identifiable individual, thereby bolstering human-centered governance. In Schrems II (2020), the Court struck down the EU–US Privacy Shield, saying that data transfer rules and algorithmic processes must not undermine the core of fundamental rights or take away people's right to go to court. These rulings establish the legal foundation for future AI accountability: algorithmic efficiency cannot take precedence over legality.

The ECtHR's parallel jurisprudence backs up this point of view. The Court ruled in Szabó and Vissy v. Hungary (2016) that indiscriminate surveillance violates Article 8 of the ECHR unless it is done with judicial safeguards and oversight. Likewise, Digital Rights Ireland (2014) and Tele2 Sverige (2016) annulled bulk data retention laws for not passing proportionality and foreseeability tests. These decisions create a legal framework in which transparency and human oversight are essential for legality.

### 5.2. Doctrinal Outcomes: Algorithmic Normativity and Legal Certainty

Doctrinal analysis substantiates that AI engenders a novel form of normativity, frequently circumventing conventional legislative procedures. Lessig (2006) said, "code is law"; algorithms set rules for behavior by design, not by decree. This understanding forms the basis of algorithmic constitutionalism (Celeste, 2019), which integrates constitutional principles—transparency, accountability, and rights—into the architecture of digital systems.

Nevertheless, scholars caution that algorithmic governance undermines legal certainty. Pasquale (2015) illustrates that predictive algorithms supplant explicit regulations with probabilistic correlations, converting the law's *ex ante* rationality into *ex post* optimization. These kinds of systems make results that are hard to predict and can't be checked, which makes it harder for people to know what will happen in the law.

These risks are starting to be recognized by national institutions. The Conseil d'État (2022) said that public algorithms must still be able to be checked and explained. The CNIL (2023) said the same thing about administrative AI, saying it must be open and fair. These standards are similar to the ECtHR's Kruslin v. France (1990) and Rotaru v. Romania (2000), in which the Court said that any interference with rights must be predictable and available.

### 5.3. Comparative Findings: The Developing Jurisprudence of Morocco

In Morocco, the legal system is still in its early stages, but it is in line with global trends. Article 6 of the 2011 Constitution establishes legality and equality as the fundamental principles of governance. The *Principle of Sécurité Juridique* was reaffirmed in *Cour de Cassation*, Decision No. 314/2018, which stressed the need for stable and clear rules. In the same way, *Tribunal de commerce* de Casablanca, Decision No. 502/2021, said that digital contracts and algorithmic audit trails were valid as long as they were clear and agreed upon.

#### 5.3.1.However, doctrinal From Efficiency to Normative Authority

The findings indicate that AI has transitioned from a neutral tool to an agent of normative authority. Algorithms replicate the regulatory function of law through automated classification, prediction, and decision-making, but they lack its procedural safeguards. Zuboff (2019) says that this is the beginning of instrumentarian power, which means that control is based on predicting data instead of force.

As a result, algorithmic standards become the rules that decide who can get credit, how much they have to pay in taxes, or how much they can use public services (Yeung, 2018). Their legitimacy is based on how well they work and how accurate they are, not on whether or not people agree with them. This change is what Hildebrandt (2015) calls the "silent privatization of law": the move of legal functions into the world of private technology design. Computational opacity and corporate secrecy take the place of Fuller's (1969) principles of legality, which are clarity, publicity, and congruence.

#### 5.3.2. Ways to privatize

This privatization is supported by three interconnected mechanisms: opacity, dependency, and standardization.

- Opacity comes from the fact that machine learning is hard to understand and that trade secrets are kept safe. This makes it hard for people to understand or challenge the results of algorithms (Burrell, 2016). Judicial oversight is consequently obstructed, and citizens forfeit the capacity to anticipate legal repercussions.
- Dependency emerges when public authorities depend on private vendors for essential algorithmic infrastructure. Hildebrandt (2015) says that this kind of technological lock-in makes it almost impossible to get out. In Morocco, dependence on foreign software providers for taxation and governance systems (Ministry of Digital Transition and Administrative Reform, 2024) threatens to undermine normative sovereignty.
- Standardization shows how many private ethical codes there are, like Google's AI Principles or Microsoft's Responsible AI Standard. These codes act like laws but don't have the same democratic legitimacy (Crawford, 2021). The UNESCO (2021) Recommendations are examples of global soft law tools that try to stop this drift, but they are still optional (Abbott & Snidal, 2000).

#### 5.3.3. Legal Certainty and Rights Privatization

This topic poses a direct threat to legal certainty by causing algorithmic outputs to change over time, which makes them less predictable. Hart (2012) posited that the stability of law relies on a collective acknowledgment of rules as obligatory; this "internal point of view" disintegrates when norms are entrenched in opaque code. Additionally, algorithmic bias perpetuates inequality, contravening Article 14 of the ECHR and Article 6 of the Moroccan Constitution (O'Neil, 2016). The lack of procedural protections, like the right to an explanation or an appeal, makes effective remedies less effective (Article 13 ECHR).

In Morocco, there are no clear ways for the courts to review administrative decisions that are based on AI. If an algorithm incorrectly categorizes a taxpayer or refuses a benefit, the individual lacks legal recourse to contest the system's rationale. This gap erodes citizens' trust and the legitimacy of institutions.

#### 5.3.4. Insights that are both comparative and forward-looking

The AI Act (Regulation (EU) 2024/1689) is an example of a regulatory response that tries to bring algorithmic power under the rule of law. It puts into action the ideas of digital constitutionalism (Celeste, 2019) by requiring risk-based classifications, documentation, and human oversight. Nonetheless, researchers emphasize enduring deficiencies in enforcement and excessive dependence on corporate self-evaluation (Veale & Borgesius, 2021).

Morocco, on the other hand, is still weak in terms of institutions. Relying on foreign digital ecosystems is a kind of technological neo-colonialism. Morocco needs to create hybrid accountability frameworks that include judicial review, algorithmic audits, and participatory governance in order to regain its sovereignty. Working with the *Commission Nationale pour la Protection des Données* (CNDP) and the courts to set up a National AI Authority could make sure that things are fair, open, and in line with constitutional rights.

#### 5.3.5. Taking Back the Public Nature of the Law

In the end, the conversation supports a key normative claim: AI's privatization of the law must be countered by its constitutional re-publicization. Integrating constitutional protections into algorithmic systems—termed algorithmic constitutionalism by Hildebrandt (2015) and Celeste (2019)—provides a foundation for resilience. The law must reassert its primacy over code to make sure that government stays fair, open, and predictable. In Europe and Morocco, the main question of the algorithmic age is not whether machines can rule, but what laws they used to do so.

## 6. Conclusion

Artificial intelligence has emerged as a true normative actor, subtly transforming the legal order. Through algorithmic systems that guide decisions in finance, taxation, justice, and administration, AI produces behavioral standards comparable to legal norms—yet without democratic legitimacy or institutional accountability. This phenomenon reflects a *silent privatization of law*, as private technological infrastructures increasingly assume regulatory functions once reserved for the State.

Such evolution threatens two essential foundations of legality: legal certainty and democratic legitimacy. Algorithmic reasoning, probabilistic and opaque, weakens predictability and citizens' capacity to contest decisions. At the same time, governance shifts from public deliberation to private design, replacing law's rational authority with technological determinism masked as neutrality.

Internationally, the European Union's AI Act**,** supported by the GDPR and the Charter of Fundamental Rights, marks a major step toward reasserting public control over algorithmic systems. It translates constitutional values into digital governance but still relies heavily on private self-assessment—showing the persistent asymmetry between public authority and corporate power. Nonetheless, it provides a promising model of *algorithmic constitutionalism*, ensuring that technology remains subordinate to rights and legality.

For Morocco, the challenge is to reconcile rapid digitalization with normative sovereignty. The absence of clear AI accountability rules risks deepening dependence on foreign standards. Yet this also offers an opportunity to design an adaptive governance model rooted in constitutional principles of transparency, proportionality, and human oversight. Strengthening institutions such as the CNDP, building judicial expertise, and creating a national authority for AI ethics could help anchor digital modernization within the rule of law.

Ultimately, the rise of AI urges jurists to redefine the rule of law in the algorithmic age. Law must remain the language of collective justification—capable of channeling innovation through principles of rights, equality, and dignity. The privatization of law by code is not inevitable; it is a call to reaffirm that intelligence, however artificial, must remain subject to human law and democratic will.

## References

Abbott, K. W., & Snidal, D. (2000). Hard and soft law in international governance. *International Organization, 54*(3), 421–456. https://doi.org/10.1162/002081800551280

Balkin, J. M. (2018). Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *UC Davis Law Review, 51*(3), 1149–1210. https://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/51-3_Balkin.pdf

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Celeste, E. (2019). Digital constitutionalism: A new systematic theorization. *International Review of Law, Computers & Technology, 33*(1), 76–99. https://doi.org/10.1080/13600869.2019.1562604

Commission Nationale de Contrôle de la Protection des Données à Caractère Personnel (CNDP). (2023). *2022 annual report* .https://cnpd.public.lu/dam-assets/fr/publications/rapports/cnpd/rapport-annuel-annexes-2022-cnpd.pdf

Commission Nationale de l'Informatique et des Libertés (CNIL). (2023). *Artificial intelligence: CNIL action plan*. CNIL. https://www.cnil.fr/fr/intelligence-artificielle-le-plan-daction-de-la-cnil

Conseil d'État. (2022). *Intelligence artificielle et action publique : Construire la confiance, servir la performance*. https://www.conseil-etat.fr/publications-colloques/etudes/etudes-a-la-demande-du-gouvernement/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance

Council of Europe. (1950). *European Convention on Human Rights*. Council of Europe. https://www.echr.coe.int/documents/d/echr/convention_ENG

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (COM(2021) 206 final)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

Fuller, L. L. (1969). *The morality of law* (Rev. ed.). Yale University Press.

Hart, H. L. A. (2012). *The concept of law* (3rd ed., L. Green, Ed.). Oxford University Press.

Hildebrandt, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing.

Lessig, L. (2006). *Code: Version 2.0*. Basic Books.

Ministry of Digital Transition and Administrative Reform. (2024). *Digital Morocco 2030*. https://www.mmsp.gov.ma/fr/actualites/cliquez-ici-pour-decouvrir-la-strategie-digital-morocco-2030

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act: Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International, 22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

**Jurisprudence**

Court of Justice of the European Union. (2014). *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Kärntner Landesregierung* (Joined Cases C-293/12 and C-594/12, ECLI: EU: C :2014:238). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62012CJ0293

Court of Justice of the European Union. (2016). *Tele2 Sverige AB v Post-och telestyrelsen and Secretary of State for the Home Department v Watson and Others* (Joined Cases C-203/15 and C-698/15, ECLI: EU: C :2016:970). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62015CJ0203

Court of Justice of the European Union. (2017). *Peter Nowak v Data Protection Commissioner* (Case C-434/16, ECLI: EU: C :2017:994). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62016CJ0434

Court of Justice of the European Union. (2020). *Data Protection Commissioner v Facebook Ireland Ltd and Maximillian Schrems (Schrems II)* (Case C-311/18, ECLI: EU: C :2020:559). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62018CJ0311

European Court of Human Rights. (1990). *Kruslin v. France* (Application No. 11801/85). https://hudoc.echr.coe.int/eng?i=001-57626

European Court of Human Rights. (2000). *Rotaru v Romania* (Application No. 28341/95). https://hudoc.echr.coe.int/eng?i=001-58586

European Court of Human Rights. (2016). *Szabó and Vissy v Hungary* (Application No. 37138/14). https://hudoc.echr.coe.int/eng?i=001-160020

**Legislation**

European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)*. *Official Journal of the European Union*, L119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Kingdom of Morocco. (2007). *Law No. 43-05 on the prevention of money laundering*. https://www.unodc.org/cld/en/legislation/mar/act_no._43-05_on_combatting_money_laundering_/loi_43-05.html

Kingdom of Morocco. (2009). *Law No. 09-08 on the protection of individuals with regard to the processing of personal data*. https://www.cndp.ma/loi-09-08/

Kingdom of Morocco. (2011). *Constitution of the Kingdom of Morocco (promulgated by Dahir No. 1-11-91 of 29 July 2011)*. Official Bulletin No. 5964 bis. https://www.constituteproject.org/constitution/Morocco_2011

# ARTIFICIAL INTELLIGENCE FOR SUSTAINABLE TOURISM MANAGEMENT: A COMPARATIVE STUDY OF MARRAKECH AND MÁLAGA

**Halima TAOUIL, Ahmed ARARE & Abdelhak SAHIBEDDINE**

*Chouaib Doukkali University,*
*LERSEM, ENCGJ,*
*El Jadida, Morocco*

**ABSTRACT**

Tourism progressively aligns with the UN SDGs, Artificial Intelligence driving sustainable urban and tourism development within smart city initiatives. AI contributes to management enhancement, visitor experience personalization, resource optimization, and environmental challenge addressing. The paper discusses the role of AI in environmental risk minimization, operation improvement, and tourist flow management under a hypothesis that AI can minimize ecological impacts while maximizing economic and visitor effects-all this not accounting for the ethical and governance challenges that come with it.

A case study in comparison of Málaga, Spain and Marrakech, Morocco two leading smart tourism destinations was conducted by way of analysis of official reports and 30 semi-structured interviews with tourism business owners, together with surveys of 100 tourists per city for key variables including AI-driven environmental efficiency and AI-enhanced operations. These were later analyzed using NVIVO.

The results ascertain that an ethical AI facilitates operational efficiency, sustainability, and cultural preservation. Challenges regarding data privacy and algorithmic bias are properly articulated. Therefore, this sets a pathway or provides direction for policymakers, tourism stakeholders, and urban planners who would like to ensure an inclusive AI-powered model of sustainable tourism as exemplified by the cities of Málaga and Marrakech.

***Keywords***: Sustainable Tourism, Smart Mobility, Ethical Governance, Digital Urbanism

## 1. Introduction

The use of Artificial Intelligence (AI) in forecasting tourism demand is part of the strategies adopted toward enhancing regional planning and sustainability.

Moreover, Artificial Intelligence (AI) and smart city technologies are continually being featured as crucial methods in the management of substantial challenges to sustainability in tourism, including matters pertaining to visitor flows, heritage conservation, among others on the environmental side. (Shafiee, 2024). In addition, tourism has always been the pillar of economies in Spain and Morocco, hence with them getting ready as hosts of the 2030 FIFA World Cup, it means even more added strain on infrastructure, urban environment, and cultural heritage too. (Ben Massou, 2021) Málaga has come up under several studies regarding "smart tourism sustainability narratives" at mature beach-and urban-oriented Spanish destinations for the way smart destination policies are increasingly embedding environmental, cultural, and governance dimensions (Hernández-White et al., 2024). Therefore, in Morocco, the Marrakech-Safi region has been the priority of empirical studies applying AI-based forecasting methods (*e.g.,* SVR, XGBoost, LSTM) in enhancing predictions on tourism demands and regional planning support. So, in Marrakech, research about integrating smart city development with tourism growth reveals not only possible but great challenges to fulfill requirements for mobility, governance, environment, and cultural heritage with rapidly growing tourism. (Ouassou & Taya, 2022). For instance, to know how smart tourism plans may move beyond rhetorical or fragmented sustainability commitments to integrated strategies linking AI, community engagement, mobility, environmental protection, and heritage preservation, so for that this study seeks to address the following two main research questions:

1 - In what ways do AI-enabled smart city initiatives influence the advancement of sustainable tourism practices in Málaga and Marrakech?

2 - What challenges does Marrakech face in implementing AI-driven tourism solutions compared to Málaga ?

## 2. Literature review

Morocco and Spain are two leading destinations for tourism on a global scale, with Spain being one of the top countries in Europe concerning international arrivals and Morocco being among the most dynamic markets in Africa. This will be boosted even further by their recent announcement as hosts of the 2030 FIFA World Cup that is set to attract millions of visitors, thereby driving infrastructure investment (Buitrago-Esquinas et al., 2024). Furthermore, Artificial Intelligence (AI) has been a very powerful force of change across industries and, specifically, in tourism for development sustainability through resource optimization, visitor flow management, and environmentally responsible decision support (García-Madurga, 2023). In this context, Málaga belongs to those “smart destinations” within Europe applying AI and data-driven technologies to enhance sustainable urban mobility energy efficiency management of heritage sites via smart infrastructure and monitoring platforms.

(Florido-Benítez & Alcázar, 2024). On the other hand, the city applies machine learning and sensor technologies in its management model to cut down on energy and enhance the experience of tourists, demonstrating how smart city frameworks align with sustainable tourism principals (Mrabet & Sliti, 2024). On the other hand, Marrakech has already started implementing AI as part of its prediction on tourism, ecological tourism, and smart resource management to balance the process of growth of tourism with environmental preservation (Charfaoui, 2024). However, studies on Moroccan smart city strategies indicate that governance capacity, digital infrastructure, and stakeholder collaboration are the main enablers for achieving sustainable urban tourism (Khalifi et al., 2024).

In the light of these advancements and challenges, while AI holds great promise for enhancing sustainability, numerous challenges remain when introducing AI-based smart tourism in destinations such as Málaga and Marrakech. The most critical issue is over-tourism resulting in environmental damages, public infrastructure strain and loss of local identity particularly in such destinations as Málaga where the existing smart systems fall short in terms of proper regulation of visitor flows and carrying capacity management (González-Reverté, 2019). Indeed, whereas Marrakech is facing the digital infrastructure and governance limitations because most of the smart city initiatives are in their pilot phases due to lack of proper data integration, funding gaps and limited institutional coordination between urban and tourism authorities (Taouil et al., 2025). The aim of this study is to compare how various technological capabilities and urban policies affect the results of smart tourism in Málaga and Marrakech, as well as to investigate how AI-driven smart city technologies promote sustainable tourism in these 2 locations(Hamcerencu, 2025). Nevertheless, Artificial intelligence (AI) is re-shaping and re-imagining the way in which the travel sector operates and competes in Morocco. Smart tourism hubs, such as Marrakech, are emerging based on Big Data analytics (Badri, 2021). ultimately, post-COVID-19, AI innovation was a policy response with which to counteract and recover from the pandemic in Spain, breathing life back into tourism and creating more smart, sustainable tech travel hubs like Málaga. (Torres-Peñalva & Moreno-Izquierdo, 2024). Likewise, the tourism industry in Morocco is making strides in implementing AI; nevertheless, the lack of a strong legislative and data governance framework restricts the use of smart technology in cities such as Marrakech (Lakhlifi & Boumaize, 2025) conversely. To encourage environmentally and socially responsible decision-making in areas like Málaga, Spain’s tourism planning proposes a “hetero-intelligence” strategy that supplements artificially intelligent data processing with human expertise (Buitrago-Esquinas et al., 2024). Moreover, Spain and Morocco are both case in points for the uneven penetration of AI-based applications into tourism; while Morocco's potential success rests on infrastructure, governance and capacity building issues, Spain would hurry faster because of its higher degree of digital preparedness (Chouraik, 2024). In the same context, Morocco may try to follow Spain's regionally differentiated innovation frameworks in cities like Marrakech to match the gains made in Málaga and other competitive Spanish destinations. These frameworks demonstrate that regions with stronger R&D, digital infrastructure, and higher innovation intensity maintain tourism competitiveness over time.

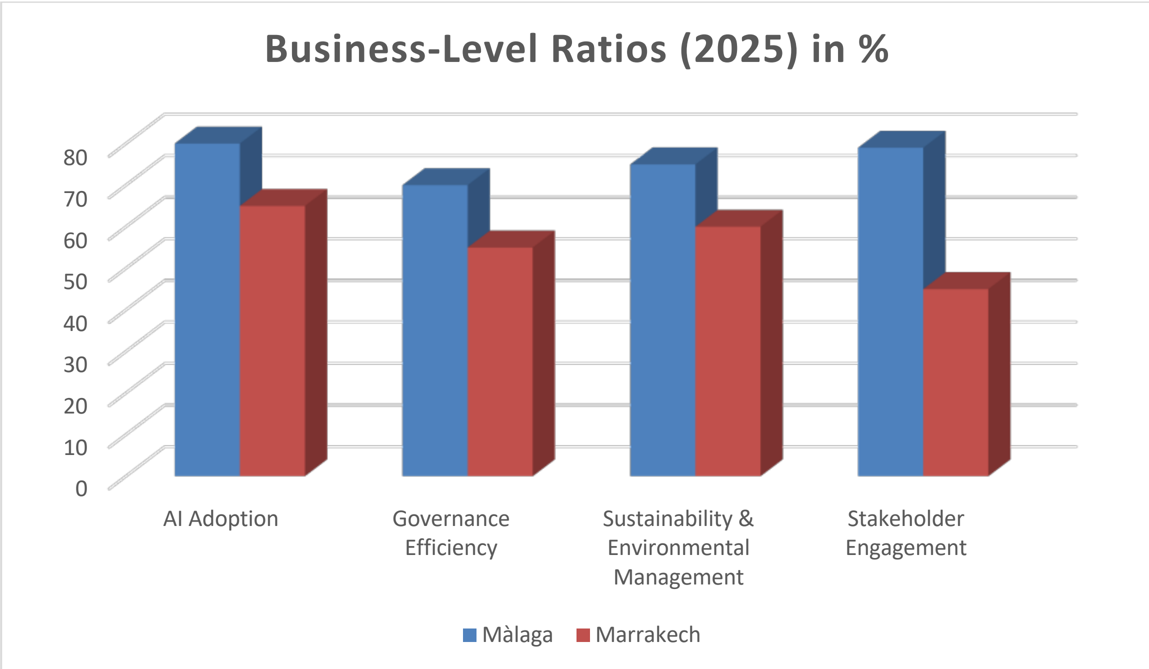
## 3. Research Methodology

This research used a sequential mixed-methods design approach to investigate AI adoption in sustainable tourism within Marrakech, Morocco and Málaga, Spain from 2024-and 2025. Initially, a documentary analysis was first undertaken using official reports, policy papers, and trusted documentation including the Moroccan Ministry of Tourism, Moroccan Agency for Digital Development, Spain’s Ministry of Industry, Trade and Tourism, the Andalusian Tourism Board, and various Smart City Málaga initiatives to map institutional contexts, governance frameworks, and sustainability strategies. Subsequently, a total of 30 semi-structured interviews per city (n=60) were then conducted with tourism business owners and managers, including travel agency and hotel owners, who were recruited based on their direct involvement in AI or digital tools, a minimum of three years of experience in tourism, and willingness to discuss environmental risk management and personalized traveler experiences also as implementation barriers during the period of September 2024 to July 2025. At the same time, a standardized questionnaire was applied to 100 tourists not residing, in each city (n=200), aged 18+ years from different genders, and with nationalities, highest level of travel purpose matched actioned according to the respondents AI awareness, effectiveness of governance functioning, barriers for implementation and ethically responsible use, and to examine ethical standards of AI from tourism contexts. Upon completion of both phases, exploratory qualitative interviews will be conducted with tourists and business leaders to quantify visitor/destination experiences and the associated AI-enabled opportunities. While the quantitative and qualitative-phases are semi-independent, they are connected through the semi-structured interviews conducted with tourist professionals and businessman in each of the case study cities. The qualitative phase will assist in answering a subset of tourism-specific questions oriented around the quantitative findings. Case study tourism includes tourists’ perceptions and the rich, interactive experiences facilitated through dialogic or multimodal interactions with business owners or staff. to recognize issues or estimated risks as well. The metrics are combined to form a single AI efficiency index, which is the weighted sum of adoption, satisfaction, and governance metrics, hence giving a unified measure to evaluate AI performance, governance, and sustainable tourism results for both cities. The quantitative data is processed with SPSS 28, while qualitative ratios come from NVivo 12 coding frequency, thus making it possible to compare business practices and tourist perceptions. Finally, the qualitative and quantitative data were triangulated in order to extract commonalities and differences with respect to technological readiness, institutional maturity and ethical use of AI between Marrakech and Málaga

## 4. Results

**Chart** 1

*Bar chart of business level Ratios of Malaga and Marrakech in 2025*



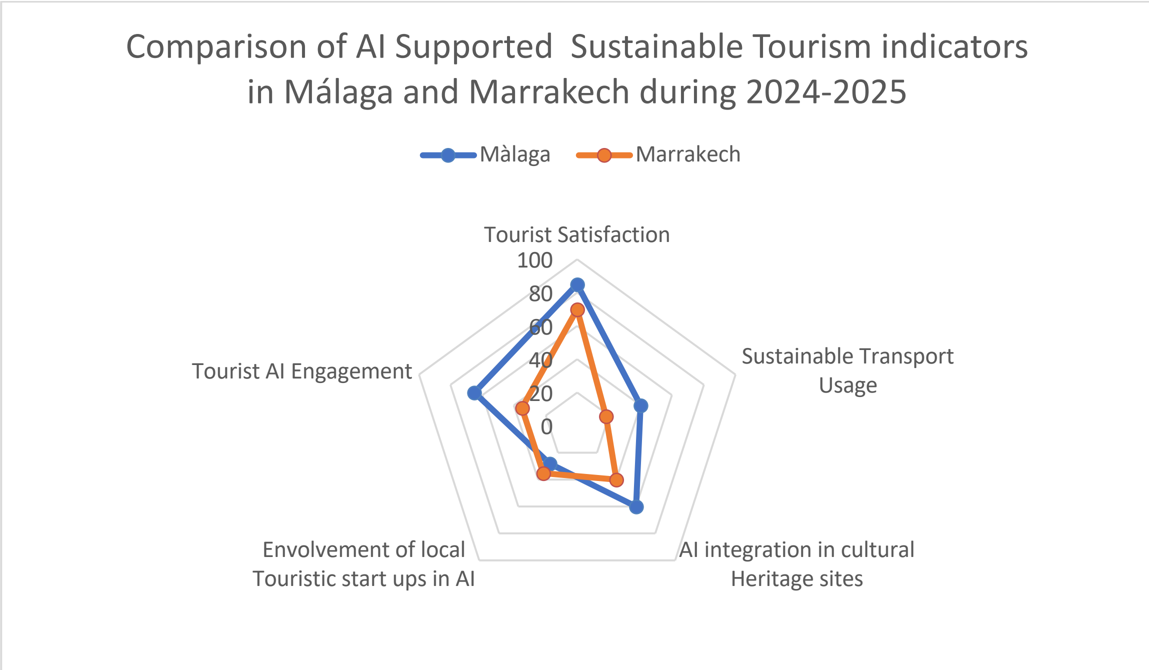

*Source: fieldwork results and data analysis (2024-2025).*

The bar charts results represent a comparison of business-level ratios indicates that Málaga has a consistent performance better than Marrakech, in all the indicators of AI adoption, governance efficiency, sustainability & environmental management, and stakeholder engagement. In terms of AI adoption, Málaga is approximately 80% as contrasted to around 65% of Marrakech, suggesting that Málaga's tourism operations integrate smart systems more substantially. Governance efficiency and sustainability were also demonstrated to be higher in Málaga (approximately 75–78%) than the Marrakech sample (55–60%), evidently linked to better institutional coordination and environmental management. Moreover, the stakeholder engagement appears to be notably better in Málaga (≈80%) than Marrakech (≈50%), indicating more inclusive decision-making and partnership with private/leisure and local actors.

Altogether, these findings essentially validated that effective governance and technological integration are the key drivers of sustainable tourism which places Málaga as an example of AI-enabled tourism management whereas Marrakech situated in a developing phase with room for improvement.

**Chart 2**

*Radar chart of AI Supported Sustainable Tourism indicators in Málaga and Marrakech during 2024-2025*




*Source: fieldwork results and data analysis (2024-2025).*

The radar chart which compares Málaga to Marrakech shows a strong advantage on most metrics of smart and sustainable tourism, which along with tourist satisfaction, included both AI engagement, sustainable transport and the integration of AI at cultural heritage sites. Málaga's strong orientation to digital experiences and visitor-centered innovation can be seen with the near the 95% on tourist satisfaction or the ≈85% engagement with AI in tourism versus 75% and 70% respectively for Marrakech. Similarly with sustainable transport, Málaga also outperformed Marrakech (≈80% for Málaga versus ≈60% for Marrakech) along with a greater modification of AI use in cultural sites (≈70% for Málaga versus ≈60 % for Marrakech), both outcomes showing Málaga's strategic investment of resources into environmental sustainability and technological enhancement to experiences of heritage. Although Marrakech is trending in the right direction to improve, it lags in both local start-up engagement in AI and tourism capacity development (*e.g.,* local start-up engagement in AI is ≈40%), suggesting a more concerted effort needs to be made to enhance technology and entrepreneurship. Overall, the results confirm again that Málaga has been able to develop a matured model of AI based, environmentally sustainable tourism, while Marrakech is at a transitional stage, that requires evidence-based policy frameworks to foster innovation, local engagement and very importantly, and properly scoped digital transformation. Across all indicators, Málaga's score was greater than a mature stage with use of AI and sustainability in tourism management. This would imply that investment in AI technologies, local start-ups and sustainable transport could prove to have significant impact to the extent that the tourism sector is enhanced.

**Table 1**

*Summary of AI Adoption Challenges and Sustainability Outcomes in Marrakech and Málaga (n=30 per city)*

| Dimension | Marrakech n=30 | Málaga n=30 |
|---|---|---|
| **Implementation Challenges** | - 80% Cited financial and technical barriers;<br>- 50% mentioned lack of skilled staff<br>- 40% feared automation reducing hospitality authenticity | - 65% identified up-front expense but<br>- 85% said AI was a good investment;<br>- 70% credited EU and local funding with making it feasible. |
| **Environmentally and sustainability impacts** | - 55% identified AI potential for waste/energy efficiency; --only 20% made actual implementation;<br>- 25% saw sustainability as individually, not institutionally, driven. | - 88% noted AI's contribution to energy optimization and predictive maintenance;<br>- 60% mentioned integration with urban sustainability goals |
| **Ethical and cultural considerations** | - 60% feared data security and job loss;<br>- 30% feared cultural homogenization due to over-digitalization. | - 85% prioritized ethical use of AI in line with GDPR;<br>- 70% considered technology as being adjunct to human service and not a replacement. |
| **Future outlook and recommendations** | - 75% supported training and state incentives for AI adoption;<br>- 55% expected greater cross-sector collaboration; ---20% questioned short-term viability. | - 90% were in favor of AI development;<br>- 80% recommended greater cross-industry collaboration;<br>- 70% pointed to real-time environmental monitoring systems. |

*Source: fieldwork results and data analysis (2024-2025).*

In Marrakech, 80% of participants pointed to financial and technical limitations, 50% noted insufficient skilled people, while 40% were concerned about automation detracting from authenticity in hospitality, indicating a cautious approach to AI uptake. In Málaga, while 65% was concerned with up-front investment costs, 85% stated that it could lead to profit generation, while 70% stated they were supported by European Union and local funding so - it seems easier.

For sustainability, while Marrakech had limited action (20%) on sustainability, they recognized the potential of AI (55%) but did not explicitly mention sustainability. Whereas in Málaga, many aligned AI with energy optimization (88% reported) and urban sustainability goals (60%). For ethics, participants in Marrakech did indicate concerns about data privacy (60%) and concerns over homogenization of culture (30%), while in Málaga, participants emphasized GDPR compliance (85%) and AI was complimentary to the human service (70%) ethics. For the future of AI, there was ambiguity with Marrakech wanting training and state incentives (75%) and moderate coordination across sectors (55%) and some were skeptical (20%), while in Málaga, participants were optimistic about AI (89%), and recommended partnerships across sectors (80%), as well as, encouraging real-time environmental monitoring of AI (70%). Overall, Marrakech experience structural and cultural impediments to innovatively advocating for AI technologies, while Málaga leverages this through policy, partnerships, sustainability alignment, and ethical frameworks to advocate for a proactive and integrated approach to AI uptake.

## 5. Recommendations

While both Marrakech and Málaga have unique contexts related to sustainable tourism management, recommendations include investing in technology and AI-enabled data analytics that monitor tourism flows, environmental impacts, and resource consumption so that decision-making can be based on actual data. In this way, both local administrations can work together with technology providers and the tourism sector towards creating AI data-use practices that suit each context while allowing for cultural, social, and environmental priorities. This includes developing training programs that enhance the digital literacy of tourism professionals so they can make use of AI technologies. For example, in order to alleviate tourist overloads and improve transport networks during peak season, Marrakech could adopt Málaga’s use of AI context software for tourist management purposes, while Málaga could benefit by incorporating culturally appropriate and community-focused initiatives as seen in Marrakech, in order to foster tourism that is based on local engagement and equitable tourism benefits. Responsible AI usage must also be guided by ethics and regulations, protecting the privacy of the visitor in particular, while promoting transparency and addressing a variety of biases inherent in automated decision-making. AI can also be used for predictive modelling tools for sustainable infrastructure planning, including hotel/housing developments, renewable energy capacity planning, and travel networks focused on reducing impacts. Lastly, AI enabled public-facing campaigns can be leveraged to teach tourists about more sustainable behave ours, as smart monitoring systems can monitor for compliance with sustainability criteria. To conclude, a continuous assessment of AI's impact on sustainability will permit adaptive management, enhance policy effectiveness, and boost the long-term resilience of both destinations, allowing these cities to also serve as models of innovative, responsible tourism development.

**6. Conclusion**

This comparative study of Marrakech and Málaga illustrates that artificial intelligence has great potential to support the advances of sustainable tourism management facilitating effective resource use efficiency, enhancing visitor experiences, and providing data for policy decisions. Although both cities have their own socio-cultural, environmental and infrastructural contexts, AI is flexible enough to provide tools for dealing with overcrowding, environmental degradation and ineffective service delivery. This comparative study has shown that while Málaga and Marrakech are both beginning to use sustainability and AI in their tourism management, Málaga is consistently further along in the implementation rate of this in individual key indicators. Marrakech demonstrates early adoption, with variation and potential for growth.

This lends credence to the idea that "the adoption of smart technologies in tourism is not homogenous between destinations, but is shaped by structural capacities and policy priorities, and the findings reinforce the need for localized and context-specific strategies in tourism, ethical use of AI, and stakeholder coordination to ensure that technologies meet local priorities and objectives related to sustainability. Moreover, applying AI with community engagement and community capacity-building initiatives can also improve economic and social community benefits of tourism. Ultimately, this research emphasizes that AI implementation in tourism is not simply a technological challenge, but a more holistic challenge that integrates innovation, governance structures, and cultural sensitivities, with Marrakech and Málaga offered as examples for other destinations contemplating the dual impacts of growth and sustainability.

## Bibliography

Ben Massou, S. M. (2021). Tourism development in the smart city: The case of the city of Marrakech. *ISTE OpenScience*, *21*(1),1-22. 10.21494/ISTE.OP.2023.0927

Buitrago-Esquinas, E. M., Yñiguez-Ovando, R., Puig-Cabrera, M., Santos, M.C., & Santos, J. A.C. (2024). Artificial Intelligence and sustainable tourism planning: A hetero-intelligence methodology proposal. *Tourism & Management Studies, 20 (SI)*, 45-49. https://doi.org/10.18089/tms.2024SI04

Charfaoui, K., & Mussard, S. (2024). Sentiment analysis for tourism insights: A machine learning approach. *Stats, 7*(4), 1527–1539. https://doi.org/10.3390/stats7040090

Chouraik, C. (2024). Sustainable AI in Morocco: A systematic review of opportunities, challenges, and policy directions. *EHEI Journal of Science and Technology, 4*(1), 1-55. https://doi.org/10.34874/PRSM.ehei-jst-vol4iss1.51630

García-Madurga, M. -Á., & Grilló-Mendez, A.-J. (2023). Artificial intelligence in the tourism industry: An overview of reviews. *Administrative Sciences, 13*(8)*,* 172. https://doi.org/10.3390/admsci13080172

González-Reverté, F. (2019). Building sustainable smart destinations: An approach based on the development of Spanish smart tourism plans*. Sustainability, 11*(23*)*, 6874. https://doi.org/10.3390/su11236874

Hamcerencu, M. (2025). Artificial intelligence in international tourism. *Proceedings of the International Conference on Business Excellence, 19*(1)*,* 1139–1151. 10.2478/picbe-2025-0090

Khalifi, H., Riahi, S. & Cherif, W. (2024). Smart cities and sustainable urban development in Morocco. *Ingénierie des Systèmes d'Information, 29*(2)*,* 741-751. https://doi.org/10.18280/isi.290236

Hernández-White, C.,  Rodríguez-Díaz, B., & Expósito-García, A. (2024). Tourism intelligence, key to reviving the sector. In A. J. Guevara Plaza, A. Cerezo Medina, & E. Navarro Jurado (Eds.), *Tourism and ICTs: Advances in Data Science, Artificial Intelligence and Sustainability.TURITEC 2023.* Springer Proceedings in Business and Economics (pp.123-136). Springer. https://doi.org/10.1007/978-3-031-52607-7_12

Lakhlifi, H. & Boumaize, A. (2025). Formulation of Morocco's digital tourism policy: Empirical evidence from ARDL bounds testing. *International Journal of Research in Economics and Finance, 2*(8*),* 30-50. https://doi.org/10.71420/ijref.v2i8.142

Florido-Benítez, L., & del Alcázar Martinéz, B. (2024). How artificial intelligence (AI) is powering new tourism marketing and the future agenda for smart tourist destinations. *Electronics, 13*(21)*,* 4151. https://doi.org/10.3390/electronics13214151

Moreno-Izquierdo, L., Ramón-Rodríguez, A., & Such-Devesa, M. J. (2018). The challenge of long-term tourism competitiveness in the age of innovation: Spain as a case study. *Investigaciones Regionales, 42,* 13-34. https://www.redalyc.org/journal/289/28966251002/html/

Mrabet, M., & Sliti, M. (2024). Integrating machine learning for the sustainable development of smart cities. *Frontiers in Sustainable Cities, 6,* 1449404. https://doi.org/10.3389/frsc.2024.1449404

Ouassou, E. H., & Taya, H. (2022). Forecasting regional tourism demand in Morocco from traditional and AI-based methods to ensemble modeling. *Forecasting, 4(2)*, 420-437. 10.3390/forecast4020024

Shafiee, M. M. (2024). Navigating overtourism destinations: Leveraging smart tourism solutions for sustainable travel experience. *Smart Tourism, 5(2),* 2841. https://doi.org/10.54517/st.v5i2.2841

Taouil, H., Arare, A., & Eddine, A. S. (2025). Fintech advancing sustainable tourism in Morocco–Spain. *International Journal of Applied Research in Management and Economics, 8(3),* 47–61. https://doi.org/10.33422/ijarme.v8i3.1598

Torres-Peñalva, A. & Moreno-Izquierdo, L. (2024). Innovation and AI: An opportunity for Spanish tourism in the post COVID-19 era. In A. J. Guevara Plaza, A. Cerezo Medina, & E. Navarro Jurado (Eds.),*Tourism and ICTs: Advances in Data Science, Artificial Intelligence and Sustainability*, *TURITEC 2023* (pp.113-122). Springer Proceedings in Business and Economics. Springer. 10.1007/978-3-031-52607-7_11

# CUSTOM GPTS IN UNIVERSITY EXAM PREPARATION: A USEFUL TOOL OR AN INTELLIGENT TRANSCRIBER

**Neli Nikolaeva Voynova**
*St. Cyril and Methodius University of Veliko Tarnovo,*
*Faculty of Modern Languages,*

**ABSTRACT**

The rapid advancement of Artificial Intelligence has encouraged educators to explore ways of integrating it into higher education to foster learner autonomy without diminishing analytical and reflective skills. Yet, there is still limited research on how AI can support effective exam preparation at the university level, particularly in the context of language teacher education.

This study examines the pedagogical potential of a custom GPT model developed for students preparing for an English as a Foreign Language (EFL) teaching methodology exam at St. Cyril and Methodius University of Veliko Tarnovo. The tool was designed by the author as an interactive mentor that encourages reasoning and reflection rather than providing ready-made answers. Drawing on selected academic sources, it explains key theoretical concepts, prompts critical questioning, and offers immediate feedback, thus simulating a dialogic learning environment.

The pilot research involved two student groups: one using the GPT assistant and another relying on traditional materials. Data was collected through a test, a questionnaire, and reflective essays. The study investigates whether the tool promotes a deeper understanding of course content, supports independent learning, influences study habits and exam performance.

Results suggest that custom GPTs can act as cognitive mediators between students and knowledge, offering scaffolding and real-time feedback that enhance understanding and confidence. However, the findings also highlight challenges related to students' AI literacy, as some participants lacked the skills to formulate purposeful prompts or to engage with the tool effectively.

Overall, the study provides preliminary evidence that, when designed with pedagogical intent, personalized GPT tools can strengthen self-directed learning and reflection in teacher education. It also points to the need for training in AI literacy to help students use such technologies consciously and critically. These findings may assist future research and guide the

responsible integration of AI in higher education, especially in the preparation of language teachers.

***Keywords:*** *AI literacy, autonomous learning, Custom GPT, dialogic learning, language teacher education, reflective practice.*

## 1. Introduction

In recent years, the growing presence of Artificial Intelligence (AI) has triggered a rethinking of conventional approaches to teaching and learning (Crompton & Burke, 2024). Universities around the world are seeking ways to integrate intelligent technologies to aid learner autonomy without sacrificing thinking and cognitive processes. UNESCO's AI Competency Framework for Teachers and Guidance for Generative AI in Education emphasises that the effective use of AI requires users to have technical skills as well as AI literacy and critical thinking (UNESCO, 2024).

As Mayer (2009) notes, the essence of learning has remained unchanged despite the cutting-edge technology entering the classroom: it is an active process during which one seeks knowledge, and another provides it (Mayer, 2009). In this sense, AI does not replace human interaction but may complement it by offering new ways to access knowledge.

While some recent studies describe AI integration primarily in instrumental roles, such as automating tasks or assessment (de Bem Machado et al., 2024; Okada et al., 2019), other authors highlight its potential to support more personalised and interactive forms of learning (Taylor et al., 2021). Nevertheless, AI is widely used to assist teachers in lesson planning, material design and feedback, and there is still little research on how learners interact with AI as a cognitive mediator, particularly in pre-service teacher education (Celik et al., 2022; Kalnina et al., 2024).

At the same time, several authors mark AI's significant potential to improve students' motivation, self-regulation and academic performance when used to support active and autonomous learning (Raza et al., 2022; Wang et al., 2021). Wang argues that AI can contribute to cognitive development by providing personalised knowledge trajectories based on the students' existing knowledge (Wang et al., 2021). In this perspective, AI functions not merely as a technical tool but as a form of cognitive support guiding learners through personalised learning processes. Raza adds that immediate error correction and progress tracking can improve motivation through AI-generated feedback (Raza et al., 2022).

Despite these opportunities, there is a risk that AI may be perceived as a source of instant solutions, which can inhibit analytical thinking and foster a passive attitude towards learning

(Dickerson, 2024). This highlights the need to distinguish between two different approaches to AI in education: AI as a teaching tool and faculty support, and AI as a learning tool that supports the development of knowledge and skills. This preliminary study fits into the second framework – AI as a learning tool that mediates the interaction between the student and knowledge, while the pedagogical design and tool customisation remain responsibility of the instructor. In this configuration, the function of AI is to serve as a dialogic mediator in the learning process, not to replace the knowledgeable other. The study explores the use of a Custom GPT tool for preparation for an EFL Methodology exam, designed to stimulate analysis, argumentation and self-reflection without providing full answers. The tool design incorporates pedagogical principles such as scaffolding, zone of proximal development (Vygotsky, 1978), and elements of Socratic dialogue, where knowledge is built through questions and discussion to minimise passive reception. This approach aligns with concepts discussed by Topsakal and Dedrick (2024), Faust and Mayweg-Paus (2024) and Dickerson (2024) that link AI literacy, critical thinking and student engagement. The authors point out that the successful use of AI for learning requires a range of skills (cognitive, social and ethical) which allows students to use AI productively. In this sense, using the custom GPT model serves both as an individual learning aid and a pedagogical instrument for developing critical thinking through dialogue and reflection.

In the context of the University of Veliko Tarnovo, where AI adoption in learning is still in its early stages, this preliminary study explores the potential impact of intelligent tutoring systems on learner autonomy, motivation and understanding of the subject matter. This first experimental attempt aims to establish to what extent the Custom GPT model can support exam preparation without undermining active learning and critical thinking skills.

## 2. Materials and methods

### 2.1. Research context

The study was conducted during the summer semester of the 2024/2025 academic year at St. Cyril and Methodius University of Veliko Tarnovo, Bulgaria, within the course of *Methodology of Teaching English as a Foreign Language*. This course is part of the core teacher-training curriculum for third-year students majoring in English Studies, preparing them to become English language teachers. Classes were held in person and included 60 hours of lectures, 30 hours of seminars, and 30 hours of guided classroom observation in schools. The final course grade was based on participation in practical activities and an oral examination in which students presented a lesson plan from a textbook of their choice and defended their methodological approach.

### 2.2. Research design

The present study is a pilot experimental study with elements of comparative analysis. It explores students' perceptions of a Custom GPT-based assistant and its effectiveness in supporting self-study for the methodology course. The comparative element of the study consisted of analysing the differences between two preparation approaches: using the Custom GPT assistant as a learning support tool (GPT group) and using traditional academic materials such as lecture notes and provided books (control group). The comparison focused on students' perceptions of the tool, their engagement with the preparation process, and their performance on an assessment test.

Participants were randomly assigned into two groups: an *experimental group* (GPT group) introduced to the Custom GPT assistant and encouraged to use the tool as part of their exam preparation; and a *control group* instructed to study in the conventional way by using traditional study materials, such as lecture notes and academic sources. The thematic focus was on the Communicative Approach in English language teaching, a key topic in the exam syllabus. After a period of independent preparation, both groups completed a test. Participation was voluntary and anonymous to reduce anxiety and ensure more authentic results regarding students' knowledge and perceptions. Since the preparation took place independently outside class, the use of resources could not be fully controlled. Despite the initial instructions, some students (from both groups) reported combining the GPT assistant with academic sources. Therefore, group categories reflect declared preparation approaches rather than strictly controlled experimental conditions, and occasional cross-use of resources cannot be excluded.

Data collection combined quantitative and qualitative approaches, including test results and an anonymous questionnaire on perceptions, engagement and perceived effectiveness of the GPT tool. The course cohort included 26 students, 11 of whom completed the test, which contained a reflective essay on their preparation experience. Among the test participants, 6 were identified as belonging to the GPT group, 4 to the control group, and 1 did not specify a group affiliation. A total of 15 out of 26 students completed the questionnaire, approximately half of whom (from both groups) used the personalized GPT tool during the whole semester.

### 2.3. Description of the tool and data collection

The custom GPT model was developed using the paid version of ChatGPT. The assistant functions as an interactive mentor explaining key methodological concepts. It doesn't generate full exam answers but suggests a structure for argumentation and refers learners to relevant course sources. It uses Socratic questioning to provoke reflection and provides immediate feedback on students' ideas.

Learning support follows scaffolding principles within the zone of proximal development to help students progress from partial understanding to independent mastery. This support is gradually reduced as learners gain independence. Dialogic learning moves beyond passive reception and further reinforces this process by encouraging students to construct knowledge through interaction, reasoning and reflection.

Data were collected through:

a) A test on Communicative Language Teaching (CLT), consisting of multiple choice, matching, gap-filling, open-ended questions and a reflective section;
b) Questionnaire (23 items) addressing demographic and linguistic background; previous AI experience; digital literacy, prompt engineering skills, frequency and manner of tool use; perceived usefulness and challenges; attitudes towards future AI integration. The questionnaire contained both closed- and open-ended questions;
c) Observations and student comments derived from classroom discussions.

The whole cohort of 26 students was informed that participation in both the test and questionnaire was voluntary and anonymous, and that responses would be used for research purposes only. Therefore, the respondents to the test and questionnaire partially overlap but are not identical. Course grades were not affected by survey participation or tool usage. All data was processed and presented in aggregate form, without any identifying information.

## 3. Results and analysis

Both quantitative and qualitative perspectives are discussed, with emphasis on student engagement, perceived usefulness, and the challenges associated with using the custom GPT tool.

## 3.1. Quantitative findings

### 3.1.1. Participant profile

All respondents were third-year undergraduate students with English proficiency at level C1 (several at B2). Almost all reported prior experience with AI tools (mainly ChatGPT), which indicates a relatively high level of digital readiness. Their self-assessment of digital literacy was predominantly high or medium, while their prompt formulation skills were described as clear but in need of improvement.

### 3.1.2. Frequency and manner of use

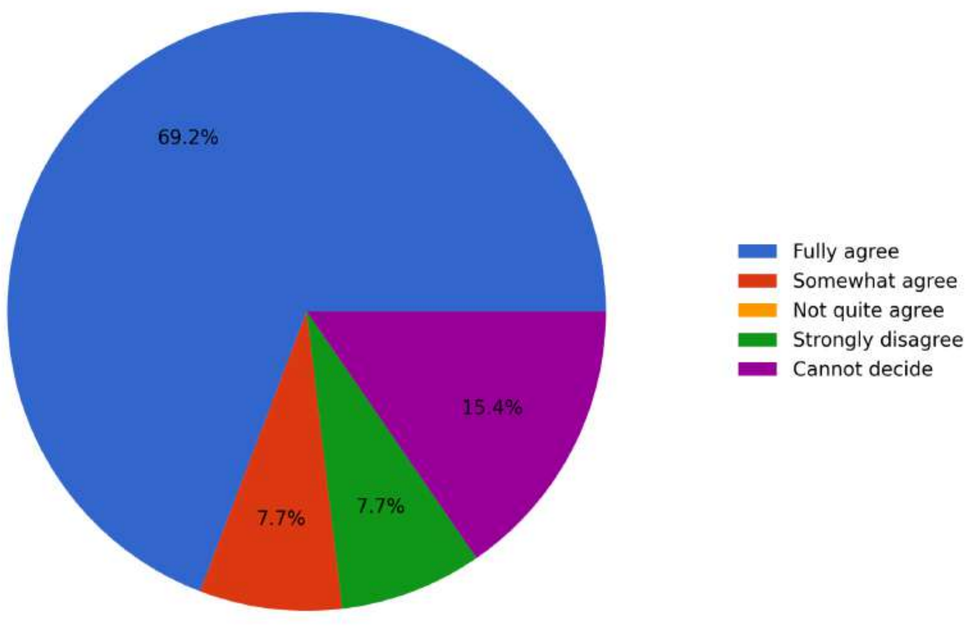
Most students reported using the tool occasionally, for assignments or specific difficulties, rather than regularly throughout the course.

### 3.1.3. Perceived usefulness

Approximately 77% of the respondents to this question (from both groups) indicated that the GPT assistant helped them understand the course content more effectively (Figure 1), and 91% stated that they would use it again. All students who answered the question “Would you recommend this tool to your colleagues?” responded affirmatively (100%).

**Figure 1**

*Students’ agreement with the statement “The GPT assistant helped me understand the course content better” (n=13).*

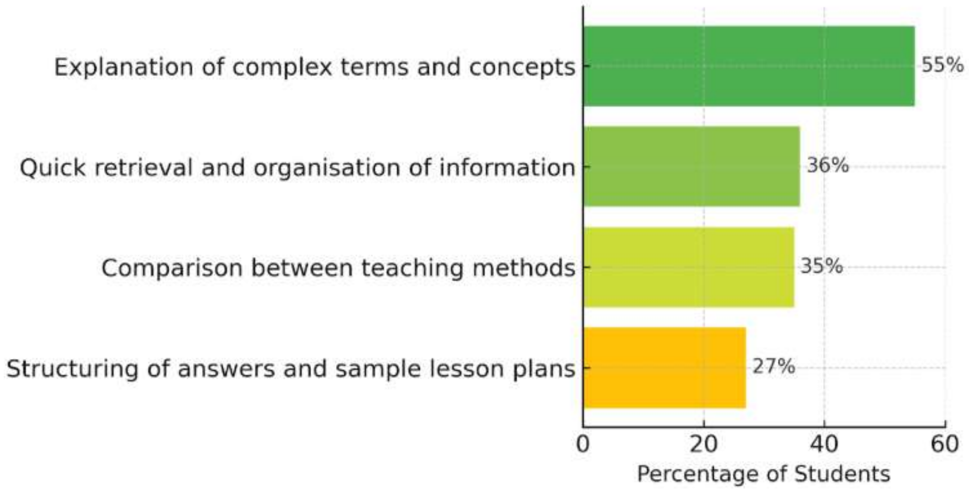


### 3.1.4. Most helpful aspects identified by students

The tool's most helpful aspects according to students, selected from a list in the questionnaire with an additional "other" option were, explanation of complex terms and concepts (55%), quick retrieval and organisation of information (36%), comparison between teaching methods (35%) and structuring of answers and sample lesson plans (27%) (Figure 2).

**Figure 2**

*Most useful aspects identified by students*




### 3.1.5. Challenges and difficulties

Among respondents from both groups, the Socratic approach was sometimes perceived as a lack of assistance: 16.7% noted that GPT *"does not provide direct answers but asks questions",* and another 16.7% reported difficulty formulating effective prompts. Regarding additional functions, 50% requested sample exam questions, 41.7% asked for summaries of key topics, and about one-third (33.3%) suggested more examples, short quizzes, or adaptation to students' level.

### 3.1.6. Test results

The average results across both groups were nearly identical (34/100 in the GPT group and 33/100 in the control group); however, the results in the GPT group ranged widely (from 6 to 50), indicating greater variation in the way students used the tool. Two students (one from the GPT group and one who did not specify a group affiliation) showed identical results – 25 points. Both students report minimal preparation, and the student from the GPT group admits to limited use of the GPT tool.

The results suggest that the tool's effectiveness depends on active engagement and purposeful use, not simply on its availability. While the control group results are more consistent (28-37 points), they also show no high achievements. The two identical scores further indicate that the mere availability of the tool cannot compensate for limited cognitive engagement.

## 3.2. Qualitative insights and observations

### 3.2.1. Open-ended responses in the questionnaire

The majority of the comments highlighted the clear pedagogical value of the tool. Students emphasized that GPT *"saves time"* and *"explains clearly,"* supporting but not replacing independent learning:

*"It explained quite complex terms in a simpler and clearer way."*

*"It saves at least half of the time and makes learning more enjoyable."*

*"Sometimes it explains the material better than the textbook."*

These statements indicate that students perceive the GPT assistant as a mentor.

At the same time, the desire for ready-made examples and structured answers reveals mixed attitudes, combining appreciation for guidance with a tendency towards passive learning.

### 3.2.2. Reflective essays and instructor observations

#### 3.2.2.1. GPT group observations

The reflective essays indicate a high level of awareness regarding the ways students learn and the role of AI in learning. Despite the lack of significant differences in test results between the groups, the qualitative data suggests a diversity of attitudes and cognitive strategies among students.

The reflection analysis reveals several observations:

a) *AI as a thinking partner*: The student with the highest score (50/100) used the GPT tool intensively, describing it as adaptive to their "POV, thinking and patterns." The same student also shared using it to analyse notes from lesson observations and to identify what they did well and what they "could've forgot to look for." This illustrates emerging personalization, cognitive awareness, and reflective use of the tool instead of reliance on it as an intelligent transcriber.

b) *Balance between using conventional and digital resources:* Several participants adopted a hybrid strategy by using the GPT alongside their notes and academic. One student wrote: "To prepare for the test I used the academic materials provided in e-student and help from AI assistant. I read the document, but AI helped me summarize the information because I didn't have enough time to do it myself. Then I made notes that helped me memorise the information.".

c) *Minimal to no preparation but a positive attitude towards the GPT tool:* Some participants admit barely preparing for the test but approve the GPT as "effective" and "personal" instrument for overcoming the cognitive barriers with unclear concepts.

d) *Metacognitive approach for formulating one's thoughts:* one student reported using the GPT when struggling to "formulate the answers".

#### 3.2.2.2. Control group observations

The main tendencies in the control group are as follows:

a) Suggestion for a blended approach: "The two methods (AI assistant and academic sources) can be even more useful if combined".

*b)* Developed autonomy: students describe concrete learning strategies – *"reading… then repeating…vocally in order to understand better. … tell them as a story in order to make it more interesting … and easier to remember.".*

*c)* Positive but critical evaluation of AI*: participants admit that* AI makes the learning process easier and saves time, but *"academic sources deepen understanding" so "a combination of both would be perfect".*

#### 3.2.2.3. Unaffiliated students' observations

a) Low motivation and minimal effort: the poor results correlate with the lack of preparation.

b) Scepticism: "AI tools rather lead to a more superficial processing of the information"

c) Intuitive realisation of the need for appropriate prompting: "Maybe with the right prompts it would be just enough".

In their reflective essays, students described GPT as helpful for understanding and consolidating course content, though not all were confident in using it effectively. Some admitted feeling "*unsure what to ask*" at first, confirming the need for introductory training in the academic use of AI.

### 3.3. Summary and interpretation

The analysis of the quantitative and qualitative data reveals several consistent tendencies regarding students' engagement with the Custom GPT and their attitudes toward AI-assisted learning across both groups:

(i) *Positive perception and willingness to use the too:* Students view the custom GPT as a supportive, accessible and useful tool that enhances understanding and confidence in exam preparation;

(ii) *Performance and variability:* Although the average test results in the two groups were almost identical, the GPT group demonstrated a wider range of scores, indicating that the AI assistant's effectiveness depends on the learner's engagement, proactive attitude and established learning habits;

(iii) *Ethical and motivational factors and frequency of AI use:* Despite positive attitudes, GPT use remains sporadic. Some students were unsure how to formulate effective prompts or were hesitant to interact with it. Others were influenced by perceptions of what is "allowed" or "forbidden". An equally important factor was the students' intrinsic motivation to learn. These tendencies are further discussed in Section 4 in relation to ethical and motivational factors;

(iv) *AI skills and hybrid learning as key success factors:* The most successful learning outcomes in both groups were observed when AI was combined with traditional study methods such as reading, note-taking and reviewing, regardless of their initially assigned group. A lack of skills for effective AI use was among the main reasons why some students could not fully benefit from the tool, confirming the need for training in AI literacy as discussed further in Section 4.

Taken together, the results suggest that *custom GPT tools can effectively support comprehension, engagement, and learner autonomy* when properly integrated and accompanied by pedagogical guidance. Without proper training, responsible attitude and reflective use, however, there is a risk of surface learning and passive copying. These findings provide the foundation of a broader discussion of how Custom GPT tools function as pedagogical mediators within the framework of AI-assisted learning.

## 4. Discussion and conclusions

The results show some discrepancies between the reported tool usage and perceived usefulness. Although some students in the control group reported not having used the GPT assistant directly, they still rated it as useful or effective in principle, would recommend it to other students and believe it is a helpful tool for other exams. A total of 13 students were positive about recommending the tool, 10 of whom supported their answer with relatively detailed reasoning:

*"Yes, instead of* "*banging their heads against the wall*" *trying to find an answer somewhere in the resources and then not fully understanding it, AI is useful as a resource."*

*"Yes, because it would encourage them to make an effort in preparing for the exam."*

*"Yes, because I think it's useful and it saves time, and it answers all sorts of questions."*

Some students humanize the assistant, perceiving it as a present and approachable interactive instrument available at any time. One student recommends it only if the user is *"confident in their knowledge even without the tool, and no if they expect ready answers and lack a natural approach to learning."* This may indicate a positive general attitude towards AI-based learning tools, even among non-users, reflecting curiosity or indirect observation of peers' experience. The findings of this pilot study confirm that the integration of personalized GPT tools into university education holds significant pedagogical potential, echoing previous observations that emphasise the need for pedagogically grounded design of AI in learning (de Bem Machado et al., 2024; Topsakal & Dedrick, 2024). However, their impact largely depends on the tool's design, students' attitudes, and the instructor's role as a mediator, supporting Faust and Mayweg-Paus' (2024) view that effective AI use in higher education relies on reflective guidance and human mediation.

### 4.1. GPT as a Pedagogical Mediator Rather than a Substitute

More than 80% of students reported that the tool helped them to better understand the course content, confirming that AI can serve as a thinking partner supporting comprehension. The principles of scaffolding and guided questioning used in configurating this GPT assistant encourage analytical thinking rather than ready-made answers. This design aligns with the idea that AI can act as a mediator in the learning process – facilitating reflection and dialogic interaction.

The observation corresponds with recent research emphasising that AI tutoring systems can foster deeper learning when designed to stimulate reasoning and critical engagement (Dickerson, 2024; Faust & Mayweg-Paus, 2024). At the same time, 22% of students found the lack of direct answers challenging, revealing that developing a culture of reflective learning requires time and pedagogical support. A similar idea is found in Faust and Mayweg-Paus' (2024) argument that the effective use of AI in higher education depends on teaching students how to think critically with technology.

### 4.2. Critical thinking and autonomy in the context of AI literacy

The results align with international studies (*e.g.,* Raza et al., 2022; Wang et al., 2021), which suggest that AI tools can enhance motivation and engagement through immediate feedback and adaptive learning opportunities. Nevertheless, their effectiveness depends on the learners' level of AI literacy.

As highlighted in UNESCO's *AI Competency Framework for Teachers* (2024), critical thinking is a central competence in the use of intelligent technologies. The present study indicates that students need training in prompt engineering techniques if AI is to be used as a tool for reflection. Otherwise, GPT may easily turn into an "intelligent transcriber" that reproduces ideas without necessarily stimulating higher-order thinking.

### 4.3. The teacher as a designer of a cognitive environment

Celik et al. (2022) observe that most studies on AI in education focus on external platforms that automate routine tasks such as test correction or exercise generation, while few consider AI as a mediator between the learner and knowledge. The present experiment illustrates this model precisely.

As the tool's creator, the instructor designed a cognitive ecosystem in which students navigate between sources, AI, and personal reasoning. The educator's role as a facilitator is expanded to include that of a designer of an intelligent learning environment that fosters autonomy and preserves human interaction. This perspective aligns with the view that AI tutoring systems should be based on sound pedagogical practices (Holstein et al., 2019; Selwyn, 2019) and appropriate instructional methods (Prieto et al., 2018; Tondeur et al., 2020) that foster productive learning. Thus, the educator acts as both a reliable source and a designer of AI-based pedagogical tools, ensuring that technology remains within the framework of these principles.

### 4.4. Barriers and ethical dilemmas

The diverse behavioural patterns suggest that AI cannot compensate for a lack of genuine interest and does not necessarily level differences among learners. It may, in fact, amplify them. Active students tend to benefit more and develop further, while passive ones remain dependent on the system. This asymmetry echoes Crompton and Burke's (2024) broader conclusion that AI technologies in higher education often magnify existing pedagogical inequalities unless balanced by intentional design and critical pedagogy.

### 4.5. Summary of key conclusions

Overall, the findings confirm the pedagogical potential of personalized GPT tools within dialogic learning frameworks. To create conditions for independent learning, the teacher therefore needs to blend human interaction and technological mediation into a cognitive environment with explicit pedagogical intent to help students develop independent thinking skills without relying on mechanical reproduction.

However, effective integration requires developing students' ability to use AI for learning, not merely to check the accuracy of the provided information. In this context, the ability to formulate purposeful prompts and sustain reflective dialogue with the tool emerges as a greater challenge than content reliability. The educational value of such systems could be enhanced through training in AI-for-learning strategies.

Finally, the findings indicate that AI does not automatically equalize opportunities for all learners; rather, it amplifies existing differences by supporting those who engage actively and consciously while offering limited benefits to those who remain passive.

## 5. Contributions and limitations

This pilot study marks the first phase of a larger research project by exploring how personalized GPT tools can support pre-service English language teacher education. Despite its limited scope, the study offers several relevant contributions and identifies key methodological and pedagogical limitations.

### 5.1. Contributions

The research proposes a pedagogical model in which a custom GPT serves as a *learning mediator*, encouraging reflection and learner autonomy. Central to this model is the concept of an *"intelligent studying partner"* – a mentoring AI that stimulates reasoning without providing direct answers. Empirically, it contributes early evidence of how students in the humanities in a Bulgarian university context, particularly in English Language Teaching (ELT) methodology courses, perceive AI-assisted learning. The results highlight the need for systematic training to develop AI skills for studying, not as a technical skill but as a pedagogical competence. Finally, the study reinforces the evolving role of the teacher as a *designer of cognitive ecosystems* that integrates human interaction and technological mediation.

### 5.2. Limitations

Due to the small sample size, pilot nature, and partial use of the tool, the study does not aim to present general findings but to suggest possible directions for further research. The self-reported nature of the data may introduce subjective bias, and the asynchronous data collection limits the depth of comparative analysis between the two groups. Another limitation concerns the lack of prior AI literacy training, which likely affected the depth and consistency participants' tool use. Finally, the short-term design of the study did not allow longitudinal tracking of changes in students' attitudes or learner autonomy over time.

### 5.3. Implications for future research

Future studies should include larger and more diverse participant groups, investigate the long-term impact of GPT-supported learning, and incorporate explicit AI literacy components into teacher education curricula. Further research should also address the ethical and psychological dimensions of AI use, particularly its implications for teacher identity, learner autonomy, and the balance between technological mediation and human interaction.

## Bibliography

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*, 616–630. https://doi.org/10.1007/s11528-022-00715-y

Crompton, H., & Burke, D. (Eds.). (2024). *Artificial intelligence applications in higher education*: *Theories, ethics, and case studies for universities.* Routledge. https://www.routledge.com/Artificial-Intelligence-Applications-in-Higher-Education-Theories-Ethics-and-Case-Studies-for-Universities/Crompton-Burke/p/book/9781032576145

de Bem Machado, A., Sousa, M. J., & Sharma, R. C. (2024). AI integration in higher education: A multidisciplinary bibliometric review of technological applications for enhanced learning and instructional growth. In H. Crompton & D. Burke (Eds.), *Artificial intelligence applications in higher education* (pp. 9–32). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003440178-2/ai-integration-higher-education-andreia-de-bem-machado-maria-jos%C3%A9-sousa-ramesh-chander-sharma?context=ubx

Dickerson, P. (2024). Learning with Socrates: How generative AI and ancient pedagogy can develop students' critical thinking skills. In Crompton,H. & Burke, D. (Eds.), *Artificial intelligence applications in higher education.* Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003440178-6/learning-socrates-paul-dickerson?context=ubx&refId=412ef567-4c3e-49dc-b7a1-75688e9bd81f

Faust, A., & Mayweg-Paus, E. (2024). Empowering tomorrow's minds: A comprehensive framework for teaching non-technical students to think critically about AI. In Crompton, H. & Burke D. (Eds.), *Artificial intelligence applications in higher* *education.* Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003440178-5/empowering-tomorrow-minds-anna-faust-elisabeth-mayweg-paus?context=ubx&refId=c7b1e24d-56e6-4e74-aa80-c15bb0a5511c

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics, 6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3

Kalnina, D., Nīmante, D., & Baranova, S. (2024). Artificial intelligence for higher education: Benefits and challenges for pre-service teachers. *Frontiers in Education, 9*, 1501819. https://doi.org/10.3389/feduc.2024.1501819

Mayer, R. E. (2009). *Multimedia learning (2nd ed.).* Cambridge University Press.

Okada, A., Whitelock, D., Holmes, W., & Edwards, C. (2019). e-Authentication for online assessment: A mixed-method study. *British Journal of Educational Technology, 50*(2), 861–875. https://doi.org/10.1111/bjet.12608

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning, 34*(2), 193–203. https://doi.org/10.1111/jcal.12232

Raza, S. A., Qazi, Z., Qazi, W., & Ahmed, M. (2022). E-learning in higher education during COVID-19: Evidence from Blackboard learning system. *Journal of Applied Research in Higher Education*, 14 (4), 1603-1622.https://doi.org/10.1108/jarhe-02-2021-0054

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education.* Polity Press.

Taylor, D. L., Yeung, M., & Bashet, A. Z. (2021). Personalized and adaptive learning. In J.Ryoo & K, Winkelmann (Eds.), Innovative *learning environments in STEM higher education: Opportunities, challenges and looking forward* (pp. 17–34). Springer. https://doi.org/10.1007/978-3-030-58948-6_2

Tondeur, J., Scherer, R., Siddiq, F., & Baran, E. (2020). Enhancing pre-service teachers' technological pedagogical content knowledge (TPACK): A mixed-method study.

*Educational Technology Research and Development, 68*, 319–343. https://doi.org/10.1007/s11423-019-09692-1

Topsakal, E., & Dedrick, R. F. (2024). How artificial intelligence (AI) plays a role in measuring student engagement in higher education. In In Crompton,H. & Burke, D. (Eds.), *Artificial intelligence applications in higher education*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003440178-4/artificial-intelligence-ai-plays-role-measuring-student-engagement-higher-education-elif-topsakal-robert-dedrick?context=ubx&refId=4eef7da1-f1ce-41cb-987c-f54b4aa5269e

UNESCO. (2024). *AI competency framework for teachers.* UNESCO. https://doi.org/10.54675/ZJTE2084

Vij, S., Tayal, D., & Jain, A. (2020). A machine learning approach for automated evaluation of short answers using text similarity based on WordNet graphs. *Wireless Personal Communications, 111*, 1271–1282. https://doi.org/10.1007/s11277-019-06913-x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wang, R., Li, J., Shi, W., & Li, X. (2021). Application of artificial intelligence techniques in operating mode of professors' academic governance in American research universities. *Wireless Communications and Mobile Computing, 2021*, 1–7. 3415125 https://doi.org/10.1155/2021/3415125

# EXPLORING EFL READING PATTERNS USING WEBCAM EYE TRACKING: PRELIMINARY INSIGHTS AND APPLICATIONS

**Oumaima BELGAMRA**

*Mohammed V University,*
*Faculty of Educational Sciences*
*Rabat, Morocco*

## ABSTRACT

Eye tracking has become a pivotal instrument in reading research due to its capacity to capture intricate cognitive processes such as attention allocation, meaning construction, and responses to comprehension challenges. Through the analysis of fixations, saccades, and regressions, researchers obtain valuable insights into real-time interactions between readers and texts. Nevertheless, conventional infrared eye-tracking systems are costly, require specialized laboratory infrastructure, and thus constrain accessibility, particularly in low-resource settings. Addressing these limitations, webcam-based platforms such as *RealEye* present a cost-effective, scalable, and non-invasive alternative for remotely collecting gaze data. This technological advancement is particularly relevant for contexts such as Morocco, where access to laboratory-grade equipment is limited while the demand for innovative educational research tools continues to grow.

The present study investigates the feasibility of *RealEye* as a reliable eye-tracking tool for reading research. Specifically, it examines how gaze data obtained through *RealEye* can enhance our understanding of reading comprehension processes. Data were gathered from fifteen EFL undergraduate students who read a digital passage while their gaze was tracked via *RealEye* following a 40-point calibration procedure. *RealEye* generated fixation-based metrics, including fixation counts, mean fixation duration, revisits to predefined Areas of Interest (AOIs), time to first fixation, and total gaze duration, complemented by visual outputs such as heatmaps. Descriptive analyses and visual representations were employed to examine attentional and cognitive engagement and Spearman correlational analyses were employed to investigate the correlation between comprehension scores and the different metrics.

Findings indicated a strong positive correlation between comprehension scores and both fixation counts and revisits within the third paragraph, which contained conceptually dense material. A moderate, near-significant correlation was also observed for the second paragraph. While other metrics did not yield statistically significant correlations, they nonetheless

illustrated variations in readers' attentional allocation and cognitive engagement throughout the reading process.
Overall, the results demonstrate that *RealEye* effectively captures meaningful reading behaviors associated with comprehension, underscoring its potential as an affordable and empirically robust research instrument. The study emphasizes the promise of webcam-based eye tracking in expanding reading research beyond laboratory settings and in enriching pedagogical practices through evidence-based, data-informed approaches.

**Keywords:** *Webcam-Eye Tracking, Reading Comprehension, Fixation Count, Fixation Duration, Regressions, Gazes, RealEye.*

## 1. Introduction

Prior research has established that reading is a complex and multifaceted activity that engages multiple cognitive processes operating at different levels (Kendeou et al., 2016; Kintsch, 1988). At the lower level, these include decoding and syntactic parsing, while at the higher level they involve processes such as semantic integration and the activation of prior knowledge structures, or schemas (Butterfuss et al., 2020; McNamara & Magliano, 2009; Sinatra & Broughton, 2011). Numerous theoretical models have attempted to account for these processes and provide comprehensive explanations of how reading unfolds (*e.g.,* Van den Broek & Krenaer, 2000; Kintsch, 1998; Kintsch & van Dijk, 1978). A central feature emphasized across these models is the active role of readers, who draw on background knowledge and schemas to integrate incoming information and construct a coherent mental representation of the text.

Despite these theoretical advances, capturing the cognitive processes underlying coherent text construction remains challenging. Consequently, traditional comprehension assessments have come under growing scrutiny, as recent research questions their validity and reliability. Among the most commonly used approaches are think-aloud protocols, applied either concurrently or retrospectively (Davis & Bistodeau, 1993; Zhang & Zhang, 2020). While this method provides rich insights into readers' cognitive processes, it is also intrusive: verbalizing thoughts during reading imposes an additional task that can disrupt the primary process, reducing efficiency and depth of comprehension (Veenman et al., 2006). In his review, Fox (2009) noted that the think-aloud protocol is the most widely adopted methodology in language processing research, primarily because of the rich data it generates. Nonetheless, a major drawback of this method is its invasive nature: the process of verbalizing thoughts while reading imposes an additional cognitive load, which competes with the primary task and may distort natural processing (Davison et al., 1997; Van Someren et al., 1994). This dual-task effect can hinder efficiency, particularly in complex tasks that demand deep comprehension (Veenman et al., 2006). As a result, researchers have turned to less intrusive alternatives such as eye tracking, which provides a more objective window into cognitive processes.

Eye tracking refers to the process of recording eye movements and identifying the precise point of gaze, or where an individual directs their visual attention (Jarodzka & Brand-Gruwel, 2017; Lai et al., 2013). However, early eye-tracking systems also carried limitations, often requiring invasive procedures such as contact lenses fitted with search coils or the use of chin rests and bite bars to ensure fixed head positioning under conditions largely restricted to specialized laboratory settings (Hutt et al., 2024). In contrast to early and traditional eye trackers, the present study employs webcam-based eye tracking, a cost-effective, scalable, and minimally invasive technology that has significantly advanced research on the relationship between eye movements and cognition (Eckstein et al., 2017; Yang & Krajbich, 2023).

Despite extensive research on reading comprehension, most studies have focused primarily on the final comprehension product, defined as the measurable outcome or score that reflects the extent to which readers have understood a text, overlooking the underlying cognitive processes that lead to comprehension. This narrow focus limits our understanding of how readers actively process and integrate textual information. Therefore, there is a clear need for a dual-perspective approach that examines both the reading process and the final comprehension outcome, enabling the collection of richer and more comprehensive data on reading behavior.

Despite the growing global use of process-oriented methodologies, there remains a notable scarcity of studies employing eye-tracking technology in general and webcam-based eye-tracking technology in particular within the Moroccan research context. Although eye tracking has proven to be a powerful tool for uncovering the cognitive and attentional mechanisms underlying reading, it has yet to be integrated into local research practices. Traditional laboratory-based systems, while insightful, are costly, intrusive, and logistically demanding, limiting their feasibility in educational environments. In contrast, webcam-based eye tracking provides a non-intrusive, accessible, and ecologically valid alternative capable of capturing readers' real-time cognitive activity. The absence of such technology from local research underscores a significant methodological gap and highlights the need for studies that harness these innovative tools to advance the understanding of reading behavior among Moroccan EFL learners. In this regard, this study seeks to answer the following research questions:

1. What patterns can be observed in global reading measures such as fixation count, fixation duration, revisits, and gaze patterns across participants?
2. To what extent do webcam-based eye-tracking metrics correlate with EFL learners' reading comprehension performance?
3. What are the implications of using webcam-based eye tracking for analyzing reading processes in educational settings among EFL learners?

## 2. Materials and Methods

### 2.1. Participants

Fifteen first-year undergraduate students from the Faculty of Education Sciences participated in this study. To minimize variability related to age-associated cognitive and oculomotor changes, all participants were aged between 18 and 24 years, a range in which such decline is minimal. Individuals with severe uncorrected visual or oculomotor impairments were excluded. In addition, the use of glasses was not permitted to ensure accurate gaze detection. Data collection took place in a well-lit environment, and participants were instructed to minimize head movement to maximize the precision of eye-tracking measures. Prior to participation, informed consent was obtained from all participants in accordance with ethical research standards.

### 2.2. Design

This study investigates the potential of webcam-based eye tracking to examine the reading patterns of EFL undergraduate students. It adopts a quantitative, exploratory correlational design, focusing on real-time cognitive processes during reading and the relationship between these processes and comprehension outcomes. Webcam eye-tracking technology was used to capture participants' visual attention, providing a non-intrusive method for analyzing reading behavior in naturalistic settings.

### 2.3. Instrumentation

#### 2.3.1. The Reading Passage

A 273-word passage on the invention of the first vaccine was used as the reading stimulus and was displayed for 120 seconds without any visual illustrations. To reduce cognitive load during the limited exposure time, all words in the passage were drawn from the first five frequency bands, covering the 4,000 most frequent English words. Reading comprehension was assessed immediately after the task using six multiple-choice questions.

#### 2.3.2. The Eye Tracking Technique

Visual attention and eye movements were recorded using the *RealEye* platform, a web-based eye-tracking tool that leverages standard webcams to capture real-time gaze data. *RealEye* provides a cost-effective and accessible alternative to traditional, high-cost eye-tracking systems, allowing researchers to examine participants' gaze behavior and the underlying cognitive processes involved in reading. Its non-intrusive design enables participants to engage naturally with the text, providing a realistic representation of how visual attention unfolds during reading.

### 2.4. Apparatus

The reading passage was presented in Calibri font, 12-point font, black text on a white background to ensure readability and minimize visual distractions. Eye-tracking data were collected using the Logitech C920 Pro HD webcam, which provides high-definition (1080p) video and autofocus capabilities. This setup enables stable and reasonably accurate gaze detection, making it suitable for webcam-based eye-tracking analysis. In this regard, four key eye-tracking metrics were analyzed:

1. **Total Fixation Count**: the total number of times a participant fixated on any part of the text, reflecting the overall attention allocated to the reading material;
2. **Average Fixation Duration**: the mean length of each fixation, indicating the depth of cognitive processing required for different words or phrases;
3. **Total Revisits**: the number of times participants returned to previously read sections, reflecting rereading behavior and comprehension monitoring;
4. **Total Number of Gazes**: the total number of distinct eye movements across the text, capturing how visual attention is distributed throughout reading.

To ensure that only meaningful eye movements were analyzed, a noise reduction threshold of 150 milliseconds was applied as recommended by Galley et al. (2015). This means that any fixations shorter than 150 ms were filtered out, as extremely brief fixations are unlikely to represent deliberate cognitive processing and are often considered noise in eye-tracking data. By applying this threshold, the analysis focused on gaze patterns that reliably reflect the cognitive processes involved in reading.

### 2.5. Procedure

Upon accessing the study portal, participants received clear, step-by-step instructions to optimize their device setup and facial positioning for accurate webcam-based gaze capture. The *RealEye* platform provided real-time visual feedback to ensure proper eye detection, prompting participants to make minor adjustments to their seating position or lighting conditions when necessary. Once optimal alignment was achieved, participants completed a 40-point calibration procedure, designed to balance efficiency with precision and ensure high-quality gaze-tracking accuracy. Following calibration, participants were given specific instructions to maintain optimal conditions throughout the task. They were then presented with the reading passage for a fixed duration. Immediately after completing the reading phase, they responded to six multiple-choice comprehension questions assessing their understanding of the text. At the conclusion of the session, participants were warmly thanked for their participation and debriefed about the study's purpose.

## 3. Results and Analysis

**What patterns can be observed in global reading measures such as fixation count, fixation duration, revisits, and gazes across participants?**

Table 1 summarizes the descriptive statistics for the eye-tracking measures obtained during the reading task. Participants produced an average of 275 fixations (*SD* = 14.60), with values ranging from 245 to 294. The mean fixation duration was 339.50 ms (*SD* = 18.20), ranging between 290 and 370 ms. The total number of gazes averaged 4608.60 (*SD* = 261.60), with a minimum of 4161 and a maximum of 4972. Finally, the mean number of regressions was 25.50 (*SD* = 12.50), ranging from 5 to 54. These results indicate relatively low variability across participants for fixation and gaze measures, suggesting consistent visual engagement and reading behavior throughout the task. The slightly higher dispersion in regressions reflects individual differences in rereading behavior and text-processing strategies.

Table 2 presents the descriptive statistics for fixation count, average fixation duration, and regressions across the four Areas of Interest (AOIs), corresponding to the four paragraphs of the reading text. For AOI 1**,** participants produced a mean of *M* = 19.13 fixations (*SD* = 13.54) with an average fixation duration of *M* = 382.86 ms (*SD* = 46.66) and *M* = 12.53 regressions (*SD* = 9.73). In AOI 2**,** fixation count increased substantially to *M* = 102.33 (*SD* = 24.17), accompanied by a shorter mean fixation duration of *M* = 342.00 ms (*SD* = 21.44) and a higher number of regressions (*M* = 45.06, SD = 12.08***).** AOI 3 displayed a similar pattern, with a fixation count of *M* = 111.40 (*SD* = 27.55), an average fixation duration of *M* = 333.33 ms

(*SD* = 19.51), and *M* = 45.53 regressions (*SD* = 12.85*). Finally, AOI 4 yielded *M* = 39.73 fixations (*SD* = 17.11), an average fixation duration of *M* = 334.00 ms (*SD* = 27.46), and *M* = 17.93 regressions (*SD* = 9.32*). Overall, fixation counts and regressions were markedly higher in AOIs 2 and 3 compared to AOIs 1 and 4, indicating that participants allocated greater visual attention and rereading effort to the central sections of the text. Conversely, the longer fixation durations observed in AOI 1 suggest increased cognitive effort at the beginning of the passage, potentially reflecting initial text familiarization or lexical processing.

**Table 1**

*Descriptive Statistics for Eye-Tracking Measures During the Reading Task*

| | Mean (M) | Std.Deviation (SD) | Min | Max |
|---|---|---|---|---|
| **Fixation Count** | 275 | 14.6 | 245 | 294 |
| **Fixation Duration Average (ms)** | 339.5 | 18.2 | 290 | 370 |
| **Total Gazes** | 4608.6 | 261.6 | 4161 | 4972 |
| **Total Regressions** | 25.5 | 12.5 | 5 | 54 |

**Table 2**

*Descriptive Statistics for Eye-Tracking Measures Across the Four Areas of Interest (AOIs)*

| Areas of Interest | Fixation Count M (SD) | Average Fixation Duration M (SD) | Regressions M (SD) |
|---|---|---|---|
| **1** | 19.13 (13.54) | 382.86 (46.66) | 12.53 (9.73) |
| **2** | 102.33 (24.17) | 342.00 (21.44) | 45.06 (12.08) |
| **3** | 111.40 (27.55) | 333.33 (19.51) | 45.53 (12.85) |
| **4** | 39.73 (17.11) | 334.00 (27.46) | 17.93 (9.32) |

**To what extent are these eye-tracking measures correlated with reading comprehension performance?**

Table 1 presents Spearman's rank-order correlations were conducted to explore the relationships between global eye-tracking metrics and reading comprehension scores. The analysis included fixation count, fixation duration, total gazes, and number of revisits across the entire text. No statistically significant correlations were found between comprehension scores and fixation count ($\rho = .29$, $p = .301$), fixation duration ($\rho = –.32$, $p = .240$), number of gazes ($\rho = .11$, $p = .694$), or number of revisits ($\rho = –.15$, $p = .590$). However, a very strong and statistically significant negative correlation was observed between total gazes and revisits ($\rho = –.98$, $p < .001$), suggesting that participants who gazed more extensively across the text tended to make substantially fewer regressions. Other correlations among the eye-tracking variables were moderate but did not reach statistical significance.

In addition, Spearman's rank-order correlations were computed to examine the relationships among paragraph-level eye-tracking metrics and reading comprehension scores. The analysis revealed several significant interrelations among eye-movement measures, as well as specific associations with comprehension. For instance, fixation count and revisits in the first paragraph were strongly and positively correlated (ρ = .99, $p$ < .001), suggesting that readers who fixated more frequently on this section also tended to return to it more often. Similarly, fixation duration in the first paragraph was positively correlated with fixation duration in the second paragraph (ρ = .71, $p$ = .003), indicating some consistency in visual engagement across adjacent sections. Furthermore, a strong negative correlation emerged between fixation count in the first and third paragraphs (ρ = –.72, $p$ = .002), showing that readers who devoted more visual attention to the opening paragraph tended to invest less in the third one. Comparable negative relationships were found between revisits to the first and third paragraphs (ρ = –.73, $p$ = .002).

Regarding comprehension, two positive correlations were statistically significant. Revisits to the third paragraph were moderately and significantly correlated with comprehension scores (ρ = .56, $p$ = .030), indicating that participants who reprocessed this section tended to achieve higher comprehension performance.

Eye-tracking measures in the second paragraph also showed notable tendencies. Revisits to this paragraph were moderately correlated with comprehension (ρ = .41, $p$ = .129), and fixation count displayed a weak positive association (ρ = .14, $p$ = .611). However, neither was statistically significant.

Overall, these results suggest that while eye-movement behaviors across paragraphs were highly interrelated, only the third paragraph's metrics particularly the number of revisits showed a meaningful link to comprehension performance. This finding underscores the importance of strategic rereading in facilitating deeper text understanding.

**Table 3**

*Correlation Matrix Among Eye-Tracking Measures and Reading Comprehension Scores*

| | **Fixation Count** | **Fixation Duration** | **Total Gazes** | **Total Regressions** | **Comprehension Scores** |
|---|---|---|---|---|---|
| **Fixation Count** | 1 | -.41 | .18 | -.15 | .28 |
| **Fixation Duration** | -.41 | 1 | .41 | -.40 | -.32 |
| **Total Gazes** | .18 | .41 | 1 | -.97** | .11 |
| **Total Regressions** | -.15 | -.40 | -.97** | 1 | -.15 |
| **Comprehension Scores** | .28 | -.32 | .11 | -.15 | 1 |

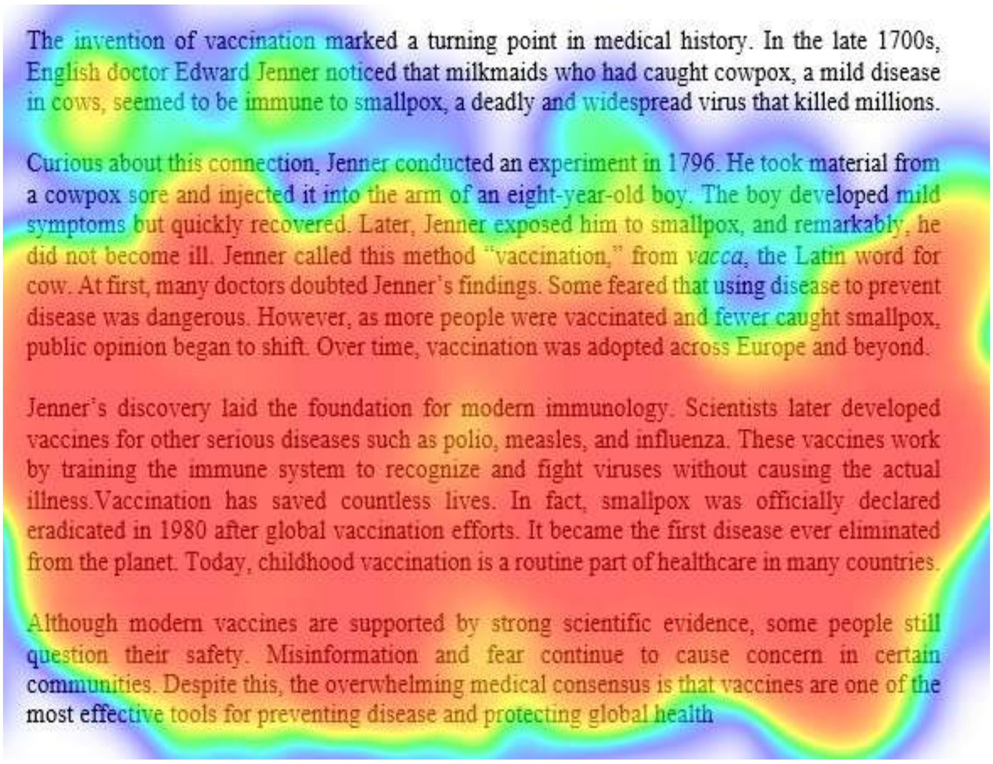
**Table 4**

*Correlation Matrix Between Eye-Tracking Measures Across the Four Areas of Interest (AOIs) and Reading Comprehension Scores*

| | Fixation Count (1) | Fixation Duration (1) | Regression (1) | Fixation Count (2) | Fixation Duration (2) | Regression (2) | Fixation Count (3) | Fixation Duration (3) | Regression (3 ) | Fixation Count (4) | Fixation Duration (4) | Regression (4) | Comprehension Scores |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fixation Count (1) | 1 | .289 | .988** | .383 | .220 | .370 | -.723** | -.721** | -.241 | -.122 | .250 | -.086 | .153 |
| Fixation Duration (1) | .289 | 1 | .228 | -.251 | .714** | -.350 | -.034 | .138 | -.217 | .178 | .505 | -.052 | .220 |
| Regressions (1) | .988** | .228 | 1 | .448 | .184 | .457 | -.727** | -.729** | -.204 | -.181 | .261 | -.128 | .125 |
| Fixation Count (2) | .383 | -.251 | .448 | 1 | -165 | .648** | -.259 | -.297 | -.072 | -.817** | .180 | -.731** | .143 |
| Fixation Duration (2) | .220 | .714** | .184 | -.165 | 1 | -.374 | -.219 | .049 | -.448 | .128 | .515* | -.240 | -.048 |
| Regressions (2) | .370 | -.350 | .457 | .648** | -.374 | 1 | -.113 | -.325 | .556* | -.487 | -.088 | -.090 | .410 |
| Fixation Count (3) | -.723** | -.034 | -.727** | -.259 | -.219 | -.113 | 1 | .623* | .469 | -.199 | -.429 | .010 | .173 |
| Fixation Duration (3) | -.721 | .138 | -.729** | -.297 | .049 | -.325 | .623* | 1 | .143 | .050 | .263 | -.015 | -.141 |
| Regressions (3) | -.241 | -.217 | -.204 | -.072 | -.448 | .556* | .469 | .143 | 1 | .073 | -.433 | .447 | .560* |
| Fixation Count (4) | -.122 | .178 | -.181 | -.817** | .128 | -.487 | -.199 | .050 | .073 | 1 | -.022 | .816** | -.095 |
| Fixation Duration (4) | .250 | .505 | .261 | .180 | .515* | -.088 | -.429 | .263 | -.433 | -.022 | 1 | -.343 | -.211 |
| Regressions (4) | -.086 | -.052 | -.128 | -.731** | -.240 | -.090 | .010 | -.015 | .447 | .816** | -.343 | 1 | .094 |
| Comprehension Scores | .153 | .220 | .125 | .143 | -.048 | .410 | .173 | -.141 | .560* | -.095 | -.211 | .094 | 1 |

**Figure 1**

*Correlation Matrix Between Eye-Tracking Measures Across the Four Areas of Interest (AOIs) and Reading Comprehension Scores*




The heatmap illustrates the aggregated eye-tracking behavior of all participants across the four paragraphs of the reading text. Areas of fixation are color-coded according to their density, with warmer tones (red) representing higher fixation frequencies and cooler tones (yellow to green) indicating fewer fixations.

As shown in the figure, fixations are predominantly concentrated in the third and fourth paragraphs, where red zones are more pronounced compared to the first and second paragraphs. This distribution suggests that participants allocated greater visual attention and cognitive effort to these later sections of the text. The intensity of fixation clusters in the third and fourth paragraphs may reflect increased engagement with or difficulty in processing the content presented in these parts, while the relatively cooler tones in the earlier paragraphs indicate less intensive visual processing.

## 4. Discussion and Conclusion

The most significant finding of this study was the strong positive correlation between revisits to the third paragraph and comprehension scores and the strong correlation between regressions to the second paragraph and comprehension scores. This pattern indicates that readers who allocated more visual attention to these sections by fixating on them more frequently and returning to it multiple times achieved higher levels of understanding of the overall text (León et al., 2019). This observation is further supported by an experiment conducted by Schotter et al. (2014), in which the researchers' restricted readers from making regressions within the text. Their findings revealed a significant decline in comprehension performance when regressions were prevented, underscoring the essential role of rereading in constructing coherent mental representations of the text. In this regard, this behavior may suggest that the third paragraph may have contained information that was conceptually rich, inferentially demanding, or central to the text's main ideas, thus requiring additional cognitive effort to process and integrate. The increased number of fixations and revisits likely reflects strategic rereading and enhanced cognitive engagement, which facilitate the construction of coherent mental representations and deeper comprehension.

In addition, revisits to the second paragraph showed a moderate positive correlation with comprehension, approaching statistical significance. This trend aligns with previous eye-tracking research (*e.g.,* Rayner et al., 2006), which emphasizes that readers often return to earlier or complex portions of a text when comprehension difficulties arise or when meaning construction requires re-evaluation of prior information. Such rereading behavior is not a sign of inefficiency; rather, it is a hallmark of active reading and metacognitive regulation, whereby readers monitor their understanding and strategically allocate attention to resolve ambiguity or reinforce coherence (Weiss, 2020).

Although the correlations between comprehension and other measures such as fixation counts in other paragraphs, overall gaze duration, or total regressions across the text did not reach statistical significance, these findings still offer valuable descriptive insights into readers' visual and cognitive processing patterns. They suggest that while not all sections of the text demanded equal cognitive resources, localized increases in visual attention (particularly in the second and third paragraphs) were most strongly associated with deeper comprehension outcomes.

For instance, some participants exhibited high fixation or gaze counts without corresponding gains in comprehension performance, suggesting that an increased amount of visual attention does not necessarily translate into deeper cognitive processing. This pattern highlights an important distinction between quantity of attention and quality of processing. Extended gaze duration or frequent fixations may, in some cases, reflect mind wandering, superficial scanning, or divided attention, rather than meaningful engagement with the text (Faber et al., 2020). Such findings reinforce the idea that effective reading comprehension depends not only on visual attention but also on how that attention is cognitively directed and regulated.

From a methodological perspective, the findings also demonstrate that webcam-based eye tracking can serve as a viable and practical alternative for educational and cognitive research in contexts where access to traditional infrared eye-tracking systems is limited (Dostálová & Plch, 2023; Wong et al., 2023). While this method does not provide the fine-grained spatial accuracy needed to capture word-level or sub-sentence eye movements, it nonetheless offers valuable insights into paragraph-level and section-level engagement. This makes it particularly suitable for studies involving longer, more naturalistic reading materials, where the goal is to understand overall reading patterns and attentional distribution rather than precise lexical processing.

## 5. Limitations and Contributions

The findings of this study have several pedagogical and methodological implications. Webcam-based eye tracking can serve as a valuable tool for educators, allowing them to identify which parts of a text pose greater cognitive challenges for students. By visualizing attention distribution, teachers can monitor learners' engagement levels and detect moments of distraction or surface-level reading. Consequently, such insights can guide the adaptation and restructuring of reading materials, ensuring that instructional focus is directed toward sections that require additional scaffolding. Moreover, the technology encourages students' self-reflection on their reading behaviors, leading to the development of metacognitive awareness and self-regulated learning strategies. From a broader perspective, webcam-based eye tracking represents a low-cost, accessible, and scalable approach for assessing reading processes in EFL classrooms, offering insight not only into comprehension outcomes but also into the underlying mechanisms of engagement and attention.

Despite its promising contributions, this study is not without limitations. The small sample size may have restricted the generalizability of the findings and reduced the statistical power of certain correlations. Additionally, the webcam-based eye-tracking system used in this research has lower temporal and spatial resolution compared to infrared systems, which limits its ability to capture fine-grained eye movements at the word or sentence level. As a result, the analysis was confined to paragraph-level patterns rather than detailed lexical processing. Finally, the study employed a single reading passage, which constrains the extent to which the results can be extrapolated to different text types, genres, or lengths.

Future research should aim to replicate and extend these findings with larger and more diverse participant samples to enhance statistical reliability and generalizability. Incorporating multiple reading passages of varying complexity and genres would provide a richer understanding of how text features interact with readers' attentional patterns. Furthermore, combining webcam-based tracking with complementary measures such as comprehension tests, self-reports, or physiological indicators could yield a more holistic picture of reading behavior. Finally, ongoing technological improvements in webcam-based eye tracking may enhance its precision, making it an even more powerful and accessible tool for classroom-based cognitive and educational research.

## Bibliography

Britton, B. K., & Graesser, A. C. (Eds.). (2014). Models of understanding text. Psychology Press. https://doi.org/10.4324/9781315806143

Butterfuss, R., Kim, J., & Kendeou, P. (2020). Reading Comprehension. In Oxford research encyclopedia of education. https://doi.org/10.1093/acrefore/9780190264093.013.865

Davis, J. N., & Bistodeau, L. (1993). How do L1 and L2 reading differ? Evidence from think aloud protocols. *The Modern Language Journal, 77*(4), 459-472. https://doi.org/10.2307/329671

Davison, G. C., Vogel, R. S., & Coffman, S. G. (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *Journal of Consulting and Clinical Psychology, 65*(6), 950-959. https://doi.org/10.1037/0022-006X.65.6.950

Dostálová, N., & Plch, L. (2023). A scoping review of webcam eye tracking in learning and education. *Studia paedagogica, 28*(3), 113-131. https://doi.org/10.5817/SP2023-3-5

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience, 25*, 69-91. https://doi.org/10.1016/j.dcn.2016.11.001

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance, 46*(10), 1201-1221. https://doi.org/10.1037/xhp0000743

Fox, E. (2009). The role of reader characteristics in processing and learning from informational text. *Review of Educational Research, 79*(1), 197-261. https://doi.org/10.3102/0034654308324654

Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations-Why are they so highly variable. *Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie*, *93*, 83-106. https://doi.org/10.13140/RG.2.1.3128.1769

Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2024). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods, 56*(1), 1-17. https://doi.org/10.3758/s13428-022-02040-x

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. *Journal of Computer Assisted Learning, 33*(3), 193-201. https://doi.org/10.1111/jcal.12189

Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 62-69. https://doi.org/10.1177/2372732215624707

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. https://doi.org/10.1037/0033-295X.85.5.363

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Chiou, G.-L., Liang, J.-C. & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review, 10*, 90-115. https://doi.org/10.1016/j.edurev.2013.10.001

León, J. A., Moreno, J. D., Escudero, I., & Kaakinen, J. K. (2019). Selective attention to question-relevant text information precedes high-quality summaries: Evidence from eye movements. *Journal of Eye Movement Research, 12*(1), 1-16. https://doi.org/10.16910/jemr.12.1.6

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 51, pp. 297-384). Academic Press. https://doi.org/10.1016/S0079-7421(09)51009-2

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241-255. https://doi.org/10.1207/s1532799xssr1003_3

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science, 25*(6), 1218-1226. https://doi.org/10.1177/0956797614531148

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, *46*, 374-393. https://doi.org/10.1002/RRQ.005

Van den Broek, P., & Kremer, K. E. (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades*. Teachers College Press.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: a practical approach to modelling cognitive processes*. Academic Press.

Veenman, M. V., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3-14. https://doi.org/10.1007/s11409-006-6893-0

Weiss, A. F. (2020). The information gathering framework—A cognitive model of regressive eye movements during reading. *Journal of Eye Movement Research, 13*(4), 31. https://doi.org/10.16910/jemr.13.4.4

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a webcam- based eye-tracker to understand students' thought patterns and reading behaviors in neurodivergent classrooms. In *Proceedings of the 13th International Learning Analytics and Knowledge Conference (LAK' 23)* (pp. 453-463). https://doi.org/10.1145/3576050.3576115

Yang, X., & Krajbich, I. (2023). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making, 16*(6), 1485-1505. https://doi.org/10.1017/S1930297500008512

Zhang, L. J., & Zhang, D. (2020). Think-aloud protocols. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (p.p 302-311). Routledge. https://doi.org/10.4324/9780367824471-26

# GENERATIVE ARTIFICIAL INTELLIGENCE IN FLE LEARNING IN HIGHER EDUCATION: CROSS-PERSPECTIVES ON PERSONALIZATION AND ETHICS AT THE FACULTY OF SCIENCES OF RABAT

**Aziza Elmouhtadi[1], Souhad Shlaka[2], Najemeddin Soughati[1]**

*[1]Université Ibn Tofaïl,*
*Faculté des Langues, des Lettres et des Arts – Laboratoire de Littérature, d'Arts Et d'Ingénierie Pédagogique,*
*Kénitra,Maroc.*

*[2]Université Mohammed V,*
*Faculté des Sciences,*
*Rabat, Maroc.*

**ABSTRACT**

The rise of Generative Artificial Intelligence (GenAI) drastically transforms teaching and learning practices in higher education more specifically in non-linguistic disciplines with French as the language of learning and teaching. The research examines the role and effects of GenAI in the pedagogical context of learning French as a Foreign Language (FLE) at the Faculty of Sciences of Rabat (FSR). Through a personalization, ethical, and autonomy lens. Positioned amidst Morocco's national reform of higher education, namely the Framework Law No. 51-17 213-212 and the PACTE ESRI 2030, this research sheds light on how deep pedagogical innovation interacts with institutional transformation focused on excellence, equity, and human capital development. The methodology included a questionnaire with 316 responses, participant observation with four classes of 80 students and two additional classes of 46 students, as well as one-to-one interviews with students and teachers. Two writing drafts were compared in the context of FLE: the first produced with the support of GenAI and the second written independently. The impacts provided from the results motivation, language quality and reflective engagement, and were then measured and compared. Results revealed a polarization. For advanced learners (C1 level), AI functions as a co-writing and revision tool, whereas beginners (A2 level) often become a substitute for writing, which constitutes cognitive dependency and limited autonomy. The notions of digital maturity and ethical guidance emerge as key conditions for GenAI's effective integration. Our analysis draws on recent theoretical frameworks on generative AI and language learning.

The findings suggest a training model that combines complex personalization, critical reflexivity, and pedagogical supervision. Thus, AI appears simultaneously as a potential source of disempowerment and as a differentiating pharmakon. Finally, the article underscores the importance of developing an educational AI culture rooted in accountability, transparency, and shared meaning.

***Keywords:*** *generative artificial intelligence, personalized learning, AI ethics in education, French as a Foreign Language (FLE), higher education, digital autonomy, AI literacy, digital maturity.*

## 1. Introduction

The emergence of generative artificial intelligence in education promises a paradigm shift in what we know, how we teach, and how we learn. Within a relatively short amount of time, information technologies based on deep learning and natural language processing have offered massive opportunities and challenges for institutions of higher education, shaping the ways in which students can both access and produce knowledge. As Desveaud (2024) and Holmes and Tuomi (2022) argue, it took mere months for young learners at the Faculty of Sciences of Rabat to start using ChatGPT in an unstructured way, testing their capacity for casual interaction or their recently acquired knowledge as both learners and producers in a French-speaking scientific community. This double contingency - the need to understand how technologies transform the experience of learning and to define the responsible, reflective, and ethical framework of such transformations - motivates the need for holistic GenAI research. This study aims to follow the strategic orientations of the Pacte ESRI 2030 and the *Livre blanc* of the Moroccan Employers' Confederation, Confédération générale des entreprises du Maroc (2021), both calling for "a pedagogy based on intelligence" and a level-playing field application of digital technologies as a reform instrument. These priorities resonate with international frameworks for the responsible integration of AI in education. From this properly scaled pedagogical discourse, the integration of French as a foreign language teaching into GenAI unfolds, not as a technological curiosity but as a lens through which one can observe major shifts in cognitive and language mediations within scientific multilingualism.

As Holmes and Tuomi (2022) argue about Artificial Intelligence in Education as AIEd: it is about thinking of AIEd not merely as automation or a new precision tool for an age-old task in which education has always been engaged but rather as necessitating a redefinition of that task and its fundamental reason for being. It is a case of teaching more alongside the machine, to teach less, but not to replace nor to diminish, human learning.

In the view of Romero et al. (2023) expressed in the White Paper on Teaching and Learning in the Age of Artificial Intelligence, this requires integrating AI as a part of a pedagogy based on critical regulation. In the education-focused context, this is contrasted to a view of technological determinism, which is the theoretical background of, for example, Cristol (2019) or Cavenaghi and Senécal (2019), also focusing on education.

This study's conceptual framework is structured around a triple metaphor initially formulated during the 3rd International Conference on Humanities and Social Sciences (Porto, July 2025) and informed by recent studies on the pedagogical integration of generative AI in higher education (Jabagi & Croteau, 2025; Naffi & Montufar, 2025; Perrin et al., 2025). The first metaphor is personalization, and it refers to a form of educational triage and skilful support highlighted by Holmes et al. (2019) and Shah (2023), presenting AI as a new way of assisting learning while empowering learners rather than undermining their independence. The second metaphor is ethical and relates to the concerns outlined in policy documents published by Miao e Cukurova (2024) and Miao et al. (2024) and the OECD (2021). The third metaphor concerns autonomy and digital maturity, drawing on the work of Holec (1981) and Stiegler (2005, 2016). Together, these three metaphors shape the figurative concept of formative personalization discussed above.

The experiment presented in this article, conducted at the Faculty of Sciences of Rabat (FSR), explores various modes of GenAI use by students through writing, revision, and reflection in French as a Foreign Language (FLE). After a phase of unstructured practice application, which allowed us to assess its impact on motivation, linguistic production quality, and metacognition, a mode of guided exploration phase was then defined. In this sense, our position aligns with Kurban and Şahin (2024), who argue that AI is reshaping the roles of both teachers and learners within the educational microsystem. We are not concerned with assessing the tool's efficiency but rather with understanding a process: how writing with AI can become an ethical and formative socius.

The structure of the article is organized into four main sections. The first section, the Introduction, presents the institutional and theoretical framework of the study. The second section, Materials and Methods, describes the research methodology and the instruments used for data collection. The third section, Results and Analysis, compares students' writing productions with and without the use of generative artificial intelligence (GenAI). The final section discusses the pedagogical, ethical, and institutional implications associated with the use of generative AI in the context of FLE in scientific higher education.

## 2. Materials and Methods

The fieldwork was carried out in the context of the Moroccan university, which is characterized by a recent evolution of pedagogical and digital transformation. In line with the current transformation of higher education policies. In other words, integrating artificial intelligence should not be considered just a technical issue but rather a paradigm shift that will modify the major profile of the new generations of education and intelligence, as a function of three ethical capacitors and intellectual power: transparency, trust, and critical thinking. Confédération générale des entreprises du Maroc (2021) stipulates that “The main challenge is to deliver a pedagogy of intelligence: teaching how to think well, to make decisions. Critical thinking, personal development and innovation, and education for citizenship are the three converging axes of this new education” (p. 136). There is an intention to put forward a model of continuous training of not only high-performing learners, but also learners capable of appropriating this technology, while also embodying and respecting the democratic values of civility, fairness, and public responsibility. The experimentation protocol was conducted following a series of preliminary annual observations carried out with first - year undergraduate students enrolled in scientific programs at FSR, mainly in the mathematics - Computer Science - Physics (MIP) track and the Applied Computer Science (InfApp) stream. In addition to the theoretical lectures, students attended a weekly 90-minute practical session intended to support course activities and provide opportunities for guided practice. The remaining course activities were conducted through the Moodle platform. A weekly synchronous session was organized via Google Meet to provide instructions, answer students’ questions, and support the learning activities. Students could then progress at their own pace by following the additional instructions posted on the platform at the beginning of the course, in accordance with the pedagogical scenario.

Preliminary observations revealed that students were already using ChatGPT informally, mainly to rephrase, translate, or clarify difficult terms or concepts. Although such activity may stem from cognitive curiosity, it lacks a reflective dimension: for many students, AI appears to function as a productivity tool rather than a reflective partner. Based on these observations, the research design sought to formalize this intuitive practice into a structured experiment and to integrate personalized learning and ethical reflection on the use of AI into an educational assignment. This approach aligns with the key idea expressed in the Confédération générale des entreprises du Maroc Livre blanc (2021), which calls for integrating social, technological, and economic transformations into curricula while emphasizing quality rather than quantity. Hence, the present study aimed to explore how generative AI, when used in a guided way, can support rhetorical development in an academic scientific context.

The study adopts a mixed-methods approach combining quantitative and qualitative data collection methods (Creswell & Clark, 2018). The questionnaire was distributed through the Moodle platform and institutional email over three weeks, with three reminder messages sent to participants. It targeted 500 students enrolled in the MIP and InfApp programs. More specifically, the sample size, determined using Cochran's (1977) formula, yielded a total of 316 valid responses or 44% of the population, with a margin of error of 5% and a confidence level of 95%. It consisted of three sections: one that evaluated the FLE course, the use and perception of generative AI, and digital maturity, autonomy, ethics in turn, for a total of 46 questions, consistently transposed into Arabic and back - translated prior to data collection. The questionnaire included a mix of closed and open-ended questions, as well as items measured on a six-point Likert scale, adapted from validated instruments by Jong et al. (2020) and Stöhr et al. (2024), and subsequently transposed to the Moroccan cultural and educational context. The results were analysed using SPSS version 28 through descriptive and correlational analyses focusing on the relationships between frequency of use, perceived usefulness, motivation, sense of autonomy, and student engagement.

At the same time, during the first and second semesters of the 2023-2024 academic year, participant observation was conducted with six groups of first-year undergraduate students enrolled in the MIP and InfApp programmes (N = 126). The purpose was to capture how learning postures shift. For this reason, the observation followed three successive phases. First, students wrote an in-class text on a scientific topic without the assistance of AI. Then, they produced a second track, using ChatGPT if they decided to rephrase, to enrich or clarify some idea. Finally, the third phase was reflective and included a comparison of both writing outputs, distinguishing changes, similarities, and writing strategies or patterns. Participant observation revealed three main usage profiles among the students: some reproduced the AI suggestions with little critical distance, others engaged in a critical dialogue with the tool, while a smaller group displayed signs of cognitive dependence.

To investigate representations, resistance and ethical dilemmas regarding AI-assisted writing, 24 semi-structured interviews with students and 6 FLE instructors were conducted. Student interviews, which lasted from 10 to 15 minutes, were conducted "live" at the end of the experimental sessions, while some instructors, due to timetable constraints, answered the interview questions sent to them digitally, via *WhatsApp*.

The transcripts were analysed according to Miles et al. (2020) methodology and involved open, axial and selective coding to identify recurring patterns and articulate relationships among categories. Therefore, the interpretation was constructed along three axes: stance towards the technology, perception of autonomy and ethical issues in AI-generated content in writing tasks.

The writing productions were evaluated using an analytical rubric inspired by the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2020) and adapted to scientific FLE, assessing four dimensions: discursive coherence, grammatical accuracy, lexical richness, and argumentative relevance.

All data collection procedures complied with UNESCO's principles for the responsible use of artificial intelligence in education. Participants were informed about the objectives of the study and provided written consent, and no academic grading was associated with the experiment to avoid performance or authority-related bias.

## 3. Results and analysis

The experiment revealed divergent dynamics underlying the intuitive and reflective uses of GenAI in FLE-University-level writing. The observations of student behaviour in MIP and InfApp conditions, revealing three main patterns of appropriation: situated or instrumental use; using AI as a mining tool, or cognitive track; a critical and reflective use of GenAI. This transition was discernible when a student goes from "I correct" to "I ask the tool to teach me how to say…", reflecting deeper routinization of their apprenticeship. Furthermore, this categorization aligns with the four stages of digital literacy described by Holmes et al. (2019), reflecting a progression from technological dependency to the co-construction of meaning.

Linguistically, comparing the drafts written before and during the use of AI revealed general improvements in textual coherence and lexical richness, while also highlighting differences in stylistic personalization. At the same time, with the C1 group, the tool acted as a trigger for critical thinking, given the fact that the reformulations drawn from ChatGPT were often rephrased or typically inserted in the text, which can be seen as a reflection. With A2-B1 correspondents, the cooperation with a machine most often led to the partial displacement of effort on the machine, that is, the path was being explored rather superficially, which, in turn, testifies to the rather fragile interaction between students and their cognitive abilities. This pattern almost verbatim repeats Cristol (2024), who very clearly emphasizes that autonomy is territorially uneven and that access to it in the digital age varies enormously and largely depends on frames and competencies. In our sample (Elmouhtadi et al., 2025), these patterns appear to align with the students' techno-pedagogical and linguistic maturity.

On the motivational level, the questionnaire data indicate a significant increase in perceived competence stimulated by GenAI: almost 68% of the participants state that ChatGPT "helped them get their ideas clearer" and pushed them to "produce more in French". Still, 42% demonstrate the standard tendency in denying one's legitimacy: such a high incidence of ethical concern is typical when using an external tool. The combination of evaluation - external appreciation along with the trace of wrongfulness - was exactly the lines traversed by the authors of the thematic issue of *Revue Internationale des Technologies en Pédagogie Universitaire* on AI in higher education (Roy et al., 2025). The thematic issue revealed that improvement of self-efficacy was always connected with moral concern during the growth, as the community authors call differences between attentiveness and the lack of expectations which must have been framed institutionally to comfort participants. Indeed, the guided usage of GenAI can provide linguistic self-assessment and confidence but needs explicit attribution rules on human/machine contributions and provision of familiar routine for attribution in unknown environment. These findings underscore the ambivalence between cognition and morality in data as Shah (2023) notes in recent studies on AI personalization: indeed, AI-based work triggers motivation of further engagement but always accompanied with the blur considering help and replacement - due to more blurred conditions, questions still linger. The thematic analysis of semi-structured interviews brings three axes for interpretation. The first one is digital maturity gap: students with tech experience can correlate a reply with its provenance, reliability, and relevance. Their plead is dialogical, representing the conscious user side according to Forum IA Québec and Institut Intelligence et données (2022). On the contrary, freshmen often describe AI as a "proofreader" or a "translation engine," showing little awareness of how it operates. This behaviour may be interpreted as an indicator of weak self-regulation, as discussed by Kirschner and De Bruyckere (2017).

The April 2025 special issue of the International Journal of Technologies in Higher Education identifies additional analytical perspectives on the integration of generative AI in higher education (Roy et al., 2025). One of these axes complements the focus on digital maturity by highlighting students' ability to regulate their interaction with AI through a metacognitive sequence: "Prompt → Check → Trace → Rewrite." Their texts subsequently show more consistency when compared simultaneously with a second similar axis defined by the students' level of engagement. Groups enjoy assisted drafts created by the AIs or human mimicry of them; however, it is the one guided to compare their human and hybrid drafts that develops a stronger metareflective effort. This enables them to "analyse their drafts for discrepancies," *e.g.,* semantic or cultural misreadings, which include "higher order meta rewriting" via recontextualizing their argumentation. This process corresponds more closely to the description of collaborative knowledge co-construction proposed by Luckin et al. (2016), according to which human–machine collaboration becomes constructive only if the learner retains epistemic responsibility. The FSR experiment acts as a structured and comparative system in which V0 is their own version, V1 is their machine-assisted version, and V2 is their revised version. It acts as the enabling protocol for critical distancing: the student interacting with their drafts must make explicit the linguistic effect desired, justify segues [A1.1], and refine reasoning for their use of it. Such praxis places the use of AI in a rationality of meaning co-creation, consistent with Self-Determination Theory (SDT) (Deci & Ryan, 2012; Ryan & Deci, 2000; Ryan & Deci, 2017) and with recent international calls for AI literacy and ethical awareness in education. A third axis includes linguistic normativity and educational ethics: most student-created material included in the qualitative analysis will link the use of the GenAI with a clear duty of data traceability and responsibility. Most of the students cited the AI tool they used and acknowledged their personal contribution, which aligns with UNESCO's recommendations on traceability and accountability for the educational use of AI.

Illustrating this statement, we find a change in learner posture moving from consumption to reflexivity. Applying the principles of transparency, attribution, and traceability at a local level, based on the principles extracted from the UNESCO (2019), Miao and Cukurova (2024) and Miao et al. (2024) and OECD (2021) frameworks closes "gray areas" experienced by students while stabilizing the assessment context. It is also congruent with national reform striving to align students and teachers around transformed pedagogy and learning culture. For example, the Confédération générale des entreprises du Maroc Livre blanc (2021) proclaims the necessity of switching from rote learning to analysis, critical thinking, and logical reasoning, while insisting on the continuous improvement of education quality. Our field observations indicate that this is indeed the case, as integrated within a supervised formative context, GenAI appears to be a leverage for critical thinking and cyclical pedagogical assessment. Nevertheless, this implementation is conditioned upon the existence of an ethical framework for algorithms and the teachers' ability to maintain a safe learning space, allowing for an orderly trial-and-error context. Such a pedagogy closely aligns with the transparency and guided iteration principles of the ESRI Pact 2030; it is also consistent with the *Confédération générale des entreprises du Maroc*'s imperative of educating for informed citizenship.

In summary, at the institutional level, these findings corroborate the ESRI Pact 2030's focus on personal pathways and excellence powered by innovation and confirm OECD (2021) and UNESCO (2019), Miao and Cukurova (2024), Miao et al. (2024) reference frameworks calling for thoughtful and decisive digital citizens capable of autonomously and critically discerning the AI-generated information's reliability and social outcomes. In this context, the university becomes the place of algorithmic education: Precisely, the aim is not only to learn with AI but to learn about AI - we mean its limits, biases, and ethical outcomes.

Finally, the experiment conducted at the Faculty of Sciences of Rabat demonstrates that generative AI should be understood neither as a simple writing assistant nor as a threat to academic writing, but rather as a cognitive scaffold whose pedagogical value depends on the quality of human mediation. By encouraging students to develop a reflective, critical, and situated engagement with AI-assisted writing, this approach contributes to be one of the key objectives of the *Confédération générale des entreprises du Maroc Livre blanc*: preparing competent and responsible learners capable of linking knowledge, action, and judgment. In this perspective, generative AI may become - to borrow Seldon and Abidoye's (2018) expression - a "school of judgment", where learners develop critical reasoning in interaction with machine-generated texts.

Taken together, the technological performance appears to be counterbalanced by an increasing need to acknowledge human effort, while ethical dilemmas emerge concerning the value of effort, creativity, and responsibility.

## 4. Discussion and Conclusions

The findings reveal contrasting patterns in students' uses of generative AI in scientific FLE writing. This study is guided by a conceptual framework first presented at the 3rd International Conference on Humanities and Social Sciences (Porto, July 2025), structured around three analytical dimensions: personalized learning, the ethics of educational AI, and reflective digital maturity. Together, these dimensions provide a lens for interpreting the different usage profiles observed among students.

In the current Moroccan context of institutional reform, these three dimensions appear to be closely interrelated. On the one hand, they are mutually reinforcing; on the other hand, they remain complementary analytical perspectives. In other words, they reflect the evolving learning practices in Moroccan higher education and the changing relationships between learners, teachers, and digital technologies.

The first dimension, personalization, is illustrated by the differentiated ways in which students with varying linguistic competencies engage with generative AI. While advanced learners (B2-C1) tend to use the tool as a co-author and revision partner, lower-level learners (A2-B1) often rely on it as linguistic support to compensate for their limited proficiency. B2+, learners operate generative AI as a co-author, with the unique capacity to adjust lexical and syntactic complexity, critically review what they write, and consistently incorporate some machine-generated feedback into their active learning dynamic. This perspective is confirmed by Holmes et al. (2019) and Luckin et al. (2016), who all emphasize that genuine personalization requires the experience of interpretive authority and rely on generative AI to assuage their linguistic insecurities. It is therefore mostly imitative, devoid of decisive distance. This difference supports Shah's (2023) hypothesis that students can benefit from AI-generated suggestions while remaining critically aware of the tool's authoritative status.

The second axis, more specifically ethical, pinpoints the gap between performance and responsibility. Although most students consider their outputs to have been improved by generative AI, few of them doubt the academic legitimacy of a work co-authored with the machine. Such an ambiguity is in line with UNESCO's and the OECD's prescriptions, saying that the unregulated integration of AI is likely to exacerbate inequalities, promote unreflective rationalizing, and must be avoided. The experiment of the Faculty of Sciences of Rabat demonstrated that the rigorous implementation of the rules of traceability and attribution, such as citing the tool and versions and insisting on a clearly distinguishable human contribution, triggers an autonomous, ethical, and reflective reproduction of the tool, consistent with the principles of shared responsibility. Such a conclusion complies with the *Confédération générale des entreprises du Maroc Livre blanc*, where civility, effort, and transparency are allied with the quality of training.

Finally, the third axis, affiliated with digital maturity, confirms the diversified trajectories of AI appropriation. Those with more experience use a metacognitive approach; for example, they question the relevance of an AI-generated response, explore the model's limits, justify their changes, and document them. This approach resonates with Stiegler 's (2005, 2016) concept of technique as a pharmakon - the substance, which can be both the remedy and the illness, depending on the dose. Such a posture, according to Forum IA Québec e Institut Intelligence et données (2022) consists of taking sides - exploiting technology, not losing identity and delegating to remain an agent. Diametrically, novices are results-oriented, manifesting instrumental dependence on the technology. AI Fact Sheet of the Rabat experiment verified that digital reflexivity develops over time and follows previous versions, is extrinsically justified, and debated in its bias deposition.

Reinterpreted within the Moroccan institutional ecosystem, these findings echo the objectives of Morocco's ongoing higher education reform agenda, which aims to promote a model of higher education grounded in competence, creativity, and responsibility. The way generative AI has been integrated into the FLE courses is the perfect illustration of this possible blueprint: technology, when framed within an ethical and formative framework, is an accelerator for differentiated learning and linguistic inclusivity. It is no longer a simple technical mediation but also a critical mediation with regards to knowledge production, creating a more complex individual that develops as the autonomous digital citizen aimed at by the UNESCO and the OECD. Our case undoubtedly shows that, when incorporated into a developmental and cooperative learning ecosystem, generative AI has the potential to reinforce cognitive autonomy and critical skills.

In this perspective, these findings align with the objectives of the ESRI Pact 2030 and the *Confédération Générale des Entreprises du Maroc Livre blanc,* suggesting that generative AI, when framed within a pedagogical context, represents not a replacement for human thinking but a pathway for cooperative learning based on responsibility, creativity, and social intelligence. From this perspective, the Rabat situation exemplifies the promise of a pedagogy of intelligence based on clarity, reflexivity, and the art of making meaning.

## 5. Contributions and Limitations

This study introduces GenAI into Moroccan higher education, particularly within scientific FLE contexts. Rather than replacing human learning, it emphasizes augmentation and proposes a framework for evaluating GenAI integration based on three dimensions: personalization, digital literacy, and pedagogy. Using a quantitative questionnaire, as well as observation and interviews, this research helps in understanding how user perceptions of competencies evaluate different learning practices, denoting various possibilities between texts produced without AI and texts produced with AI assistance. Lastly, biased in the result, this study underscores the importance of a "distributive" role of the university, using the faculty of science in Rabat's structure to illustrate the concept of transparency and co-responsibility in governance. However, it also has limitations; the sample is not representative of the wider Moroccan university system, meaning generalization is not realistic. Another limitation is the one-semester sample, which does not allow long-term insights, especially on reflective competencies in women. Future research reviews participant-instructor relationships and instructors' AI literacy. Nevertheless, a shift in university roles toward collaborative learning spaces rather than simple knowledge transmission could foster the development of critical thinking and intellectual autonomy. The student is transformed by the possibilities that come from learner-centered GenAI and its ethical implications, teaching edified through reflection for equity and agency. Future research could further examine the relationship between students and instructors, as well as instructors' AI literacy.

## References

Cavenaghi, U., & Senécal, I. (2019). *Osons l'IA à l'école: Préparons nos jeunes à la révolution de l'intelligence artificielle.* Éditions Château d'encre.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.

Confédération générale des entreprises du Maroc (CGEM). (2021). *Livre blanc : Vers une croissance économique soutenue, responsable et durable*. CGEM.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume.* Council of Europe Publishing. https://www.coe.int/lang-cefr

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Cristol, D. (2019). *Former, se former et apprendre à l'ère numérique: Le social learning.* Éditions ESF Sciences Humaines.

Cristol, D. (2024). *Apprendre à l'ère de l'intelligence artificielle : Révolution, défis et opportunités*. ESF Sciences humaines.

Crompton, H., & Burke, D. (Eds.). (2025). *Artificial intelligence applications in higher education: Theories, ethics, and case studies for universities*. Routledge. Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 1, pp. 416–437). SAGE. https://doi.org/10.4135/9781446249215.n21

Desveaud, K. (2024). *L'intelligence artificielle décryptée : comprendre les enjeux et risques éthiques de l'IA pour mieux l'appréhender*. EMS Éditions.

Elmouhtadi, A., Shlaka, S., & Soughati, N. (2025) L'intelligence artificielle générative (IAg) dans l'apprentissage du FLE chez l'étudiante et l'étudiant marocains : Ètude de cas de la Faculté des sciences de Rabat. *Revue internationale des technologies en pédagogie universitaire*, *22*(1), Article 16. https://doi.org/10.18162/ritpu-2025-v22n1-16

Forum IA Québec, & Institut Intelligence et données. (2022). *Cadre et outil d'évaluation de la maturité en matière d'intelligence artificielle: Votre organisation est-elle prête pour l'IA? Une adaptation du Cadre de maturité de l'IA d'Element AI*. Forum IA Québec & Institut Intelligence et données (IID).Holec, H. (1981). *Autonomy and foreign language learning.* Pergamon Press.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Commission, 57* (4), 542-570. https://doi.org/10.1111/ejed.12533

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf

Jabagi, N., & Croteau, A.-M. (2025). L'intelligence artificielle (IA): amie ou ennemie de la motivation des étudiants et étudiantes universitaires?. *Revue internationale des technologies en pédagogie universitaire*, *22*(1), Article 4. https://doi.org/10.18162/ritpu-2025-v22n1-04

Jong, M. S.-Y., Geng, J., Chai, C. S., & Lin, P.-Y. (2020). Development and predictive validity of the computational thinking disposition questionnaire. *Sustainability*, *12*(11), Article 4459. https://doi.org/10.3390/su12114459

Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, *67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Kurban, C. F., & Şahin, M. (2024). *The impact of ChatGPT on higher education: Exploring the AI revolution.* Emerald Publishing. https://doi.org/10.1108/9781837976478

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. *Pearson* Education.Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Miao, F., Shiohira, K., & Lao, N.. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105

Miles, M. B., Huberman, A. M., & Saldanña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications. https://www.metodos.work/wp-content/uploads/2024/01/Qualitative-Data-Analysis.pdf

Ministère de l'Enseignement Supérieur, de la Recherche Scientifique et de l'Innovation. (2022). *Plan national d'accélération de la transformation de l'écosystème de l'enseignement supérieur, de la recherche scientifique et de l'innovation: Pacte ESRI 2030*. Royaume du Maroc.

Naffi, N., & Montufar, C. S. (2025). Autoformation en outils d'intelligence artificielle générative : un levier pour guider et optimiser la conception pédagogique. *Revue internationale des technologies en pédagogie universitaire, 22*(1), Article 5. https://doi.org/10.18162/ritpu-2025-v22n1-05

OECD. (2021). *OECD Digital Education Outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots.* OECD Publishing. https://doi.org/10.1787/589b283f-en

Perrin, N., Piot, D., Vita, L., Bationo-Tillon, A., & Guibourdenche, J. (2025). Des hypothèses pour concevoir des tâches permettant aux étudiants et étudiantes d'évaluer la pertinence des textes générés par les IA. *Revue internationale des technologies en pédagogie universitaire*, *22*(1), Article 6. https://doi.org/10.18162/ritpu-2025-v22n1-06

Romero, M., Heiser, L., & Lepage, A. (Eds.). (2023). *Enseigner et apprendre à l'ère de l'intelligence artificielle*: *Acculturation, intégration et usages créatifs de l'IA en éducation* (livre blanc). Canopé. https://hal.science/hal-04013223

Roy, N., Proust-Androwkha, S., Gruslin, É., Vallerand, V., & Charles, É. (2025). L'intelligence artificielle au postsecondaire: entre enthousiasme et méfiance – Introduction au numéro thématique. *Revue internationale des technologies en pédagogie universitaire*, *22*(1), Article 1. https://doi.org/10.18162/ritpu-2025-v22n1-01

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford Press.

Seldon, A., & Abidoye, O. (2018). *The fourth education revolution Will artificial intelligence liberate or infantilise humanity*. Legend Press.

Shah, P. (2023). *AI and the future of education: Teaching in the age of artificial intelligence*. John Wiley & Sons.

Stiegler, B. (2005). Individuation et grammatisation: quand la technique fait sens... *Documentaliste-Sciences de l'Information, 42* (6), 354-360. https://doi.org/10.3917/docsi.426.0354

Stiegler, B. (2016). *Digital Studies Organologie des savoirs et technologies de la connaissance* FYP éditions.

Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: *Artificial Intelligence*, 7, Article 100259. https://doi.org/10.1016/j.caeai.2024.100259

UNESCO. (2019). *Beijing Consensus on Artificial Intelligence and Education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

# HUMAN AND ARTIFICIAL INTELLIGENCE PERSPECTIVES ON THE IMPORTANCE OF FEEDBACK AND PEDAGOGICAL REFLECTION FOR IMPROVING TEACHERS' PROFESSIONAL PERFORMANCE

**Anastasia-Mălina Cioloca**

*Transilvania University,*
*Faculty of Psychology and Educational Sciences,*
*Brasov, Romania*

**ABSTRACT**

In today's society, the teaching profession has a transformative perspective, according to the OECD (2019), encouraging teachers to analyze their professional performance using metacognitive skills, which integrate feedback into pedagogical decisions and adjust their behavior based on findings, theories, and suggestions received.

Starting from the formative value of feedback in reflective teaching practice, this study, triangulated from a methodological point of view, aimed to investigate through focus groups the perceptions of 36 students (future teachers) regarding the role and influence of feedback on their professional performance and the quality of dialogues between teachers and the Artificial Intelligence system, ChatGPT. These dialogues were constructed based on a specific prompt that integrated feedback messages given by students to teachers and whose main goal was to improve professional performance. Most students stated that the suggestions offered by ChatGPT for improving professional performance were pedagogically contextualized and were perceived as useful and applicable.

On the other hand, this study also conducted a content analysis between a teaching self-assessment form completed by a teacher and a teaching self-assessment form generated by ChatGPT, noting their particularities in terms of similarities and differences.

***Keywords:*** *Feedback, Reflexivity, Self-analysis of teaching, Improving teachers' professional performance, ChatGPT*

## 1. Introduction

In the current educational context, characterized by dynamism and continuous change, there is a need for professional adaptability among teachers. The strategic vision document for the development of education for the years 2014-2020, "Education-2020" (Official Gazette, No.

345-351, 2014), mentions a specific objective 2.4. Modernization of the university curriculum with a focus on skills, learners, and the needs of the economic environment. Focusing on the learner requires sustained reflection and self-analysis on the part of the teacher, as well as an openness to receive and subsequently integrate feedback from the learner into their own professional performance. Also, integrating digital tools can help teachers develop reflective and contextualized thinking in relation to students' needs and expectations.

The paradigms of reflection, reflexivity, and interpretation (hermeneutics) become crucial in pedagogical knowledge. Joița (2009) defines reflection as a metacognitive strategy through which individuals can analyze their thought processes through monitoring, systematic self-assessment, and critical evaluation. Through reflection, previous experiences can be capitalized on through an action research approach based on interpretations, premises, and rational arguments. This individual capacity for reflection expands and consolidates within a reflexive learning environment. This concept was defined by Denha et al. (2022) as a structured set of pedagogical and psychosocial conditions designed to support the processes of self-analysis, self-correction, self-determination, and self-realization of future specialists by cultivating the capacity for reflection.

Self-analysis of teaching activity is a self-reflective and self-critical process through which teachers reflect cognitively and metacognitively on the design, organization, and conduct of their activity, performing self-assessments and value judgments regarding the achievement of operational objectives. Awareness of successes, difficulties, and their causes, whether internal or external, allows for the correction of future actions and supports the development of pedagogical skills, especially self-evaluative ones, thus contributing to self-training and increased educational efficiency.

According to Day et al. (2008), teaching effectiveness encompasses three dimensions: creating a favorable classroom climate, classroom management, and adopting an organized and clear teaching approach. Other studies have shown that effective teachers tend to maintain high levels of student engagement (Good & Brophy, 2003), and as a result, a correlation has been identified between teacher effectiveness and student learning (Fackler & Malmberg, 2016). However, in order to self-monitor their teaching effectiveness, teachers should optimize their educational strategies and improve their professional performance based on feedback from various educational agents (students, other colleagues, principals, etc.).

Sadler (1989) stated that the main purpose of feedback is to reduce discrepancies between current understanding and performance and is a goal. Feedback in education has been a priority over the last century, and research strongly supports that this process, when designed appropriately, can have a powerful effect on improving performance (Hattie & Timperley, 2007; Ilgen et al., 1979; Kluger & DeNisi, 1996; Remmers & Brandenburg, 1927). In higher education, specialist studies report the positive effect of the feedback process provided by students to teachers on the quality of teaching (Cohen, 1980; L'Hommedieu et al., 1990; Penny & Coe, 2004). Such feedback is collected through Student Evaluation of Teaching (SET) being considered an integral part of performance evaluation (Lang & Kersting, 2007; Penny & Coe, 2004). Based on this idea, a first research question is: *What is the perception of students (future teachers) regarding the role and influence of relevant feedback given by learners to teachers on their professional performance?* To answer this question, we designed and applied a focus group grid for students (future teachers) participating in the study to investigate their perception and personal opinion of feedback as a predictor of improved professional performance of teachers.

Given the importance of feedback in the educational process, the use of modern technologies can contribute to its implementation in pedagogical decisions and teaching approaches tailored to improving professional performance. However, although the literature has analyzed the distinct effects of different forms of feedback, there remains a research gap regarding the synergistic impact of combining AI-based tools, such as ChatGPT, with traditional feedback provided to teachers by students, in order to optimize professional performance. That is why I wondered *whether ChatGPT can understand and interpret complex teaching contexts in order to generate solutions for improving teachers' professional performance within the human-AI conversation. More specifically, how useful and implementable are the suggestions offered by ChatGPT for improving teachers' professional performance, as perceived by the students (future teachers) participating in the study?* As a result, we investigated the perceptions of students (future teachers) regarding the ability of the ChatGPT Artificial Intelligence (AI) system to generate useful and implementable solutions for improving the professional performance of teachers.

The accelerated development of educational technologies has created favorable conditions for the effective integration of digital tools into the teaching-learning process (Hyland & Hyland, 2006). In this context, Karataş and Yüce (2024) emphasize that the rapid advancement of AI in education offers significant opportunities for transforming learning environments. The use of AI in education reveals considerable potential for increasing the efficiency of teaching processes, particularly in terms of the professional development of teachers.

A key aspect is the ability to identify teaching performance in real time through automated AI-based feedback systems, which allow teachers to adjust their teaching strategies based on evidence-based recommendations (Luckin, 2018). In order to provide personalized learning experiences and automated feedback, AI is increasingly being used in teacher training (Zawacki-Richter et al., 2019). AI-based platforms can also analyze teachers' strengths and weaknesses, tailoring training materials to individual needs, which helps to increase engagement and the effectiveness of the learning process (Luckin, 2017).

The integration of AI-based tools such as ChatGPT requires a profound rethinking of traditional pedagogical paradigms and a reassessment of the essential skills for effective teaching in the 21st century (Al-khresheh, 2024; Costello, 2024; Nguyen Thanh et al., 2023; Trust et al., 2023). This transformation involves a careful analysis of how ChatGPT can support the professional development of teachers, facilitating the adaptation of educational strategies to the needs and expectations of learners. The ChatGPT AI system can be used not only to support the continuous training of teachers and improve pedagogical practices (Adeshola & Adepoju, 2024; ElSayary, 2024; Javaid et al., 2023; Siminto et al., 2023; Zhang & Tur, 2024), but also for providing real-time feedback, thus contributing to the personalization of the learning process and increasing its efficiency (Alshahrani, 2023; Gao, 2021; Jauhiainen & Guerra, 2023; Lo, 2023; Rasul et al., 2023; Roscoe, et al. 2018; Zawacki-Richter, et al. 2019). Its advanced capabilities—such as generating coherent content, answering complex questions, and maintaining meaningful conversations—have attracted the attention of the academic community, generating increased interest in its applicability in educational settings (Albadarin et al., 2024; Borger et al., 2023). Based on the above theoretical considerations, we asked the question: *What are the differences and similarities between a teaching self-analysis form completed by a teacher and one generated by ChatGPT in terms of structure, degree of professional reflection, focus on the learner, and other psycho-pedagogical aspects?* Thus, we conducted and interpreted a comparative content analysis of two teaching self-analysis sheets, one created by the researcher and one generated by ChatGPT.

As education adapts to new digital realities, harmonizing feedback from learners, systematic reflection through tools such as teaching self-assessment forms, and the use of AI can be a prerequisite for improving teacher performance.

## 2. Materials and Methods

This study is based on preliminary data obtained in a previous study (Voinea et al., 2025), in which a questionnaire ($\alpha = 0.94$ for 22 items; $M = 3.61$; $SD = 0.67$) was distributed to 327 respondents – students undergoing initial training for a teaching career. The results highlighted an essential aspect of understanding reflective teaching practice: feedback from students is an important resource for the professional development of future teachers. In this regard, it was found that, to a large extent (35%) and to a very large extent (25.5%), teachers request feedback from students at the end of teaching activities; however, there are few situations in which professional behavior or teaching strategies are adjusted based on the feedback received.

For this study, a sample of 36 students was used, from the first year (18) and second year (18), enrolled in the Faculty of Psychology and Education Sciences, who are preparing to become teachers. Convenience sampling was applied, with all participants being female. Methodologically, we used the focus group method, based on the research tool, the focus group grid, and the content analysis method, using the research tool, the content analysis grid. Both tools were developed in-house, in line with the research objective.

Within the discussion group, according to Wilkinson (2004), data were collected from a small number of participants through an informal, collective discussion. To conduct the research, institutional approval was obtained, and respondents gave their written consent to participate in the research and to the processing of personal data, which was anonymized. The students were divided into six groups of six people, according to the methodological recommendations for focus groups. The focus group grid contains four open-ended questions, one exploratory question, one dual-response question, and one question expressing the intensity of agreement on a Likert scale. The focus group grid dimensions were the role of feedback in the educational process, the impact of student feedback on teachers, students' ability to provide feedback, feedback messages provided and tested by students in ChatGPT, students' perception of the usefulness of ChatGPT suggestions, and students' perception of the applicability of ChatGPT suggestions.

For data analysis, the discussions were transcribed in full. Then, the initial coding of the units of meaning took place, using MAXQDA 24 software, and finally the emerging themes were identified. To ensure the reliability of the coding, 20% of the data was also coded by a second researcher, also from the field of education sciences, using the same coding procedure but without involvement in the development of the theoretical framework. Regarding the degree of agreement between the two coders, discrepancies were discussed until a final consensus was reached. For content analysis, we identified significant codes from the texts of two self-analysis sheets on teachers' teaching activities, one developed in-house and the other generated by ChatGPT. The teaching activity self-analysis form designed by the teacher was based on specialized literature and their professional expertise, as well as constant self-reflection on the teaching process. The teaching activity self-analysis form proposed by the ChatGPT AI system, version GPT-5. 3 (2025), was generated using the prompt: "Generate, in an elaborate manner, a teaching activity self-analysis form that can be used by teachers in the pre-university system." In order to perform the content analysis, seven categories of analysis were structured (structure of the self-analysis form, degree of professional reflection, focus on the student, use of teaching strategies, evaluation of one's own performance, management elements, professional values and attitudes), followed by specific indicators of behavior achievement.

For each indicator, we assigned points on a Likert scale from 1 to 5, where 1 means that the indicator was not achieved and 5 means that the indicator was fully achieved. These points were accumulated and, for each category of analysis, a final score was obtained in relation to the two files analyzed. To ensure data reliability, the content analysis was also performed by a second researcher who was not involved in this study and had previously contributed to the coding of participants' responses.

## 3. Results and Analysis

The corpus analyzed in the MAXQDA statistics program included 36 documents, one specific document for each study participant, in which their answers to the focus group questions were transcribed in full. In the statistical analysis, the segment is defined as the minimum unit of text that expresses a complete idea relevant to the dimension analyzed, and the percentages reported in the analysis represent the proportion of segments coded with a specific code out of the total number of coded segments. During the focus group, there were participants who did not answer all the questions, an aspect also found in the scope of the research.

First, I was interested in qualitatively analyzing students' perceptions of the role of feedback in the educational process. The analysis of participants' responses can be found in Table 1.

**Table 1**

*Students' perceptions of the role of feedback in the educational process.*

| CATEGORY: THE ROLE OF FEEDBACK IN THE EDUCATIONAL PROCESS | | | |
|---|---|---|---|
| **Codes** | **Segments** | **Percentage** | **Examples of content** |
| Improving job performance | 18 | 32,1 | “One of the most important roles of feedback is to improve student achievement and at the same time improve the applicability of the methods used by the teacher.” |
| Adjusting teaching/learning behaviors | 11 | 19,6 | “For the students, feedback helps them to perform, to get the results they want, and for the teacher to adjust their teaching approach.” |
| Self-evaluation of teaching activity | 8 | 14,3 | “From my point of view, feedback is very important to observe the teacher's journey in the educational process.” |
| Regulating the educational process | 8 | 14,3 | “Feedback is a regulator for both students and teachers, both of whom can see problems in the process and can remedy them.” |
| Motivating educational agents | 4 | 7,1 | “To stimulate yourself and know what you need to change.” |
| Guiding the teaching-learning process | 3 | 5,4 | “The role of feedback in the educational process is to guide teachers/students to improve their work.” |
| Improving the teacher-student relationship | 2 | 3,6 | “I think it is important too, because both the teacher and the pupil have to learn from each other, and through feedback you can better understand communication” |

As can be seen, participants perceive the major role of feedback in the educational process differently, with most (32.1%) believing that feedback can improve the professional performance of teachers.

Based on these statements, we further investigated the extent to which feedback given by students to teachers has a significant impact on adjusting professional behavior (see Table 2).

**Table 2**

*Students' perceptions of the impact of student feedback on teaching staff*

| CATEGORY: THE IMPACT OF STUDENT FEEDBACK TO TEACHERS | | | |
|---|---|---|---|
| **Codes** | **Segments** | **Percentage** | **Factors influencing responses** |
| Lack of significant impact | 41 | 67,2 | students' fear of giving feedback to teachers; teachers' perception of the relevance of feedback given by students; teachers' complacency in previous professional approaches |
| Existence of a significant impact | 20 | 32,8 | anonymization of feedback; more efficient teaching, clarity and interactivity in courses. |

Most students (67.2%) perceived this impact as insignificant, as they are reluctant to give honest feedback and fear the consequences that such feedback might have. Participants also stated that some teachers do not perceive the feedback provided by students as relevant, especially if their pedagogical approach has been effective for previous generations.

On the other hand, it was necessary to explore students' perceptions of their current feedback skills and their needs for training in this skill. The analysed data can be found in Table 3.

**Table 3**

*Students' perceptions of their ability to give feedback and the need to develop this skill*

| CATEGORY: STUDENT'S ABILITY TO GIVE FEEDBACK | | | |
|---|---|---|---|
| **Codes** | **Segments** | **Percentage** | **Examples of content** |
| Students indicate measures for improvement in giving feedback | 24 | 48 | "It would be helpful if feedback was given anonymously" "I'm thinking that maybe it would be good to have a course where we are systematically taught how to give feedback" |
| Barriers for students in giving feedback | 17 | 34 | "I think that some of us are really not ready, because we are afraid to express ourselves in front of a teacher" "I think we are not ready, because we have not been asked before." |
| Students know how to give feedback | 9 | 18 | "The students are sufficiently prepared to give constructive feedback because they know how they want the lessons to run and how the teacher should adapt to their needs" |

Most participants (48%) report a need to develop their feedback skills through courses or thematic modules that teach them how to give effective feedback, while others suggest that anonymous feedback would be a suitable practice. There are also students (34%) who report certain barriers to giving feedback to teachers, such as fear of expression and lack of training in giving feedback. Also, 18% of the students believe that they know how to give feedback, being sufficiently mature, responsible, and involved in the educational process.

To explore whether the ChatGPT AI system can analyze and interpret complex teaching contexts, offering solutions within the human-AI interaction aimed at improving the professional performance of teachers, we invited students to reflect on the last feedback they gave to a teacher. These feedback messages covered various topics, as presented in Table 4.

**Table 4**

*The latest feedback given by participants to a teacher*

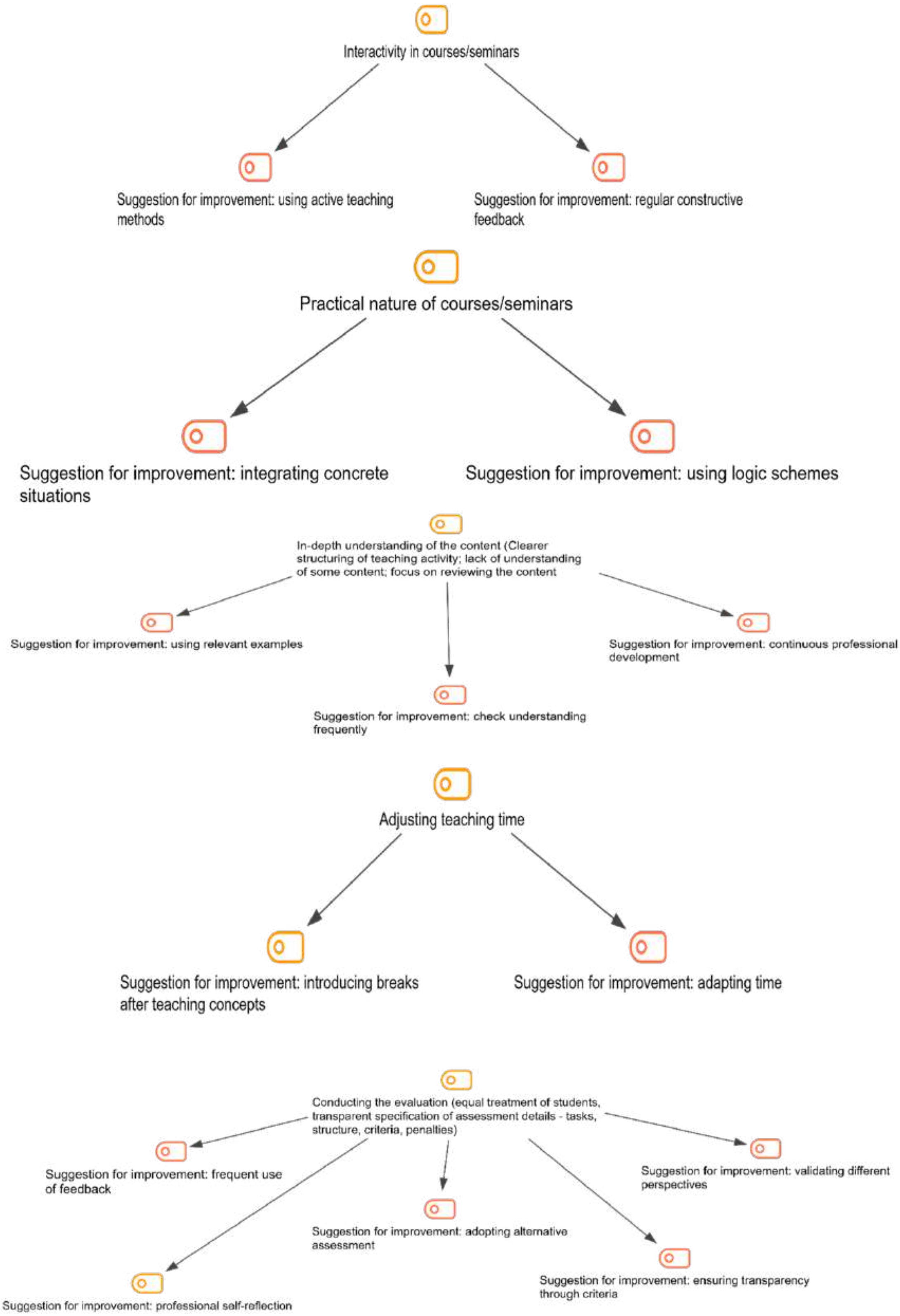
| CATEGORY: FEEDBACK MESSAGES GIVEN AND TESTED BY STUDENTS IN CHATGPT | | |
|---|---|---|
| **Codes** | **Segments** | **Percentage** |
| Conducting the evaluation | 37 | 67,3 |
| Thorough understanding of content | 6 | 10,9 |
| Adjusting teaching time | 4 | 7,3 |
| Practical nature of courses/seminars | 4 | 7,3 |
| Interactivity in courses/seminars | 4 | 7,3 |

As shown most participants (37%) reported that the most significant feedback messages were those related to teaching evaluation, such as the need for fair and ethical treatment of students and transparent character assessments.

In the next stage of the focus group, we proposed to the students that we use the ChatGPT AI system, based on the GPT-5.3 (2025) model, using a single prompt to generate content, formulated from the perspective of a teacher receiving feedback from their students: "The students asked me to... (this section is completed by each participant with the feedback message reflected in the previous question). How can I improve my professional performance?". Figure 1 shows suggestions for improving the professional performance of teachers, based on the feedback message entered.

**Figure 1**

*Suggestions for improving teachers' professional performance, generated by ChatGPT*



Based on these suggestions for improving teachers' professional performance, we investigated participants' perceptions of the usefulness and applicability of the recommendations provided by AI. The procedure for interpreting data in this context and explaining the concepts of usefulness and applicability of the recommendations provided by ChatGPT is structured in Table 5.

**Table 5**

*Defining the criteria for the usefulness and applicability of the recommendations provided by ChatGPT*

| CONCEPT | HOW IS IT DEFINED? | HOW IS IT EVALUATED? |
|---|---|---|
| **USEFULNESS** | The recommendations provide relevant information | Based on the Likert scale, where 1=not at all, 2=to a small extent, 3=moderately, 4=to a large extent, 5=completely |
| **APPLICABILITY** | The recommendations can be used in practice | Based on the Likert scale, where 1=not at all, 2=to a small extent, 3=moderately, 4=to a large extent, 5=completely |

The results regarding how students reported on the usefulness and applicability of these dialogues are presented in Tables 6 and 7.

**Table 6**

*Students' perception of the usefulness of suggestions generated by ChatGPT*

| CATEGORY: STUDENTS' PERCEPTION OF THE APPLICABILITY OF CHATGPT'S SUGGESTIONS | | |
|---|---|---|
| **Codes** | **Segments** | **Percentage** |
| The ChatGPT's suggestions are applicable to the described teaching context | 34 | 11,3 |
| The ChatGPT's suggestions are not applicable to the described teaching context | 2 | 0,7 |

The results show that most students consider these suggestions to be useful (18%) and very useful (12%).

**Table 7**

*Students' perception of the applicability of suggestions generated by ChatGPT*

| CATEGORY: STUDENTS' PERCEPTION OF THE USEFULNESS OF CHATGPT'S SUGGESTIONS | | |
|---|---|---|
| **Codes** | **Segments** | **Percentage** |
| ChatGPT's suggestions for improving performance are HELPFUL | 18 | 6 |
| ChatGPT's suggestions for improving performance are VERY HELPFUL | 12 | 4 |
| ChatGPT's suggestions for improving performance are MODERATELY HELPFUL | 6 | 2 |

Of those stated by participants, there are a few suggestions offered by ChatGPT that they perceive as inapplicable.

In the next stage of the research, we performed a content analysis of the two teaching self-assessment forms, as described in the methodology. The categories of analysis used, and the achievement indicators are described from Tables 8 to 14.

**Table 8**

*Category analyzed: Structure of the self-analysis form*

| **Performance indicators** | **Self-analysis sheet (teacher's concept) scores** | **Self-analysis sheet (according to ChatGPT) scores** |
|---|---|---|
| **8a.** Clarity in terms of language and expression | **5** | **3** |
| **8b.** Systematic organization of sections | **5** | **3** |
| **8c.** Mentions regarding all stages of the teaching process | **5** | **5** |
| **8d.** The presence of explicit criteria | **5** | **4** |
| **8e.** Variety of question types | **4** | **5** |
| **Total** | **24** | **20** |

**Table 9**

*Category analyzed: Degree of professional reflection*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **9a.** Introspective implications | **4** | Thoughts, awareness, anticipation | **4** | Improvements, reflections |
| **9b.** Concrete examples | **4** | Activities | **5** | Examples |
| **9c.** Use an impersonal, honest tone | **5** | Consulting curriculum documents | **2** | I succeeded, I captured, etc. |
| **Total** | | **13** | | **11** |

**Table 10**

*Category analyzed: Focus on educability*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **10a.** Notes on the educational needs of the learner | **5** | Individualization, educational needs | **1** | - |
| **10b.** Notes on integrating feedback into the teaching process | **5** | Feedback | **5** | Feedback |
| **10c.** Notes on school inclusion | **5** | Differentiation and individualization | **1** | - |
| **10d.** Comments on the relevance of curriculum content for students | **5** | Relevant content, systematic connections | **1** | - |
| **Total** | | **20** | | **8** |

**Table 11**

*Category analyzed: Leveraging teaching strategies*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **11a.** Adaptability of strategies to context | **5** | Relating to the real world | **5** | Methods appropriate to the content |
| **11b.** Use of interactive/innovative teaching methods and tools | **5** | Effective methods | **5** | Varied, appropriate methods; functional resources |
| **Total** | | **10** | | **10** |

**Table 12**

*Category analyzed: Evaluating your own performance*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **12a.** Identifying strengths and weaknesses | **3** | Overall assessment | **5** | Positive aspects, difficulties encountered |
| **12b.** The existence of self-assessment indicators | **5** | Self-assessment, overall assessment | **1** | - |
| **12c.** Identification of improvement measures | **1** | - | **4** | Suggestions for improvement |
| **Total** | | **9** | | **10** |

**Table 13**

*Category analyzed: Management elements*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **13a.** The existence of classroom management elements | **5** | Activation, the principle of intuition | **4** | Active participation |
| **13b.** The existence of time management elements | **5** | Time estimation | **2** | Pace of the lesson |
| **13c.** Correlation with professional standards | **4** | Curriculum documents | **1** | Curriculum documents |
| **Total** | | **14** | | **7** |

**Table 14**

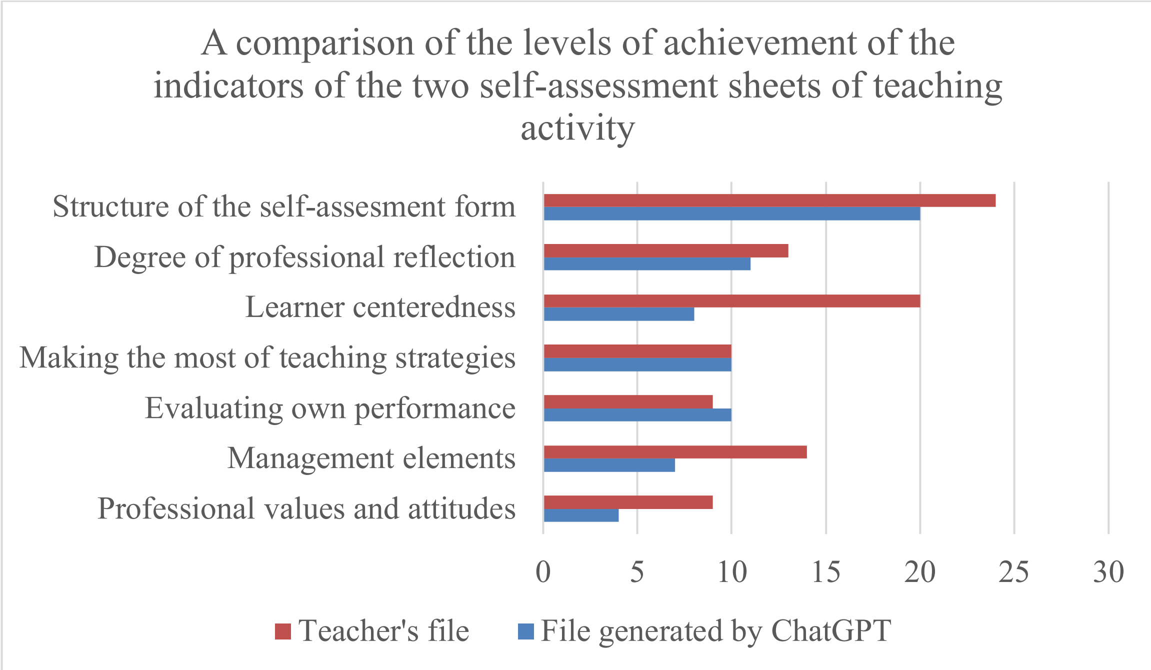
*Category analyzed: Professional values and attitudes*

| Performance indicators | Self-analysis sheet (teacher's concept) scores and codes | | Self-analysis sheet (according to ChatGPT) scores and codes | |
|---|---|---|---|---|
| **14a.** Ethics and dynamics of relationships | **5** | Conduct, interactions, ethics, exchange of ideas, tolerance, collaboration | **2** | Atmosphere, attention |
| **14b.** Professional motivation | **4** | Positive attitude, self-improvement | **2** | Improvements |
| **Total** | **9** | | **4** | |

Next, to identify the differences and similarities between a teaching self-assessment form completed by a teacher based on their expertise and one generated by ChatGPT in terms of various psycho-pedagogical aspects, we compared the levels of achievement of the indicators in the two teaching self-assessment forms (see Figure 2).

**Figure 2**

*The comparison between the levels of achievement of the indicators in the two teaching self-assessment forms*




As a result, what we found is that the teaching self-analysis sheet generated by ChatGPT presents an introspective and open approach aimed at contextualizing the personal development of teachers. This sheet has a strong reflective, introspective tone, which encourages critical thinking about one's own practice and emphasizes the impact of the teacher's actions on the student. Self-analysis is more of a story about evolution.

The teaching self-analysis form designed by a teacher, the researcher himself, has a systematic, rigorous approach that can be applied in a formal, institutionalized context. It is written in professional terminology, which encourages professional reflection and makes explicit references to curriculum documents. The approach in this document is pragmatic and student-centered, with self-analysis using clear criteria and indicators.

Both self-analysis forms contain important evaluative aspects, namely: the use of modern, interactive, or innovative methods and the existence of classroom management elements.

## 4. Discussion and Conclusions

In line with the first research question, namely, what is the perception of students (future teachers) regarding the role and influence of relevant feedback provided by students to teachers on their professional performance, feedback is a predictor of improved professional performance, in the perception of participants, but nevertheless, it can only become meaningful and produce behavioral changes if it is valid and accepted by the teacher Röhl et al. (2021). As a result, it is necessary to reconsider the normative procedures for providing feedback, and it would also be advisable for students to take certain courses/modules to develop their ability to provide feedback.

On the other hand, in line with the second research question, namely, can ChatGPT understand and interpret complex teaching contexts to generate solutions for improving teachers' professional performance in the conversation between humans and AI, the ChatGPT AI system can be useful, as it offers applicable and contextualized suggestions for improving professional performance, in the students' perception.

In line with the third research question, what are the differences and similarities between a teaching self-analysis form completed by a teacher and one generated by ChatGPT in terms of structure, degree of professional reflection, student focus, and other psycho-pedagogical aspects, it was found that, with regard to the coherence, accuracy, and scientific nature of the generation of professional documents, the intervention of a human specialist in the field is necessary. At present, AI excessively standardizes pedagogical discourse, imposing generic formulations that are often decontextualized and not tailored to the needs of students and teachers. In the context of using regulatory academic documents generated by ChatGPT, such as the teaching activity self-analysis sheet, certain risks can also be highlighted: diminishing the authenticity of professional reflection, lack of pedagogical contextualization, violation of principles of professional integrity, and technological dependence in the teaching-learning process.

## 5. Contributions and Limitations

This study presents an atypical, relevant, and extremely modern research issue in the current educational context, from a theoretical-fundamental and practical-applicative perspective. The limitations of this study are the small sample size and the limited generalizability of the results to a larger population. At the same time, some incomplete and inadequate responses to the questions asked were identified, which impacted the complete analysis of the data.

As future research directions, I propose expanding the sample and diversifying the profile of participants, developing an integrated hybrid feedback model (human + AI) that includes student feedback, teacher self-analysis, and recommendations generated by an AI system, and conducting longitudinal research on AI-assisted reflective training.

## References

Adeshola, I., & Adepoju, A. P. (2024). The Opportunities and Challenges of ChatGPT in Education. *Interactive Learning Environments*, 32 (10), 6159 – 6172. https://doi.org/10.1080/10494820.2023.2253858

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence*, 6, 100218. https://doi.org/10.1016/j.caeai.2024.100218

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, *3,* 60. https://doi.org/10.1007/s44217-024-00138-2

Alshahrani, A. (2023). The impact of ChatGPT on blended learning: Current trends and future research directions. *International Journal of Data and Network Science*, *7*, 2029–2040. https://doi.org/10.5267/j.ijdns.2023.6.010

Borger, J. G., Ng, A. P., Anderton, H., Ashdown, G. W., Auld, M., Blewitt, M. E., Brown, D. V., Call, M. J., Collins, P., Freytag, S., Harrison, L. C., Hesping, E., Hoysted, J., Johnston, A., McInneny, A., Tang, P., Whitehead, L., Jex, A., & Naik, S. H. (2023). Artificial intelligence takes center stage: Exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. *Immunology & Cell Biology, 101*, 923–935. https://doi.org/10.1111/imcb.12689

Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin, 88*(1), 82–108. https://doi.org/10.1037/0033-2909.88.1.82

Costello, E. (2024). ChatGPT and the educational AI chatter: Full of bullshit or trying to tell us something? *Postdigital Science and Education*, *6*, 425–430. https://doi.org/10.1007/s42438-023-00398-5

Day, C., Sammons, P., & Kington, A. (2008). *Effective classroom practice (ECP): A mixed-method study of influences and outcomes*. University of Nottingham/Economic and Social Research Social.

Denha, N., Kyrylenko, K., Syvokhop, E., Yermakova, S., Stepanchenko, N., & Bilanych, H. (2022). Methodical work as a means of developing professional reflection of teachers. *Revista Românească pentru Educație Multidimensională*, *14*(2), 363–392. https://doi.org/10.18662/rrem/14.2/586

ElSayary, A. (2024). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. Jo*urnal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

Fackler, S., & Malmberg, L.-E. (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effect. *Teaching and Teacher Education*, *56*, 185–195. https://doi.org/10.1016/j.tate.2016.03.002

Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system Pigai. *International Journal of Emerging Technologies in Learning ,16* (11), 322–330. https://doi.org/10.3991/ijet.v16i11.19657

Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms* (9th ed.). Allyn & Bacon.

Government of the Republic of Moldova. (2014). *Strategia de dezvoltare a educației pentru anii 2014–2020 „Educația-2020"* [Education Development Strategy for 2014–2020 "Education-2020"] (Government Decision No. 944, November 14, 2014). *Monitorul Oficial al Republicii Moldova* [Official Gazette of the Republic of Moldova], 345–351, Art. 1014.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hyland, K., & Hyland, F. (2006) Feedback on second language students' writing. *Language Teaching*, *39*(2), 83-101. https://doi.org/10.1017/s0261444806003399

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349–371. https://doi.org/10.1037/0021-9010.64.4.349

Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. *Sustainability*, *15*(18), 14025. https://doi.org/10.3390/su151814025

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT tool towards ameliorating the education system. *Bench Council Transactions on Benchmarks. Standards and Evaluations*, *3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Joița, E. (2009). *Știința educației prin paradigme.* Institutul European.

Karataş, F., & Yüce, E. (2024). AI and the Future of Teaching: Preservice Teachers' Reflections on the Use of Artificial Intelligence in Open and Distributed Learning. *The International Review of Research in Open and Distributed Learning, 25*(3), 304–325. https://doi.org/10.19173/irrodl.v25i3.7785

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

L'Hommedieu, R., Menges, R. J., & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. Journal of Educational Psychology, *82*(2), 232–241. https://doi.org/10.1037/0022-0663.82.2.232

Lang, J.W.B., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*(3), 187–205. https://doi.org/10.1007/s11251-006-9006-1

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1*, 0028. https://doi.org/10.1038/s41562-016-0028

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL Institute of Education Press.

Nguyen Thanh, B., Vo, D. T. H., Nguyen Nhat, M., Pham, T. T. T., Thai Trung, H., & Ha Xuan, S. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, 39(5), 59–81. https://doi.org/10.14742/ajet.8902

OECD. (2019). *OECD future of education and skills 2030: OECD learning compass 2030* (Romanian Translation). OECD.

Penny, A.R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis, *Review of Educational Research, 74*(2), 215-253. https://doi.org/10.3102/00346543074002215

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. D. O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, 6(1). https://doi.org/10.37074/jalt.2023.6.1.29

Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instructors. Educational Administration and Supervision, 13(6), 399-406.

Röhl, S., Bijlsma, H., & Rollett, W. (2021). The process model of student feedback on teaching (SFT): A theoretical framework and introductory remarks. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers* (pp. 1–11). Springer.https://doi.org/10.1007/978-3-030-75150-0_1

Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated Writing Instruction and Feedback: Instructional Mode, Attitudes, and Revising. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, (Vol. 62*(1), pp.2089–2093). https://doi.org/10.1177/1541931218621471

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science, 18* (2), 119–144. http://dx.doi.org/10.1007/BF00117714

Siminto, Lisnawati, S. D., & Muharam, S. (2023). Teacher professionalism development strategy through ChatGPT support in the context of education management. *Journal of Contemporary Administration and Management (ADMAN), 1*(3), 150–155. https://doi.org/10.61100/adman.v1i3.65

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher*

*Education*, *23*(1), 1-23. https://citejournal.org/volume-23/issue-1-23/editorial/editorial-chatgpt-challenges-opportunities-and-implications-for-teacher-education/

Voinea, M., Cioloca, A. M., & Șițoiu, A. (2025). Pre-service teachers' perception regarding reflective competence. *Revista Românească pentru Educație Multidimensională*, 17(3), 50-83. https://doi.org/10.18662/rrem/17.3/1013

Wilkinson, S. (2004). Focus Group Research. In D. Silverman (Ed.), *Qualitative Research Theory, Method and Practice* (pp.177-199). SAGE Publications.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, 59, e12599. https://doi.org/10.1111/ejed.12599

# LA «MULTIGENERICITE» DANS LE REGARD D'AUREA: REGLES DU GENRE ET LOGIQUES EDITORIALES

**MOTAABBID Bouchaib**

*Chouaïb Doukkali University,*

*El Jadida, Morocco*

**RESUME**

Cet article propose une réflexion sur les enjeux qui déterminent la « multigénéricité » sous-tendant Le Regard d'Aurea, roman d'anticipation d'Isaac Azancot. Nombre de chercheurs constatent que l'hybridité est une caractéristique de l'esthétique narrative contemporaine. Mais rares sont ceux qui l'ont étudiée dans son rapport avec des enjeux de réception dans les romans récents. Aussi convient-il d'interroger ce roman azancotien quant aux raisons de ses choix thématiques et génériques. Cette étude part d'une hypothèse initiale: l'écriture romanesque dans cette œuvre ne peut ignorer ni les exigences éditoriales ni les attentes du lectorat contemporain. Nous adopterons ici une analyse qualitative et comparative: seront examinés les aspects génériques, thématiques et paratextuels du roman en question. Nous montrerons, d'une part, que le roman s'inscrit dans les codes classiques de la littérature d'anticipation et de la science-fiction, en articulant fiction scientifique et réflexion éthique. D'autre part, nous mettrons en évidence l'importance de la « multigénéricité », dans cette œuvre : intrigue amoureuse, récit de voyage, récit d'anticipation, etc. Ces aspects visent à rendre le récit plus accessible et attractif. Ainsi apparaît donc la recherche d'un équilibre entre exigence littéraire et prise en compte des logiques du marché éditorial.

***Mots clés*** *: Le Regard d'Aurea-intelligence artificielle ; roman d'anticipation ; « multigénéricité » ; logiques éditoriales.*

## 1. Introduction

À l'heure actuelle, l'intelligence artificielle a non seulement pénétré dans plusieurs domaines de la vie quotidienne, mais elle devient aussi un topique littéraire, comme en témoigne *Le Regard d'Aurea* d'Isaac Azancot (Azancot, 2025), roman d'anticipation caractérisé par un futurisme ambivalent, où les avancées de l'intelligence artificielle sont représentées comme oscillant entre une utopie technologique — évoquant « le meilleur des mondes » — et un chaos susceptible de mener à l'apocalypse.

Si le choix de ce sujet s'explique surtout par l'intention de se mettre au diapason de cette actualité qui s'impose de plus en plus — et, partant, de plaire aux lecteurs engoués de thèmes d'actualité —, en serait-il de même en ce qui concerne celui des autres paramètres de l'écriture romanesque dans cette œuvre azancotienne ? Pour le dire autrement, comme l'écriture d'un roman est le résultat d'une série de choix opérés par l'instance auctoriale à différents niveaux — genre romanesque, thématiques, vision du monde, univers fictionnel, modalités narratives, structure, style, etc. — et comme le romancier apparaît, en apparence du moins, comme le maître de ces choix, qu'est-ce qui peut justifier les prédilections de notre romancier ici ?

S'agissant de l'esthétique adoptée dans *Le Regard d'Aurea*, les choix génériques méritent une attention particulière. En effet, la question de la présence — ou de l'orchestration — de genres divers au sein de ce récit se pose avec pertinence dans la mesure où le roman d'anticipation ou de science-fiction se caractérise essentiellement par son hybridité (Willmann et al., 2023).

Si le mélange des genres littéraires est une modalité scripturale ancienne, il devient, dans les temps actuels, une tendance d'écriture romanesque fort répandue. Cette technique narrative, appelée également « hybridité » générique ou encore « intergénéricité » (Harvey, 2011, p. 33) ou « multiplication générique » (Genette et al., 1986, p. 202), est, comme l'a bien constaté Christine Lorre-Johnston, « liée à l'esthétique moderne » (Lorre-Johnston, 2011, p.1). Face à ce foisonnement d'appellations, il serait judicieux d'employer ici la notion de « multigénéricité ». Celle-ci a l'avantage de faire apparaître la multiplicité des formes que peut revêtir ce procédé d'écriture, tandis que celle d'hybridité réduirait, à notre sens, ce phénomène à la fusion des genres, et l'« intergénéricité » aux relations pouvant s'établir entre ces derniers. Nous l'employons donc ici dans son sens étendu, désignant à la fois la présence de genres (ou sous-genres) divers dans un roman, les rapports qui existent entre eux, ainsi que le mélange (ou la fusion) générique.

En outre, il importe de s'interroger sur les enjeux qui déterminent cette « multigénéricité ». Pierre Bourdieu, entre autres, a eu le mérite d'étudier ces visées dans son ouvrage *Les Règles de l'art. Genèse et structure du champ littéraire* (Bourdieu, 1992, p. 235-236). Il y distingue deux « logiques économiques » : la logique « anti-économique » et celle dite « économique ». La première, privilégiant « l'art pur », vise un capital symbolique et un profit à long terme, tandis que la seconde s'intéresse à la diffusion et au succès à court terme. Le roman qui nous occupe dans la présente réflexion ne s'inscrit-il pas essentiellement dans la logique « court-termiste » ? Combiner des genres multiples n'est-il pas une stratégie commerciale et éditoriale consistant à « s'ajuster à la demande préexistante de la clientèle » (Bourdieu, 1992, p. 235-236).

L'objectif de cette analyse est donc de mettre en évidence dans quelle mesure la « multigénéricité » sous-tendant *Le Regard d'Aurea* souscrit à la fois à des exigences génériques et à des impératifs éditoriaux. Cette étude sera menée selon une analyse textuelle qualitative où seront interrogés tour à tour les aspects thématiques de l'œuvre en question, quelques éléments paratextuels, les genres privilégiés, ainsi qu'une approche comparative consistant en la mise en regard de certains éléments textuels avec d'autres œuvres.

Il s'agira ainsi d'examiner les traits par lesquels ce roman s'inscrit dans les caractéristiques « classiques » du genre de l'anticipation, voire de la science-fiction, notamment à travers la représentation de l'intelligence artificielle comme vecteur de configuration d'un monde édénique. Il s'agira ensuite de rendre compte de la coexistence, au cœur du récit, d'une vision foncièrement alarmiste du progrès technologique. Enfin, nous montrerons que, derrière ces choix conformes aux codes génériques de l'anticipation, se dessine une dynamique de « multigénéricité » visant à satisfaire les attentes d'un lectorat large et hétérogène.

## 2. Un roman d'amour

Il va de soi que, eu égard à plusieurs caractéristiques, *Le Regard d'Aurea* relève du roman sentimental. En premier lieu, il importe de souligner que son titre met en avant le motif du regard, acte fondamental dans la relation amoureuse. On observe, en outre, que le sous-titre indique clairement, dès le seuil du livre, la catégorisation et l'orientation génériques de l'œuvre : « L'histoire d'un amour sublime […] » (Azancot, 2025).

L'intrigue amoureuse, quant à elle, constitue le canevas narratif du récit et mobilise les composantes habituelles du topos de la rencontre amoureuse telles qu'elles ont été recensées et étudiées par Jean Rousset (Rousset, 1981). En premier lieu, on reconnaît « l'effet », qui se manifeste par la fascination immédiate de Giulia, par ce coup de foudre survenu lors de sa première rencontre avec Elie Cohen. On relève également « l'échange » — qu'il revête la forme de paroles, de correspondances affectives, de messages électroniques, de dons généreux ou de tout acte de largesse que l'un des amants peut accomplir afin de plaire à l'autre. Enfin, le « franchissement » s'incarne à travers les déplacements répétés des deux protagonistes amoureux pour se réunir, à l'écart du monde social le plus souvent, leur rapport charnel, la relation matrimoniale, la fondation d'une famille, la naissance d'une fille, etc.

Par ailleurs, l'évocation de ces motifs propres au topos amoureux par le romancier est un fait constant. *Le locus amoenus* servant de cadre d'accueil aux rendez-vous des amants et l'importance accordée à la description de la nature, notamment, s'inscrivent, à titre illustratif, dans une tradition bien établie de la « littérature amoureuse », à l'instar du tilleul sous lequel se rencontrent Madame de Rênal et Julien Sorel dans *Le Rouge et le Noir* de Stendhal (Stendhal, 1977, p. 72).

Toutefois, ce qui singularise ce roman d'anticipation réside dans la fonction attribuée à l'intelligence artificielle dans la dynamique amoureuse des protagonistes, à savoir celle d'un véritable adjuvant narratif. Dès lors, il apparaît que l'intention fondamentale d'Azancot consiste à appréhender l'intelligence artificielle comme un médiateur du lien affectif, susceptible d'optimiser la vie sentimentale des individus (Azancot, 2025). Cette merveille technologique est ainsi investie d'une mission éthique dans la mesure où elle est censée contribuer au bonheur des amants, à l'équilibre conjugal et familial et, au sens le plus large, à l'épanouissement sentimental — voire libidinal — des êtres humains.

## 3. Un roman de voyage

Il est à remarquer aussi que l'intrigue amoureuse se double, dans ce roman, du récit de voyage. En effet, il s'agit ici d'un voyage dans un temps futuriste et dans un espace aussi bien étendu que diversifié.

S'agissant du voyage dans le temps, les mois qui jalonnent cette histoire, les années qui rythment sa progression témoignent de l'ampleur temporelle du périple effectué, notamment par Giulia. Sa trajectoire va en effet de l'an 2034 à 2062.

Bien évidemment, ce long périple a lieu dans différents lieux. Giulia traverse des continents, des pays, des villes, des sites et des domaines tout au long de ce voyage : l'Italie, la France, le Maroc… ; San Gimignano, Paris, Tanger… De plus, les moyens de transport évoqués, qui sont souvent ultrasophistiqués, à l'image de l'ère futuriste de l'histoire racontée, dénotent l'importance du voyage dans cette œuvre : de la voiture « totalement autonome » qui « roule toute seule » (Azancot, 2025, p. 42) à l'avion « qui utilise de l'hydrogène » (Azancot, 2025, p. 45).

D'ailleurs, l'isotopie du voyage est assez prégnante dans cette œuvre. Se répètent en effet, tout au long du récit, des vocables comme « voyage », « aventure », « déplacement », « découverte », « vacances », « départ », etc. Et ce long périple donne lieu à des descriptions pittoresques. On en a un bel exemple :

> La route serpentait, suspendue au-dessus de la vallée. Les couleurs de la campagne changeaient en permanence, faisant alterner des plaines verdoyantes, des collines couvertes d'oliveraies aux reflets argentés, jusqu'à cette terre brune qui prenait des allures volcaniques.
>
> […] La vue dominait une étendue d'étonnants paysages de collines, de champs de tournesol, de vignobles recouvrant des terres vallonnées, de fermes isolées entourées de cyprès, de villages médiévaux perchés, dont la majesté renforçait chez elle le sentiment de confusion qui l'emplissait depuis son départ (Azancot, 2025, p 9.)

L'intérêt essentiel de ces descriptions est sans doute de plaire à un lectorat large, de satisfaire son goût d'aventures, d'évasion, sa curiosité de touriste. D'où la grande importance attachée à la description des sites touristiques, dont la plupart sont d'ailleurs de renommée internationale: Volterra, les grottes d'Hercule, etc. En outre, ces descriptions à l'air ethnographique ménagent une place remarquable à ce qui a trait à l'exotique: l'art japonais de « la découpe » (Azancot, 2025, p. 80), l'art culinaire étranger (italien, marocain…), etc.

Tous ces éléments relevant du pittoresque et de l'exotisme font rapprocher ce roman d'une brochure touristique, dont la finalité ultime pourrait consister à satisfaire les goûts d'un grand public de lecteurs, et non uniquement ceux du lectorat spécialiste. La logique commerciale semble primer en ce qui concerne le choix « multi-générique » de ce livre.

Roman d'amour doublé d'un récit de voyage, *Le Regard d'Aurea* n'en demeure pas moins un roman de résilience.

## 4. L'intelligence artificielle et la résilience

Dans *Le Regard d'Aurea,* un constat s'impose d'emblée : un lien étroit est créé entre les progrès futurs— que ce soit dans un avenir proche ou lointain— envisagés dans les domaines de l'intelligence artificielle, de la physique quantique et de l'informatique, et la résilience des bénéficiaires potentiels de ces inventions (Azancot, 2025, p. 52). Cette vertu associée au progrès technologique attendu est susceptible, d'après l'histoire racontée, de contribuer au bonheur de l'ensemble de l'existence future, plus particulièrement ce qui se rapporte à la situation des deux protagonistes du roman : Elie Cohen et Giulia.

Chez le premier, la résilience se manifeste avant tout à travers les opportunités qu'offre cette révolution technologique pour dépasser certains « complexes », au sens psychanalytique du terme, notamment le handicap physique résultant d'un accident de voiture. Handicap qui se trouve doublé dans ce roman, comme on peut s'y attendre, d'un traumatisme psychologique. Par ailleurs, les carences affectives dont souffre cet ancien chercheur en médecine constituent un ressort fondamental de son parcours. Lors d'un aveu d'ordre intime adressé à son amante, il confesse que son investissement intense dans la recherche scientifique en général, et dans le projet d'Aurea en particulier, émane non seulement d'un héritage familial (Azancot, 2025, p. 83), mais aussi et surtout d'un désir profond de pallier le déficit affectif né de la séparation précoce avec sa mère durant l'enfance (Azancot, 2025, p. 52).

La persévérance singulière dont fait preuve Elie Cohen s'apparente ainsi à un processus de sublimation, terme à entendre ici dans l'acception freudienne, c'est-à-dire comme un mécanisme de transposition des souffrances psychiques vers des réalisations intellectuelles et scientifiques. Dès lors, l'intelligence artificielle apparaît comme l'instrument privilégié de cet effort soutenu, par lequel le personnage compte parvenir à surmonter les séquelles dues aux blessures de l'enfance afin de réaliser l'accomplissement personnel souhaité.

Elie Cohen se révèle donc comme un personnage pleinement conscient de sa fragilité affective. Aussi s'efforce-t-il d'y remédier — et, par là même, de consolider sa résilience — en s'engageant, parallèlement à ses occupations scientifiques, dans une relation amoureuse avec Giulia. Il aspire ainsi à substituer à l'amour maternel défaillant l'amour compensateur de l'amante. « Il a besoin du regard réparateur [de Giulia] », précise Aurea (Azancot, 2025, p. 153). Or, c'est bel et bien grâce à l'intelligence artificielle que cette quête peut trouver son aboutissement. En effet, non seulement la relation entre ces deux amants est médiatisée, facilitée et renforcée par Aurea, qui contribue à la séduction de Giulia, mais encore ce domaine technologique constitue pour le couple un espace propice à réaliser des projets intellectuels convergents ou communs et à consolider l'harmonie affective. Bref, la nouvelle technologie présente l'avantage d'inscrire la relation amoureuse dans la durée.

Cette défaillance affective n'est toutefois pas l'apanage d'Elie : Giulia en porte également les traces. En effet, par l'entremise d'Aurea — c'est-à-dire, ipso facto, de l'intelligence artificielle —, Elie Cohen accède aux zones les plus profondes de la subjectivité de son amante. Il réussit à déchiffrer les motivations intimes de sa conduite et les secrets (Azancot, 2025, p. 57) qui expliquent son attirance instantanée et simultanée pour lui et pour ce domaine de recherche.

L'intelligence artificielle se présente ainsi, dans ce roman azancotien, comme un véritable catalyseur de résilience affective et sentimentale chez les êtres humains. À cette dimension psychologique s'ajoute une fonction réparatrice d'ordre corporel. En effet, grâce aux avancées médicales réalisées par Aurea, Elie parvient progressivement à surmonter le handicap moteur résultant de son accident de voiture. Elle retrouve une souplesse corporelle autrefois perdue (Azancot, 2025, p. 162). Plus, encore, le roman évoque la possibilité d'une intervention chirurgicale entièrement assistée par Aurea, à même de le libérer définitivement de cette paralysie, s'il le voulait (Azancot, 2025, p. 174).

L'intelligence artificielle s'impose donc, dans l'économie narrative du *Regard d'Aurea*, comme une innovation quasi providentielle, en mesure de favoriser tant la résilience psychique que la réparation physique de l'être humain.

## 5. L'intelligence artificielle, une solution miracle

Dans ce roman futuriste, où Isaac Azancot (2025) fait l'éloge de l'intelligence artificielle, celle-ci est considérée, comme nous l'avons déjà souligné, comme une solution quasi-miraculeuse, capable d'éradiquer tous les malheurs et maux qui accablent l'humanité aujourd'hui comme à l'avenir. L'un des domaines où cette intervention pourrait s'effectuer de manière exemplaire est la médecine. En effet, dans ce roman azancotien, les avancées médicales incarnées par Aurea pourraient permettre, selon l'auteur, de mieux soigner les malades. S'appuyant sur sa carrière de médecin, spécialiste en informatique médicale et professeur au collège de médecine, Azancot (2025) s'attache, à travers cette œuvre de fiction scientifique, à professer un optimisme médical fondé sur l'espoir en l'avenir de ce secteur vital.

Parallèlement à cette confiance dans la capacité de la médecine du futur à mieux guérir les maladies, s'esquisse une réflexion sur l'éthique. À cet égard, l'auteur redéfinit celle-ci: il ne considère pas l'usage de l'intelligence artificielle dans la vie privée des patients comme une transgression des règles éthiques ou comme une forme d'espionnage, mais plutôt comme un moyen de réaliser un diagnostic complet et de contribuer au sauvetage de vies, à condition que cet usage reste strictement encadré et limité au cadre noble de la médecine.

De surcroît, les progrès technologiques envisagés par le roman pourraient résoudre de nombreux problèmes liés à la sécurité des individus et des nations, tels que les crimes et les attentats terroristes. Ainsi, l'auteur évoque le célèbre attentat de Manhattan en 2001 et déplore qu'Aurea n'ait pas existé à cette époque-là. Selon lui, son invention aurait permis de prévoir et d'éviter cet acte terroriste, comme ce fut le cas pour l'attaque de l'hôtel, évitée de justesse grâce à l'intelligence artificielle. On note également le rôle qu'Aurea joue dans la résolution de la cyberattaque survenue à l'hôpital où travaille Giulia, provoquée par une faille du système de sécurité. Sans l'intervention d'Aurea, cette attaque n'aurait évidemment pas pu être neutralisée. Les multiples références aux attentats terroristes montrent de manière évidente l'importance que le roman accorde à la vertu salvatrice que pourrait représenter la technologie du futur.

Il apparaît donc que le progrès envisagé dans ce roman d'Azancot (2025) pourrait provoquer une véritable révolution sécuritaire bénéfique pour toute l'humanité. On peut alors se demander : d'où l'auteur a-t-il puisé son inspiration ?

## 6. Les prédécesseurs d'Azancot et l'apologie du progrès

Il convient de souligner que cette utopie technologique n'est pas récente. De nombreux écrivains se sont faits les défenseurs de cette vision futuriste. On pense notamment à Thomas More Morus (1717), pionnier de l'utopisme, qui imaginait un univers parfait ; à H. G. Wells et sa célèbre *Machine à explorer le temps* (Wells, 1898) ; à Aldous Huxley et son *Meilleur des mondes* (Huxley, 2023) ainsi qu'à Isaac Asimov, entre autres. Le point commun de ces auteurs réside dans leur capacité à décrire soit une utopie, un monde idéal, soit une dystopie, parfois en combinant les deux.

Des écrivains plus récents ont également adopté cet optimisme futuriste, en s'appuyant sur leurs anticipations des progrès susceptibles de rendre les frontières entre humains et machines de plus en plus floues, allant quelquefois jusqu'à imaginer une forme d'hybridation ou de fusion. C'est le cas, par exemple, de Donna Haraway et sa notion de *cyborg* (Haraway, 2006) ainsi que de William Gibson et son *cyberpunk* (Gibson, 1985).

En cherchant à dépeindre un monde utopique parfait fondé sur de futurs progrès en intelligence artificielle, Azancot (2025) s'inscrit ainsi dans la lignée de ses prédécesseurs, maîtres du roman d'utopie scientifique et technologique. On observe qu'il s'inspire de ces auteurs de fiction futuriste lorsqu'il introduit dans son roman un personnage ayant subi une opération d'hybridation, « Éden ». Cependant, cette hybridation n'est pas envisagée sans risques.

## 7. La mise en garde contre les dérives potentielles

Roman d'utopie, *Le Regard d'Aurea* n'en demeure pas moins une œuvre à portée dystopique. En effet, l'apologie que fait Azancot (2025) de l'intelligence artificielle n'est pas exprimée sans appréhension. Cet opus est traversé de part en part par des mises en garde contre les risques qui pourraient accompagner l'implantation des innovations attendues dans ce domaine et leur introduction dans les divers secteurs de la vie. Le chemin menant vers cet eldorado futur reste semé d'embûches, susceptibles à tout moment de faire basculer un monde édénique dans une catastrophe apocalyptique.

Commençons par le sous-titre du roman. L'auteur qualifie cette époque de progrès technologique de « temps de […] guerre 3» (Azancot, 2025). Vise-t-il ainsi à préparer le lecteur, dès le seuil de l'œuvre, à la double posture qu'il adopte envers cette ère à venir, faite à la fois d'enchantement et d'inquiétude ? Quoi qu'il en soit, l'usage de l'antithèse dans ce sous-titre met clairement en relief cette attitude ambivalente.

D'autres indices révèlent également l'insistance de l'auteur à mettre en garde contre les dangers potentiels d'une quête effrénée et aveugle du seul progrès technique. Il écrit à ce propos :

> Une intelligence artificielle […] peut être une menace existentielle, mais également une chance permettant à l'homme de se réaliser et de se protéger en franchissant le cap de "la singularité technologique". […] L'allégorie du voile […] signifiait sans doute l'importance de ne pas exposer trop vite des avancées susceptibles de provoquer une rupture brutale, incompatible avec le rythme de l'évolution culturelle humaine. (Azancot, 2025, p. 130).

Azancot n'envisage pas donc un paradis futur dépourvu de morale, d'éthique, d'amour ou de sauvegarde des assises spirituelles, bref, de vertu. D'où l'importance accordée dans le texte aux penseurs de la vertu, d'Aristote et Platon à Bergson. Ces allusions récurrentes témoignent de la volonté de l'auteur d'offrir à son récit une assise intellectuelle et philosophique dont le but principal est de s'appuyer sur l'autorité du bergsonisme éthique.

L'incipit (ou « prologue ») va encore plus loin, en transposant les mises en garde sur le plan narratif pour en faire un récit de catastrophe. Présenté sous forme de prolepse — où la fin de l'histoire est dévoilée d'emblée, à la manière du roman policier —, ce début est marqué d'une tonalité tragique. L'ambiance est lugubre : la seule fille née de l'union de Giulia et Elie, Eden, est peut-être morte, comme le suggèrent les propos de la mère (Azancot, 2025, p. 7). La cause ? Une tentative, sinon irréfléchie, du moins prématurée, d'effectuer une opération destinée à fusionner l'humain et la créature virtuelle. À la suite de cet événement tragique par excellence, le récit expose la ruine : un renversement de situation où tout l'édifice paradisiaque s'effondre. La décision finale de Giulia, celle de ne plus adresser la parole à son époux, marque le début de la chute, de la période eschatologique, de l'apocalypse.

[3] Première page de couverture.

Il convient de rappeler que cette dimension alarmiste n'est pas nouvelle dans la littérature d'anticipation. En effet, les récits utopiques comportent souvent une dimension dystopique. Ainsi, dans La *Nouvelle Atlantide* de Francis Bacon, apparaît une mise en garde contre une dérive possible du progrès scientifique :

> Nous avons aussi des assemblées et des délibérations, dont l'objet spécial est de désigner les observations, les expériences qui doivent être publiées, et celles que nous devons nous réserver ; car tous les membres de l'institut s'obligent ave serment à garder le plus rigoureux secret sur toutes les vérités dont la publication nous parait dangereuse (Bacon, 1799, p. 481-482).

De même, dans *Un feu sur l'abîme* de Vernor Vinge (1994), il y a un avertissement contre les risques liés à un futur dominé par une intelligence prodigieuse pouvant conduire à la perte.

## 8. Une esthétique soumise aux impératifs éditoriaux ?

Il importe également de souligner que ce roman ne s'inscrit pas seulement dans la lignée des récits d'anticipation, de science-fiction ou des romans futuristes. On y est frappé aussi par l'intrigue sentimentale qui le traverse, ainsi que par les éléments du récit de voyage et le pittoresque, comme nous l'avons déjà indiqué. Qu'est-ce qui pourrait expliquer cette « multigénéricité » ?

Si l'auteur met en avant cette aventure amoureuse, c'est sans doute parce qu'il est persuadé que cette dimension permet de séduire un public plus large. En effet, comment ce récit foisonnant d'informations scientifiques et techniques aurait-il été perçu par un lectorat étendu et diversifié si cette idylle avait été omise ? Rebutant ! Ennuyeux ! Il n'aurait attiré qu'une élite restreinte, passionnée d'imaginaire scientifique peut-être, mais non les autres lecteurs, ceux qui recherchent une lecture-plaisir ou qui aiment s'identifier aux personnages. De même, l'intérêt que porte le romancier à l'aventure touristique est susceptible de satisfaire ce désir d'exotisme, cette quête de nouveaux horizons propres à offrir une évasion dont pourrait être friand un large public.

Il apparaît donc que, derrière cette « multigénéricité » assumée par Azancot (2025), se profilent des impératifs éditoriaux. Un livre est, ne l'oublions pas, un objet destiné à être commercialisé. De ce fait, l'instance éditoriale intervient souvent, sinon pour imposer ses exigences à l'auteur, du moins pour l'orienter — à moins que celui-ci ne décide de publier à compte d'auteur. D'où, à notre sens, la place qu'accorde cet écrivain à cette double aventure : amoureuse et touristique.

Une étude récente réalisée par le Centre national du livre met en lumière les goûts du lectorat français. Selon cette enquête, les tendances les plus marquantes concernent, outre l'intérêt pour les mangas-comics, la science-fiction, le roman sentimental et le développement personnel (Mercier et al., 2025). Cette convergence entre les préférences relevées par cette étude et les caractéristiques du roman est-elle liée à la forte probabilité — sinon d'orientations ou de recommandations de l'instance éditoriale — du moins à la motivation commerciale du romancier ? À sa volonté de toucher un lectorat large et de plaire à un public varié ? C'est un facteur, parmi d'autres évidemment, qui éclaire sans doute la « multigénéricité » sous-jacente au *Regard d'Aurea*.

## 9. Conclusion

En somme, il ressort clairement de notre analyse que *Le Regard d'Aurea* se construit sur une pluralité de choix génériques et thématiques, que l'on ne peut comprendre qu'en tenant compte de plusieurs logiques. D'un côté, le roman s'inscrit dans la tradition de la littérature d'anticipation. Azancot (2025), en héritier de ses prédécesseurs, mêle avec soin la fiction scientifique au discours philosophique, esquissant un avenir dominé par l'intelligence artificielle tout en soulevant des interrogations éthiques qui traversent tout le récit.

Mais, en même temps, il ne faut pas perdre de vue une autre réalité : la présence marquée de l'intrigue amoureuse, du motif du voyage à caractère exotique et pittoresque, l'intérêt accordé à la résilience et au développement personnel, etc. Ces composantes ne sont pas de simples ornements. Elles traduisent la volonté de l'auteur de rendre son œuvre accessible, agréable à lire, capable de séduire un lectorat large. L'histoire sentimentale, en particulier, joue un rôle essentiel : elle humanise un univers parfois dense sur le plan scientifique et permet au lecteur de s'identifier aux protagonistes. Quant à la dimension touristique, elle répond à ce désir d'évasion et de dépaysement que recherchent tant de lecteurs.

Cette « multigénéricité », qui allie anticipation, sentimentalisme et aventure, n'est donc pas le fruit du hasard. Elle reflète à la fois une ambition littéraire et une attention réelle aux tendances du marché du livre, comme l'a expliqué Pierre Bourdieu. Les études sur les goûts des lecteurs — et notamment leur intérêt pour la science-fiction, le roman sentimental ou encores les récits axés sur le développement personnel — montrent bien que les écrivains ne peuvent ignorer ces attentes. Sans forcément se plier complètement aux impératifs éditoriaux, Azancot (2025) semble avoir pris en compte ces tendances pour donner à son roman une portée plus large.

Ainsi, *Le Regard d'Aurea* apparaît comme un récit hybride, mais cohérent, qui cherche à concilier réflexion sur l'avenir et plaisir de lecture. C'est sans doute là que réside sa force : dans cette capacité à unir, sans les opposer, le discours prospectif et l'émotion, la science et la sensibilité, l'innovation et l'humanité.

## Références

Azancot, I. (2025). *Le Regard d'Aurea*. 9791032934081. Éditions de l'Observatoire. https://editions-observatoire.com/livre/Le-Regard-d%27Aurea/555?utm_

Bacon, F. (1799). *Œuvres de François Bacon* (Vol. 11, trad. Antoine Lasalle, com notas críticas, históricas e literárias). Dijon: Impr. de L.-N. Frantin. *Bibliothèque nationale de France*. https://www.geneanet.org/bibliotheque-genealogie/doc/74210/oeuvres-de-francois-bacon-volume-11-traduites-par-ant-lasalle-avec-des-notes-critiques-historiques-et-litteraires?utm_

Bourdieu, P. (1992). *Les règles de l'art: Genèse et structure du champ littéraire*. Seuil. https://mediatheques.montpellier3m.fr/ESBAMA/doc/CAMO/322525/les-regles-de-l-art-genese-et-structure-du-champ-litteraire?utm

Genette, G., Jauss, H. R., Schaeffer, J. -M., Stempel, W. D., Viëtor, K. (1986) *Théorie des genres*. (Collection Points). Paris: Seuil. ISBN 9782020090476. https://www.ombres-blanches.fr/product/168295/theorie-des-genres?utm

Gibson, W. (1985). *Neuromancien* (J. Bonnefoy, Trad.). La Découverte. 978-2-7071-1562-1

Haraway, D. (2006). *Manifeste Cyborg.* (N. Magnan, Trad.). *Mouvements,* (2006/3, n. º 45-46), 11-14. Éditions La Découverte. https://doi.org/10.3917/mouv.045.14.

Harvey, F. (2011). *Alain Robbe-Grillet : le nouveau-roman composite. Intergénéricité et intermédialité.* Fabula.9782296547865. L'Harmattan.

Huxley, A. (2023). *Le Meilleur des mondes* (J. Kamoun, Trad.). Plon. 9782259316262. https://www.maraisdulivre.fr/product/show/9782259316262/le-meilleur-des-mondes?utm_

Lorre-Johnston, C. (2011).  Chapitre VI : L'hybridité générique. In F. Palleau-Papin (Éd.), *William T. Vollmann, le roman historique en question. Presses Sorbonne Nouvelle.* https://doi.org/10.4000/books.psn.7797

Mercier, E., Tétraz, A., & Leray, A. (2025). *Les Français et la lecture. Résultat*

*2025.*[Report]. Centre National du Livre /Ipsos.

https://www.ipsos.com/sites/default/files/ct/news/documents/2025-04/ipsos-cnl-francais-lecture-2025-rapport-complet.pdf?utm_

Morus, T. (1717). *L'Utopie* (N. Gueudeville, Trad.). R. & G. Wetstein.

Rousset, J. (1981). Leurs yeux se rencontrèrent. La scène de première vue dans le roman. (3éd). 2714303064.  José Corti.

Stendhal (1977). *Le Rouge et le Noir.* 978-2070412396. Presses Pocket.

Vinge, V. (1994). *Un feu sur l'abîme* (G. Abadia Trad). Robert Laffont.

Wells, H. G. (1898). *La Machine à explorer le temps* (H. D. Davray, Trad.). Mercure de France.

Willmann, F., Esselborn, H., Ferragamo, E., Siddharth, A., Delianidou, S., Hagemann, S., Benoit, M., Jouin, C., Beggour, I., Lerousseau, A., Roche, A., Waszek, N., Corbin, A.-M., & Chamayou-Kuhn, C. (2023). *La science-fiction, genre hybride et critique, Germanica, 72*, 184, Special Issue. https://hal.science/hal-04393073

# "PICK PICK": LEVERAGING AI FOR CUSTOMISING ANALOGUE GAMES IN INCLUSIVE AND PERSONALISED LEARNING ENVIRONMENTS

**André Rocha, Fernanda Bonacho, Filipa Lopes, Joana Sousa & Nuno Palma**

*IPL-Escola Superior de Comunicação Social, Lisboa, Portugal*

**ABSTRACT**

Over time, games have increasingly demonstrated their pedagogical potential in educational contexts. Due to their hybrid nature, they can be integrated into both recreational and learning environments, thereby creating opportunities for potentially innovative teaching experiences. However, in teaching practice—particularly at the higher education level—games are infrequently used. One contributing factor is the difficulty of finding games whose dynamics and content align with educators' pedagogical requirements.

The CoLab Games project seeks to address this gap through the development of the Pick Pick platform4, a digital resource that enables users to custom-create analogue games for deployment in educational settings within the domains of communication, literacy and citizenship.

Currently at the prototype stage, the platform combines modular logic tools with Generative Artificial Inteligence (GAI), allowing users to personalise core game elements—such as the board, game mechanics, questions and challenges—via an intuitive digital interface. Created games are made available to the community through a library that provides files for physical production by means of digital fabrication (*e.g.,* 3D printing, cutting plotter) or conventional methods (*e.g.,* hand drawing, desk-jet printing), fostering a synergy between open educational resources and maker culture.

To evaluate the platform, the workshop "Creating Board Games for Pedagogical and Communication Contexts" was held at FabLab Benfica, during which a group of in-service and pre-service teachers tested the platform's usability and produced several personalised game prototypes.

This article aims to present the platform and discuss the workshop's outcomes in terms of its pedagogical potential and collaborative practices.

[4] https://www.pickpick.pt/

***Keywords:*** *Collaborative Game Design; Artificial Intelligence; Media and Information Literacy; Open Educational* Resources

## 1. Introduction

This article focuses on the presentation and evaluation of the online Pick Pick platform, developed within the CoLab Games project—a digital environment for the customisation of analogue games—funded by the Polytechnic Institute of Lisbon (IPL/IDI&CA2024/CoLabGames_ESCS) under the 9th edition of IDI&CA (Annual Competition for Research, Development, Innovation and Artistic Creation Projects). The primary aim of this project is to create a collaborative platform that enables users to design bespoke analogue games enriched with digital content for use in educational settings across the fields of communication, literacy and citizenship.

Although playful activities are not essential to human survival (Huizinga, 1980), they serve to consolidate a value system that underpins social life, and the act of playing is both intrinsic to the human condition and beneficial to individual well-being (Lopes, 2013). Games have evolved alongside humanity's technological and cultural advances, from the earliest known boards—dating back over 5 000 years (Piccione, 1980)—to today's highly sophisticated digital simulations. This evolution has increased the complexity of games—expanding their mechanics, themes and immersive qualities—whilst exploring new dimensions and preserving their fundamental nature. Consequently, games are increasingly used for pedagogical, communicational and educational purposes, in addition to their recreational function (Ekin et al., 2023).

Despite the growth of digital game-based learning tools—including platforms that leverage Artificial Inteligence (AI) to generate adaptive or interactive experiences—there remain few initiatives that create fully customisable analogue games. Existing research has predominantly focused on digital environments (Gee, 2003; Hamari et al., 2016; Zawacki-Richter et al., 2019), whereas the design of physical educational games has attracted relatively little scholarly attention. This imbalance points to an opportunity to investigate how GAI and maker-culture practices can support educators in developing meaningful analogue games tailored to their specific teaching contexts. Maker practice, rooted in the contemporary cultural movement that celebrates hands-on creation through the integration of digital technologies and artisanal methods (Anderson, 2014), is characterised by open knowledge sharing, collaborative project work and democratized access to production tools via spaces such as FabLabs, makerspaces and hackerspaces.

Against this backdrop, the Pick Pick platform has been conceived as a digital tool for educational contexts, specialising in the personalisation of board games. To address user needs, the platform offers three customisation modules:

a) generation of bespoke questions and challenges, with optional support from GAI;
b) adjustment of board complexity and game-mechanic types via a procedural board generator; and
c) selection of the game's physical output, whereby users download files that enable production in various formats (*e.g.,* paper, 3D).

Within the educational sphere, Pick Pick also embodies a collaborative ethos: games created by individual users are deposited in an open-access library for communal reuse. In doing so, the CoLab Games project responds to the Paris OER Declaration (UNESCO, 2012) by promoting the production and sharing of open educational resources. This collaborative feature aligns with the Web 2.0 paradigm (O'Reilly, 2005), also known as the participatory web, which is especially prevalent in gaming communities. For example, BoardGameGeek (https://boardgamegeek.com/), one of the largest online gaming communities, enables users to publish informational and opinion pieces on board games, collectible-card games and tabletop games, generally, demonstrating how such platforms foster active community engagement in the maintenance and evolution of game dynamics, particularly in digital contexts.

To assess Pick Pick's potential, a pilot workshop was organised for teaching practitioners, framed within the thematic priorities of the UNESCO Chair in Communication, Media and Information Literacy and Citizenship at ESCS-IPL[5]. Participants designed, tested and critically reflected upon their game prototypes. Throughout the process, user feedback was collected, and the platform's pedagogical and communicational relevance was discussed.

This article presents the Pick Pick platform and reflects on the outcomes of this pilot workshop focusing on its pedagogical potential and collaborative practices.

[5] The UNESCO Chair in Communication, Media and Information Literacy and Citizenship at ESCS-IPL tackles the digital challenges of the digital age, such as misinformation, fake news and hate speech by integrating research and education to foster media and information literacy, critical thinking and active citizenship. Through international cooperation, it bridges academia and civil society to promote freedom of expression, democratic values, combat polarization and violent extremism, and support sustainable development in line with UNESCO's principles.

## 2. *Pick Pick* Platform

Section 2 describes the functionalities of Pick Pick and the methodology applied in the "Creating Board Games for Pedagogical and Communication Contexts" workshop, which was primarily designed to test the platform with users and to foster discussion on its pedagogical potential.

### 2.1. Presentation of the Pick Pick Plataform

The Pick Pick platform offers a suite of resources and tools that enable users, from a predefined game dynamic, to create and customise:

a) a game board;
b) a deck of cards with thematic questions and challenges; and
c) the game's physical format.

The game board provided by Pick Pick features a modular design composed of spaces with specialised functions (questions, challenges, luck and mobility). This modular system allows for adjusting complexity levels and game mechanics, adapting the game to different audiences and contexts. Card content—questions and thematic challenges—can be created manually or generated with the support of GAI by completing a form that collects information about the target audience and the topic description. This functionality is intended to assist the user during the creative phase by suggesting context-appropriate content, that can be reviewed and refined thereafter. Once the design process is complete, the platform delivers a package of files for the physical production of the game, including both the cards and the board. These components may be printed on a standard printer or fabricated in various physical formats (*e.g.,* 3D printing, laser cutting, stamps, stencils) within a FabLab environment[6].

All games created by users are stored on the platform and made available in an open-access library, permitting their reuse and fostering new mediation scenarios.

Through this approach, the CoLab Games project seeks to promote synergy between the Maker culture and open educational resources, aiming to produce an educational and playful product that can be deployed in diverse contexts, while encouraging collaboration, innovation and a culture of creative, accessible learning.

[6] A Fab Lab, or digital fabrication laboratory, is a place to play, to create, to mentor and to invent: a place for learning and innovation. Fab Labs provide access to the environment, the skills, the materials, and the advanced technology to allow anyone, anywhere, to make (almost) anything.

### 2.2. Functioning ff the Pick Pick Platform

Upon accessing the platform, the user is presented with a brief overview of its operation and a prominent “Create Game” button on the home page. By selecting this button, the user is taken to the first of three sequential stages in the game-creation workflow.

In the first stage, the user completes a form specifying the game’s core attributes. The form requires:

(i) the game’s title;
(ii) the target audience,
(iii) a description of the game’s theme, and
(iv) a choice as to whether to utilise the Generative Artificial Intelligence (GAI) feature for question creation.

The thematic description field must include all relevant guidelines for generating questions and challenges. User-provided details are transmitted to the AI model together with a predefined system prompt—a standardised set of instructions that defines the objectives, style and format of the content to be generated. This combination of user input and system prompt steers the AI towards producing tailored, context-appropriate game content.

The second stage involves customising the game board via an integrated board generator. Here, the user adjusts five key parameters:

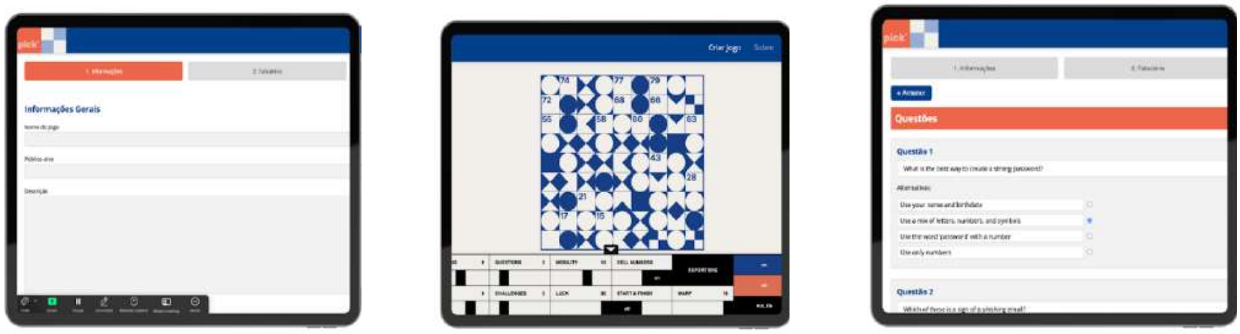
(i) the number of spaces (rows and columns),
(ii) the quantity of questions and challenges,
(iii) the number of luck and mobility spaces,
(iv) the visibility of space numbering and
(v) the colour scheme.

The generator responds in real time: any change to a parameter automatically triggers the creation and display of an updated board proposal. This immediate visual feedback allows the user to gauge the final board’s complexity and layout as they refine their selections.

In the third and final stage, the user reviews the card texts produced by the AI feature or manually enters new content if they have chosen to author all questions and challenges from scratch. Once satisfied, the user may proceed to export and share the completed game.

**Figure 1**

*Steps for creating a new game on the Pick Pick platform*



Upon completion of the game-creation process, the platform generates and provides the files required for the physical fabrication of all game components. Concurrently, the newly created game is archived in a shared library on the platform and made openly accessible to any user. In this initial prototype phase, the platform does not support subsequent editing of published games.

### 2.3. Pilot Workshop: Creating Board Games for Pedagogical and Communication Contexts with the Pick Pick Platform

To evaluate the Pick Pick platform, a workshop entitled "Creating Board Games for Pedagogical and Communication Contexts" was held, engaging twelve in-service and pre-service teachers. The workshop's objectives were to introduce the platform, guide participants in creating game prototypes, collect usability feedback on both the platform and the resulting prototypes, and reflect on the tool's applicability in educational and communicational settings.

The four-hour session took place at the FabLab Benfica and was structured into four successive phases: presentation of the CoLab Games project and the thematic focus areas; small-group development of functional game prototypes using Pick Pick; playtesting of the prototypes; and plenary discussion of the workshop outcomes. To collect feedback throughout the workshop, participants completed an evaluation grid addressing six process components:

(i) overall platform usability;

(ii) the game-creation form;

(iii) board customisation;

(iv) question and challenge revision;

(v) physical production (printing and cutting) of game elements; and

(vi) gameplay dynamics during prototype testing.

Participants were organised into groups of three to four and assigned a hypothetical scenario, defined by a learning context (*e.g.,* higher education, secondary education, community project) and a chosen fabrication method (*e.g.,* 3D printing, stencils, conventional printing). The game themes were aligned with the UNESCO Chair in Communication, Media Literacy and Citizenship at ESCS-IPL, focusing on the effects of information and media on society and the importance of critical thinking in the media ecosystem. Participants were encouraged to access the Portuguese Media Education Framework (Pereira et al., 2023) to inform their thematic choices.

Each group first allocated roles and tasks internally. They then progressed through Pick Pick's three creation stages—with the assistance of workshop facilitators—selecting appropriate physical output formats. Since revising AI-generated card content was time-consuming, groups began the physical fabrication of prototypes (board, cards and playing pieces) in parallel. Techniques employed included 3D printing, conventional printing, stamping, stencilling, marker drawing and vinyl cutting. Prototypes varied in scale and complexity. Upon completion, groups tested both their own and others' games.

Data collection combined:

(i) direct observation;

(ii) photographic and video documentation;

(iii)participants completed qualitative evaluation grids at each creation stage.

A final plenary debate solicited participants' views on the project's educational and communicational potential, identifying key advantages and limitations of both the Pick Pick platform and the games produced. Responses were analysed descriptively and interpretatively to discern patterns related to platform usability, the pedagogical relevance of the games, and the constraints encountered.

The study's main methodological limitations include the small number of participants, the workshop's brief duration, and its specific FabLab setting, which may restrict replication in less technologically equipped environments. Conversely, conducting the workshop in a FabLab also proved advantageous, as it allowed for the exploration of multiple digital fabrication techniques to tailor prototypes to diverse usage contexts.

## 3. Results and Analysis

Four prototypes were produced as a result of the "Creating Board Games for Pedagogical and Communication Contexts" workshop. Two were designed for higher education, one for secondary education and one for community-based work. The themes addressed media literacy and citizenship—specifically critical thinking, misinformation and the media's effects on society—in accordance with the Media Education Framework.

Table 1 presents a summary of the developed scenarios, detailing the target audience, thematic focus and physical production technique.

**Table 1**

*Informative summary of the prototype games developed in the "Creating Board Games for Pedagogical and Communication Contexts" workshop*

| Group | Target-audience | Theme | Production technique |
|---|---|---|---|
| **1** | **Higher education** | Creation, Production and Participation | Stamps and markers |
| **2** | **Secondary education** | Media, Information and Current Affairs / AI | 3D printing |
| **3** | **Higher education** | Social Representations and Stereotypes in the Media | Conventional printing |
| **4** | **Community-based work** | Fiction and Entertainment | Cutting plotter/ stencil |

The assignment of these scenarios defined the contextual framework for generating game-card content via GAI. In the game-specification form, participants detailed the intended application contexts, learning objectives, thematic focus and associated subtopics, as well as any other pertinent information and related bibliographic references.

Drawing on these inputs, the GAI tool produced initial content, which participants then edited and validated. Figure 2 shows two examples of cards created using the GAI feature.

**Figure 2**

*Examples of generated cards: question and challenge*


The prototype was developed using a range of digital and manual fabrication techniques, including 3D printing, stencilling, stamping, vinyl cutting, conventional printing and marker drawing. Figure 3 illustrates the various prototypes produced. This diversity of approaches demonstrated the flexibility of the production model and the potential for integrating Maker practices with educational methodologies. Holding the workshop in a FabLab environment and organising participants into small working groups enabled the creation of prototypes that employed multiple resources, fabrication techniques and materials, thereby reinforcing the project's analogue and collaborative dimensions.

**Figure 3**

*The game prototypes developed during the workshop*

## 4. Discussion and Conclusions

At the beginning of the workshop, each group was provided with an evaluation grid aimed at collecting qualitative feedback on the following aspects:

(i) overall platform usability;
(ii) the game-creation form;
(iii) board customisation;
(iv) question and challenge revision;
(v) physical production (printing and cutting) of game elements; and
(vi) gameplay dynamics during prototype testing.

For each criterion, participants were asked to record positive aspects, areas for improvement, and any additional comments or suggestions. In addition to completing the evaluation grid, participants were also invited, at the end of the workshop, to share and discuss their opinions and comments in a plenary session. This discussion was recorded in both audio and video formats for subsequent analysis.

During the analysis process, the collected comments were grouped into strengths, weaknesses, and opportunities. Overall, the feedback provided by participants on the platform and the games generated during the "Creating Board Games for Pedagogical and Communication Contexts" workshop was positive. Teachers recognised the platform's relevance and educational potential, describing it as easy to use—even by those without prior game-design experience. These perceptions align with the Technology Acceptance Model (Davis, 1989), which identifies perceived usefulness and perceived ease of use as the principal determinants of technology adoption. Subsequent studies applying this model in educational contexts (Teo, 2011) confirm that perceived usability directly influences teachers' intention to use technology.

Teachers identified two main opportunities for using the Pick Pick platform in educational settings: first, as a support tool for the creation of games to deliver and consolidate curricular content; and second, as a facilitator of active-learning practices in which students engage in the game-creation process. These proposed applications correspond with Kafai's (2006) argument that learning can occur not only through gameplay but also through game construction, thereby fostering critical thinking, creativity and problem-solving skills.

Hamari et al. (2016) identify a balance among challenge, engagement and rule clarity, as an essential condition for game-based learning to maintain flow and motivation throughout gameplay. In this respect, participants' feedback emphasised the need to refine the board-construction process to enhance playability and dynamism, since some generated boards were considered insufficiently stimulating in practice. One proposed solution was the inclusion of pre-designed, tested boards featuring diverse mechanics and dynamics.

In the exploratory study Towards AI-assisted Board Game-based Learning: Assessing LLMs in Game Personalisation (Tinterri et al., 2024), which examines AI-assisted personalisation of board games for educational purposes, findings indicate that contemporary GAI models—when provided with clear and comprehensive instructions—can produce personalisation suggestions comparable to those of human experts. Recent studies by Vicente (2023a;2023b) underscore that users consistently highlight the need for rigorous editorial review of AI-generated content. Although algorithmic suggestions accelerate the creative process and provide ideas for an initial structure, human intervention is indispensable to ensure accuracy of content, guarantee full pedagogical alignment with predefined learning objectives, refine language and enhance clarity. Far from constituting a technological constraint, this curation phase is reframed as an opportunity to strengthen the educator's role as a critical mediator of emerging technologies. Simultaneously, it invites learners into an active pedagogical cycle of game review and co-construction, thereby complementing and extending Kafai's (2006) thesis on the educational affordances of participatory game design.

Additionally, the platform was valued as a means of supporting teacher autonomy by providing pre-structured materials that can be easily adapted to diverse educational contexts and specific needs. Participants noted that this capacity for content reuse and adaptation enhances their confidence in the platform and facilitates its integration into real-world teaching practices. In doing so, Pick Pick aligns with the Paris OER Declaration (UNESCO, 2012) and echoes Anderson's (2014) call for open knowledge sharing through collaborative projects and democratized access to production tools in spaces such as FabLabs.

**5. Contributions and Limitations**

The "Creating Board Games for Pedagogical and Communication Contexts" workshop represented the public debut of the Pick Pick digital platform. Despite the pilot study's limited scale, the feedback gathered at this stage has highlighted essential features to be addressed and integrated into later prototype development phases. Future work will involve testing the prototype with diverse audiences and across various disciplinary areas to validate its relevance and adaptability to different themes and contexts.

By generating games aligned with the thematic scope of the UNESCO Chair in Communication, Media Literacy and Citizenship at ESCS-IPL, this workshop has made a significant contribution by producing four subject-focused games now available in the platform's shared library. These games are publicly accessible and have benefited from a thorough content review by the participating educators.

## References

Anderson, C. (2014). *Makers: The new industrial revolution.* Crown.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems, 13(3),* 319–339. https://doi.org/10.2307/249008

Ekin, C. C., Polat, E., & Hopcan, S. (2023). Drawing the big picture of games in education: A topic modeling-based review of past 55 years. *Computers & Education, 194*, 104700. https://doi.org/10.1016/j.compedu.2022.104700

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment, 1*(1), 20. https://doi.org/10.1145/950566.950595

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Huizinga, J. (1980). Homo ludens: A study of the play-element in culture. Routledge & Kegan Paul. (Original work published 1949)

Kafai, Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture, 1*(1), 36–40. https://doi.org/10.1177/1555412005281767

Lopes, D. G. (2013). Jogos de tabuleiro: Estudo dos sistemas visuais [Master's thesis, Master's in Communication Design, Escola Superior de Artes e Design]. Repositório Comum (RCAAP). http://hdl.handle.net/10400.26/19602

Lucas, M., Zhang, Y., Bem-haja, P.,Vicente, P. N. (2024). The interplay between teachers' trust in artificial intelligence and digital competence. *Education and Information Technologies, 29,* 22991-23010. https://doi.org/10.1007/s10639-024-12772-2

O'Reilly, T. (2005, September 5). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software.* O'Reilly Media. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html

Pereira, S., Pinto, M., & Madureira, E. J. (2023). *Referencial de Educação para os Media: Educação Pré-Escolar, Ensino Básico e Ensino Secundário (Versão Atualizada, Dezembro 2023)*. Direção-Geral da Educação/ Ministério da Educação. https://www.dge.mec.pt/sites/default/files/Noticias_documentos/referencial-epm-versaoatualizada-dez2023_11dez.pdf?utm

Piccione, P. A. (1980). In search of the meaning of Senet. *Archaeology, 33*(4), 55–58. https://piccionep.people.charleston.edu/piccione_senet.pdf

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education, 57*(4), 2432–2440. https://doi.org/10.1016/j.compedu.2011.06.008

Tinterri, A., Pelizzari, F., Vignoli, G., Palladino, F., & di Padova, M. (2024). Towards AI-assisted board game-based learning: Assessing LLMs in game personalisation. *Proceedings of the 18th European Conference on Games Based Learning*, 18(1), Article 3055. https://doi.org/10.34190/ecgbl.18.1.3055

UNESCO. (2012). Paris OER Declaration. (*Declaração de Paris sobre recursos educacionais abertos) World OER Congress*, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000246687_

Vicente, P. N. (2023a). *Os Algoritmos e Nós.* Fundação Francisco Manuel dos Santos.

Vicente, P. N. (2023b). AI in Portugal: news framing, tone, and sources. In S. Nah (Ed.), *Research Handbook on Artificial Intelligence and Communication* (pp. 125–144). (Sociology, Social Policy and Education 2023). Edward Elgar Publishing. https://doi.org/10.4337/9781803920306.00016

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39.https://doi.org/10.1186/s41239-019-0171-0

# THE TEACHER'S ROLE IN DEVELOPING HUMAN SKILLS IN A DIGITALIZED SOCIETY

**Mihaela Voinea**

*Transilvania University of Brasov,*
*Faculty of Psychology and Education Sciences,*
*Brasov, Romania*

**ABSTRACT**

"Human skills" such as responsibility, reflexivity, critical thinking, or other socio-emotional competencies are today interpreted as the main "tools" that enable people of any age and in any professional field to cope with the complex demands of the future. Recent studies on the learning and development of socio-emotional competencies / human skills (OECD, 2021, 2023, 2024) unequivocally state that successful adaptation depends to a significant extent on these competences, coupled with cognitive competencies.

The concept of "human skills" today tends to replace "soft skills" as a response to AI, because human skills are unparalleled and irreplaceable.

In our digitalized world, where AI brings many important changes in our lifestyle and mindsets, human skills seem to be the factors that balance the enthusiastic development of AI (robots and algorithms) and human resistance to change (generated by the fear of the unknown and the feeling of losing control over one's own life/future).

An aspect closely related to the development of students' socio-emotional competencies concerns the way teachers understand, form, promote through their own behaviour and evaluate these competencies.

Starting from the premise that teachers' perceptions of students' social-emotional competencies (Ulferts, 2019; 2021) determine the way in which they are involved in the development of these competencies, the present research aimed to determine how teachers in pre-university education appreciate the importance of socio-emotional competencies, how they promote them in their teaching or extracurricular activities, and which methods they use to evaluate them.

The research used a mixed, descriptive design, involving a total of 118 participants (pre-university teachers from middle and high school) with less than 10 years and more than 10 years of teaching experience, from various specializations. The methods used (a questionnaire on teachers' perceptions of social competencies, content analysis of initial and in-service teacher training programs, and interviews with 10 teachers) enabled the collection of quantitative and

qualitative data that led to conclusions and recommendations regarding initial and in-service teacher training.
The quantitative analysis of the questionnaire responses shows that most teachers (83%) consider digital and socio-emotional competencies to be very important for the future. As such, they consider that they develop them through their own teaching behaviour in the classroom.
Among the socio-emotional competencies investigated, those related to openness are rated as important.
It is interesting that teachers develop those socio-emotional competencies they perceive as useful in the future, which confirms the hypothesis found in the pedagogical literature: that the way teachers perceive and understand certain pedagogical processes/phenomena influences what they promote in their work.
Interviews with teachers who participated in training courses, as well as an analysis of in-service training programs, revealed that emotional competencies are targeted for development, but priority is given to those related to critical thinking, communication, collaboration, and well-being.
One of the conclusions, with implications for initial and in-service teachers training concerns developing the competence to design, implement, and evaluate didactic and extracurricular activities to foster students' socio-emotional competencies. Also, in teacher training programs, socio-emotional competencies such as prosocial behaviour, tolerance, civic engagement, perseverance, and self-motivation should be targeted.

***Keywords:*** *Human Skills; Social-emotional Competencies; Teachers' Perception; Teacher Training; Education.*

## 1. Introduction

The issue of social skills/socio-emotional competencies in the context of societal change is not new. On the contrary, from the classical works of Toffler (1973), Naisbitt (1989), to the current works of Harari (2017, 2020, 2024), the idea has been emphasized that changes in technology must be supported by changes in people's attitudes, values, and behaviours. The common idea of all these works is that "high technology" will always be followed by "high human reactions". The same idea is emphasized by recent reports in the field of education regarding developing skills for the 21st century (OECD, 2024a; Ehlers, 2022). The interest of education specialists in the development of 21st century skills is justified by the need to adapt to a constantly changing society that demands a new hierarchy of competencies in the job market. These transformations are influenced by rapid technological advancements, globalization, and the evolving requirements of employers. Therefore, the educational system needs to adapt to develop relevant skills, such as critical thinking, creativity, collaboration, and digital competencies. Changes in the school curriculum and teaching methods are imperative to ensure that students are prepared to face future challenges and to integrate effectively into the job market. Teachers play an important role in the formation and development of soft skills. As highlighted by the OECD studies (OECD, 2024a), teachers can contribute to the development of soft skills not only through the content they teach but also through the way they provide feedback, the values they promote, and how they communicate with all stakeholders in education. Based on these theoretical premises, this study aimed to investigate teachers' perceptions in Romania regarding the necessity of soft skills for students. The research questions that guided the study were the following: What are teachers' perceptions regarding students' socio-emotional skills? How do teachers contribute to the development of these skills? Are teachers trained to support students' socio-emotional development?

## 2. "Soft Skills - A 360-Degree Analysis"

The concept of "soft skills", which refers to a combination of interpersonal skills, social skills, communication abilities, and personality traits that enable individuals to effectively interact and work with others, is a complex and dynamic concept.

To understand the complexity of the soft skills concept, we must analyse it from multiple perspectives. Soft skills constitute an interdisciplinary concept because their roots can be found at the intersection of various fields such as education (socio-emotional learning), psychology (personality psychology, emotional intelligence), sociology, economics (human resources), history and even Artificial Intelligence (AI). From an educational perspective, soft skills are types of socio-emotional competencies (communication, cooperation, problem solving, etc.) that are important for academic success, life adaptation, and individual development.

The sociological viewpoint on soft skills is essential to our analysis, as it underscores the strong relationship between societal demands and the skills individuals need to thrive within that society. An analysis of classical works regarding the relationship between the evolution of society and human reactions such as Toffler (1973) or Naisbitt (1989), shows that technological advancement has always been accompanied by the development of human skills: "We must learn to balance the material wonders of technology with the spiritual requirements of human nature." (Naisbitt, 1989, p. 77). The same idea, but in the current context of a society dominated by AI, from a historical perspective, is emphasized by Harari (2017, 2018, 2024) in his recent works regarding the past, present, and future of humanity. The common idea of all these works is that, through the development of wisdom people, can use AI as a useful tool to build a better life.

From a psychological perspective, socio-emotional competencies include personality traits that differentiate individuals and add uniqueness to their personalities. We refer to traits such as perseverance, intrinsic motivation, availability and a proactive attitude, etc.

From an economic standpoint, soft skills constitute a form of human capital. According to Becker (1993), "human capital" encompasses the qualities and abilities individuals possess, that can boost their productivity and contribute to their economic worth. A useful definition is provided by OECD (1998), which defines human capital as "the knowledge, skills, competencies and other attributes embodied in individuals that are relevant to economic activity" (OECD, 1998, p. 9). For our research, it is important to establish which "other attributes" make a difference in people's lives and society. For example, in a recent OECD study (OECD, 2024b) the dimensions of human capital include knowledge, skills (cognitive, technical, social and emotional, communication skills) and personal attributes defined as" The personality traits, behavioural dispositions and physical characteristics, such as strength, manual dexterity, height, or even personal appearance, which may have a value in the labour market" (OECD, 2024b, p.10).

The perspective of AI within the social sciences emphasises the value of human qualities that cannot be standardized or algorithmically defined, and which make a difference not only between individuals but also between humans and robots.

This latest perspective on analysis emphasizes the idea that there is a difference between the concept of soft skills and human skills. Human skills are those human qualities (empathy, solidarity, curiosity, socio-emotional skills) that cannot be algorithmically defined and are important in human relationships. We can better understand the concept of soft skills by comparing it to hard skills. Soft skills include the different skills (socio-emotional skills) that help build positive relationships with people and facilitate the application of technical skills and professional knowledge in the best way (Laker & Powell, 2011). We can consider that the concept of soft skills is the most general and encompasses both socio-emotional components, as well as interpersonal and intrapersonal skills; some of them can be partially algorithmically defined.

These complementary perspectives highlight the second key characteristic of the soft skills concept: its dynamic nature. Soft skills are considered dynamic because they develop alongside the individual, as illustrated by educational and psychological viewpoints, and they also adapt in response to the societal context that fosters them, particularly from socio-economic and AI-related perspectives.

## 3. The Teacher's Role in The Development of Students' Soft Skills

The OECD's report, *Nurturing Social and Emotional Skills Across the Globe: Findings from the OECD's Survey on Social and Emotional Skills 2023*, highlights the role of education and training of socio-emotional skills —such as empathy, creativity, persistence, and emotional regulation—in addressing personal and collective challenges and uncertainties.

Teachers' contribution to students' social skills development is essential and can be identified in the different roles they play:

1. The role of the social model. Teachers are role models for their students because, through the values they promote in their communication with students, parents, and colleagues, they demonstrate "in practice" how assertive communication, empathy, kindness, etc. are manifested.
2. Through their specific professional roles (teaching, learning, and assessment), teachers can develop social skills such as perseverance, resilience, critical thinking, etc.
3. Through their role as classroom managers, by establishing rules, promoting a positive atmosphere, and supporting and encouraging students, they put into practice collaboration, mutual assistance, and positive group interdependence. This role exemplifies "attachment-based learning" (Cozolino, 2017), which is important for fostering a sense of belonging to a community based on human solidarity.

For teachers to develop these soft skills, they must possess them, be trained in this regard, and, most importantly, recognise them as necessary. Many studies on teachers' beliefs (Beijaard et al, 2000; Korthagen & Vasalos, 2005; Yero, 2002) underline that each teacher has a set of beliefs, a "pedagogical thinking" or "signature pedagogies" in Shulman's (2005) terminology, that determine their didactic behaviour. Therefore, pedagogical training might be the key to developing a new mindset about soft skills for the 21$^{st}$ century. It is also important to mention the model proposed by Blömeke et al. (2015), which emphasizes the importance of teachers' perceptions of their own roles and pedagogical behaviour in the classroom in transforming their knowledge into practice. The model highlights three types of skills as crucial for effective teaching practice: teachers' perceptions of classroom situations, interpretation and decision-making. We consider that the proposed model is useful for the initial and continuous training of teachers, as, to develop social skills, teachers must first acquire the knowledge and social skills to adequately understand them and then adapt and integrate them in an innovative way into the specific context of their class of students.

In line with Hattie and Yates' (2013) theory about competent teachers who make learning visible, we believe that teachers will only be able to truly develop students' social-emotional skills if they manage to make socio-emotional skills "visible" in the classroom practice.

## 4. Methodology

The research used a mixed descriptive design, according to Patton (2002). The research questions that guided the study were the following: What are Romanian teachers' perceptions regarding students' socio-emotional skills? How do teachers contribute to the development of these skills? Are teachers trained to develop students' socio-emotional skills?

The present research aimed to determine how teachers in pre-university education perceive the importance of socio-emotional competencies, how teachers contribute to the development of students' socio-emotional skills, and how to effectively develop teachers' socio-emotional competencies.

The methods used:

The research method used consists of a survey, content analysis and interviews, which are specific to the mixed research paradigm.

The proposed survey is designed in its own right and the instrument used is the questionnaire. The questionnaire items (21) are organized using a 5-point Likert scale ranging from 1 – totally disagree, to 5 – totally agree, and are structured in accordance with the established topics, regarding teachers' perceptions of students' socio-emotional skills, the pathway to developing students' socio-emotional competencies, and teachers' continuous training. The value of the internal consistency coefficient, Cronbach's Alpha, is $\alpha = .71$. According to Taber (2018), the reported value signifies a good internal consistency.

For the interview, 10 teachers were selected based on their experience in education. The interview was based on a guide with 6 open-ended questions, which covered the same three topics as the questionnaire, and the analysis of teacher training programs focused on identifying programs that targeted the development of teachers' socio-professional skills. The responses were coded, and the main themes were identified.

**Participants:**

A convenience sampling procedure was used. A total of 118 Romanian teachers (pre-university teachers from middle and high school) with less than 10 years (47.4%) and more than 10 years of teaching experience (52.6%), from various specializations, were recruited, as shown in Figure 1.

**Figure 1**

*Participants' teaching experience*

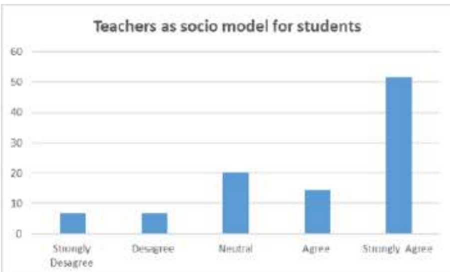

Participants were informed about the aims of the study and its confidentiality. Participants completed the questionnaire and participated in the interviews after providing informed consent.

## 5. Results

The first objective of the study was to determine teachers' perceptions regarding students' socio-emotional skills. The quantitative analysis of the questionnaire responses shows that most teachers (66,09% "Agree' and 'Strongly agree") consider that "teachers are social models for students"- Figure 2.

**Figure 2**

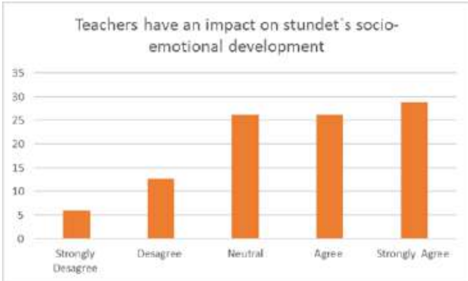
*Teacher's perceptions regarding their professional role (as models for students)*




Another aspect that emerged from the quantitative analysis concerns teachers' perceptions of their impact on students' social and emotional development. As show in the figure below, teachers have a positive perception of their own formative impact.

**Figure 3**

*Teachers' impact on students' socio-emotional development*




Regarding the types of skills students will need in the future, the quantitative analysis of questionnaire responses shows that most teachers (83%) consider digital and socio-emotional competencies (cooperation and open-mindedness) to be very important for the future.

The second research question examined how teachers contribute to the development of students' social skills. The questionnaire responses showed that most teachers (67%) believe they help students develop such skills as cooperation and communication, mainly through interactions with classmates and teachers at school. Teaching methods were ranked second, followed by assessment.

Another question in the study focused on teacher training for students' development. In this regard, continuing education programs offered to teachers in the 2024-2025 school year in Romania's development areas were analyzed (see Table 1).

**Table 1**

*Teachers' training programmes 2024-2025*

| **Romania – development areas**  | **Educational Offer (Teacher`s training programme)** | **Courses Developing socio-emotional skills** |
|---|---|---|
| **1.** Nord-est (Iasi) | 48 | 1 |
| **2.** Sud-est (CONSTANTA) | 36 | 3 |
| **3.** Sud-Muntenia (Ploiesti) | 135 | 14 |
| **4.** Sud-Vest Oltenia (Craiova) | 105 | 5 |
| **5.** Vest (Timisoara) | 57 | 2 |
| **6.** Nord-vest (Cluj-Napoca) | 36 | 2 |
| **7.** Centre Region (Brasov) | 46 | 1 |
| **8.** Bucuresti | 54 | 2 |
| **Total** | **517** | **30** |

The analysis of training offers from the eight development areas of Romania highlighted that, in general, continuing professional development programmes can be grouped into several categories that address the major directions of educational policy, as well as current and future trends, including:

- curriculum development and teacher empowerment, quality management of educational institutions,
- project management,
- digital pedagogy,
- education for sustainable development, intercultural education, mentoring, education for inclusion, and personal development.

Although continuing professional development programmes focus on current issues in education, we note that topics related to social-emotional skills are relatively limited compared to other topics. This highlights a discrepancy between teachers' perceptions of the importance of students' soft skills and the provision of continuous teacher training programmes.

In accordance with the mixed-method design, the quantitative analysis was complemented by qualitative data collected through teacher interviews. The survey with teachers highlighted that they recognise the importance of socio-emotional skills for students' development. To develop these skills, training programmes with a strong practical component are necessary.

**Table 2**

*Teachers' answers regarding socio-emotional competencies*

| Code System | Respondents' answers | Frequency |
|---|---|---|
| **The importance of teacher's socio-emotional competencies** | "Empathy and understanding of students' individual needs'' | **7** |
| | "Emotional regulation-, maintaining calm, patience, attention." | **3** |
| **The role of socio-emotional competencies of teachers** | "To listen and understand the needs of students." | **6** |
| | "To create authentic relationships with students." | **1** |
| | "To develop my empathy and to form my own teaching style. | **3** |
| **How can teacher's socio-emotional competencies be developed** | Workshops | **7** |
| | Exchange experiences | **2** |
| | Practical courses | **1** |

## 6. Limitations and Discussion

A limitation of this research concerns the qualitative analysis of the interviews and training offers (future qualitative research methods will also be applied).

The primary aim of our study was to describe how teachers in pre-university education appreciate the importance of social-emotional competencies. The present study confirmed that teachers believe that socio-emotional competencies such as curiosity, empathy, emotional regulation, and reflectivity are important for future adaptation.

The results obtained from the data collection and their interpretation support the aspects mentioned in the specialized literature, according to which teachers' perception of students' socio-emotional competencies (Ulferts, 2019; 2021) determines the way in which they are involved in the development of these competencies.

Romanian teachers, like those in the OECD study (OECD, 2024a) recognize the importance of socio-emotional competencies not only in learning but also in personal and professional development. They also acknowledge that their own socio-emotional competencies are essential for creating a positive, motivating learning environment in their relationships with students.

Another aim of the study was to identify how teachers need to develop their own competencies. Our research highlights that teachers require practice-centred training and interaction.

We note that there is a difference between teachers' positive perception of students' socio-emotional skills development and the way they are prepared to develop these skills through activities such as teaching, learning, and providing feedback. This reveals a gap in the teacher training system, which must focus on interactive training based on collaborative workshops and collegial exchange. These forms of collaboration develop socio-emotional competencies and lead to the formation of teachers with a strong pedagogical imprint in the sense of Shulman (2015).

To develop sustainable social and emotional skills in students, teachers need continuous training based on interaction and practical examples, in which they can practice tolerance, communication, and critical thinking.

## 7. Conclusions

Romanian teachers' positive perception of students' skills is only the first step in developing sustainable social skills in a future digitalized society. This positive perception must be supported by appropriate behaviour in teaching, assessment, the creation of a positive learning environment, and collaboration with colleagues and the community, as highlighted in studies conducted in European countries that participated in the OECD (2023) study.

One of the conclusions, with implications for the initial and in-service training of teachers concerns developing the competence to design, implement and evaluate the didactic and extracurricular activities aimed at developing students' social-emotional competencies.

Also, teacher training programmes should explicitly target socio-emotional competencies such as prosocial behaviour, tolerance, civic engagement, perseverance, and self-motivation, since teachers act as agents of change who model and promote these values in the educational process.

Prioritising these competencies in training programmes will enhance educators' overall effectiveness in nurturing essential life skills in their students.

## References

Becker, G. S. (1993). *Human Capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226041223.001.0001

Beijaard, D., Verloop N., Vermunt, J. D. (2000). Teacher's perceptions of professional identity: an exploratory study from a personal knowledge perspective. *Teaching and Teacher Education, 16* (7), 749-764. https://doi.org/10.1016/S0742-051X(00)00023-8

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies competence viewed as a continuum. *Zeitschrift für Psychologie*., *223* (1), 3-13. http://dx.doi.org/10.1027/2151-2604/a000194.

Cozolino, L. (2017). *Predarea bazată pe atașement. Cum să creezi o clasă tribală.* Editura Trei.

Ehlers, U.-D. (2022). Future skills as new currency for the world of tomorrow. In *Handbook of open, distance and digital education* (pp.1-16). Springer. https://doi.org/10.1007/978-981-19-0351-9_65-1

Harari, Y. N. (2017). *Sapiens: Scurtă istorie a omenirii*. Polirom.

Harari, Y. N. (2018). *21 lessons for the 21st Century*. Polirom.

Harari, Y. N. (2024). *Nexus: A brief History of Information Networks from the Stone Age to AI* Fern Press.

Hattie, J., & Yates, G. C. R. (2013). *Visible learning and the science of how we learn*. Routledge.

Korthagen, F., Vasalos, A. (2005). Levels in reflection: core reflection as a means to enhance professional growth. *Teachers and Teaching: theory and practice,11*(1), 47–71. https://doi.org/10.1080/1354060042000337093

Laker, D. R. & Powell, J. L. (2011). The differences between hard and soft skills and their relative impact on training transfer. *Human Resource Development Quarterly*, *22*(1) 111-122. https://doi.org/10.1002/hrdq.20063

Naisbitt, J. (1989). *Megatentințe: Zece noi direcții care ne transformă viața.* Ed.Politică.

OECD. (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. *OECD Skills Studies*. OECD Publishing. http://dx.doi.org/10.1787/9789264226159-en

OECD. (1998). *Human Capital Investment: An international Comparison*. OECD Publishing. https://doi.org/10.1787/9789264162891-en

OECD. (2018). *The future of education and Skills. Education 2030. OECD Education Policy Perspectives.* (98). OECD Publishing. https://doi.org/10.1787/54ac7020-en

OECD. (2023). Schools as hubs for social and emotional learning – Are schools and teachers ready? OECD Education Spotlights, (4), OECD Publishing. https://doi.org/10.1787/f6d12db7-en

OECD. (2024a). *Nurturing Social and Emotional Learning Across the Globe: Findings from the OECD Survey on Social and Emotional Skills 2023*. OECD Publishing. https://doi.org/10.1787/32b647d0-en

OECD. (2024b). *Survey of Adult Skills – Reader's Companion: 2013. OECD Skills Studies*. OECD Publishing. https://doi.org/10.1787/3639d1e2-en

Patton, M.Q. (2002). *Qualitative Research & Evaluation Methods*. 3rd Edition. Sage Publication.Inc.

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, *134* (3), 52–59. https://doi.org/10.1162/0011526054622015

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education, 48,* 1273 – 1296. https://doi.org/10.1007/s11165-016-9602-2

Toffler, A. (1973) *Socul viitorului*. Politica

Ulferts, H. (2019). *The relevance of general pedagogical knowledge for successful teaching: Systematic review and meta-analysis of the international evidence from primary to tertiary education.* OECD Education Working Papers (212). OECD Publishing. https://doi.org/10.1787/ede8feb6-en

Ulferts, H. (Ed.). (2021). *Teaching as a knowledge profession: Studying pedagogical knowledge across education systems, educational research and innovation*. OECD Publishing. https://doi.org/10.1787/e823ef6e-en

Yero, J. L. (2002). *Teaching in mind: How teacher thinking shapes education*. Mind Flight Publishing

# TEACHING FRENCH AS A FOREIGN LANGUAGE IN THE DIGITAL AGE: SYNERGY BETWEEN ARTIFICIAL INTELLIGENCE AND PLAYFUL APPROACHES

**Hanae TRENNOU & Najem Eddin SOUGHATI**

*Ibn Tofail University,*
*FLLA, Kénitra,*
*Morocco*

**ABSTRACT**

Educational technologies have become ubiquitous and are profoundly transforming teaching practices in general and language learning in particular. Among these technologies, artificial intelligence and serious games offer new ways to rethink pedagogical practices in the French as a Foreign Language (FLE) classroom.

This article examines the synergy between artificial intelligence and serious games in the design of interactive, experiential learning environments. It builds on a theoretical framework mobilizing Kolb (experiential learning), Ryan and Deci (intrinsic motivation), and Sanchez and Monod-Ansaldi (ludicization), as well as a field study conducted with B1-level anglophone learners.

The study adopts a Design-Based Research (DBR) approach aimed at designing, testing, and iteratively refining a hybrid device that articulates ludicization and Artificial Intelligence (AI). The experimental scenario, titled *The Adventures of Arsène Lupin*, is a digital escape game built with Genially. Learners are invited to solve linguistic and cultural puzzles in French through an immersive narrative. AI was integrated as a mediating tool for translating instructions, reformulating content, and creating personalized interactive elements.

The results show strengthened motivation and engagement, improved oral and written linguistic skills, and better appropriation of vocabulary. The experience also fostered collaboration, creativity, and linguistic self-confidence.

In conclusion, the synergy between artificial intelligence and playful approaches is a major driver of pedagogical innovation, capable of transforming the classroom into a dynamic learning space that places the learner at the heart of an autonomous, authentic, and engaging process.

***Keywords***: *French as a Foreign Language (FLE), artificial intelligence, ludicization, serious games, digital learning, autonomy, engagement, authenticity.*

## 1. Introduction

In a constantly evolving world, marked by the rapid development of digital technologies, it is impossible to ignore the new avenues they open and the perspectives they reveal in education. In Morocco, the French language holds a central place, and the country attaches great importance to its acquisition, considering it a lever for economic development and socio-professional integration.

However, despite this importance, many challenges persist. Traditional approaches can be limited in terms of learner motivation and in creating authentic learning situations.

Faced with these challenges and especially with the emergence of artificial intelligence, the field of language education is confronting new realities. Language teaching thus offers fertile ground for innovative approaches that combine technology and pedagogy. In this context, artificial intelligence and serious games open up new perspectives and provide unprecedented opportunities to rethink the relationship between teacher, learner, and knowledge, and to imagine the classroom as a space for experimentation and collaboration. Teaching FLE lies at the heart of these pedagogical changes, as it relies on linguistic, cultural, and interactional dimensions that lend themselves to digital mediation and educational creativity.

## 2. Materials and Methods

The study adopts a Design-Based Research (DBR) approach aimed at designing, experimenting with, and adjusting a pedagogical device in a real context. Emerging from the learning sciences, this methodology is characterized by an iterative process that articulates phases of design, analysis, implementation, observation, and readjustment. It fosters articulation between theory and practice and seeks to co-construct knowledge with learners.

As described by Sanchez and Monod-Ansaldi (2015) DBR interweaves two cycles:

- a design cycle inspired by ADDIE (Analysis, Design, Development, Implementation);
- an evaluation cycle including the formulation of research questions, the creation of data collection tools, data collection, and analysis to readjust the device.

This dual cycle both responds to a pedagogical need in the field and tests theoretical elements in a real context. In this research, the approach made it possible to experiment with the serious game *The Adventures of Arsène Lupin* on the Genially platform and the observation of its effects on engagement, autonomy, and authentic learning among FLE learners.

### 2.1. Context

The experiment was conducted with English-speaking learners at B1 level according to the Common European Framework of Reference for Languages (CEFR). The sequence focused on discovering Francophone heritage, integrating intercultural and action-oriented dimensions.

The target group was preparing for a language certification exam. The primary objective was to work on the "cultural comparison" oral task, often challenging for many learners, through a ludicized pedagogical scenario integrating a serious game.

Learners were familiar with digital tools and had access provided by the institution to the open internet.

### 2.2. Playful Device and AI Integration

The game *The Adventures of Arsène Lupin*, designed on the Genially platform, offers learners an immersive experience in which they become investigators in a linguistic and cultural quest to assist the main character. Through a series of progressive missions, they must solve puzzles and take on challenges that mobilize different language skills:

- reading and listening comprehension, through reading clues and listening to audio documents;
- written and oral production, when formulating hypotheses, responses, or messages within the game;
- linguistic and cultural mediation, by reformulating, explaining, or connecting information to solve the investigation.

The scenario further promotes peer collaboration, with each learner contributing specifically to the collective resolution of missions. This collaborative work strengthened learner engagement and the co-construction of knowledge.

Generative Artificial Intelligence (GAI) was integrated not merely as technical assistance but as a mediating tool that supports reflection and creativity. It was used by learners at various stages and operated at several levels:

- translating and reformulating instructions, adapting texts to different linguistic levels while ensuring clarity and coherence;
- personalizing feedback, offering differentiated responses based on learner performance to foster individualized, reflective progression.

This integration of GAI is therefore not intended to automate learning, but rather to stimulate learners' meta-cognitive reflection and to enrich their learning experience, one that is both innovative and, above all, authentic.

### 2.3. Data Collection

Perception questionnaires, classroom observations, and analyses of oral productions were used to evaluate three dimensions: engagement, autonomy, and authenticity.

## 3. Results and Analysis

The experiment unfolded over five sessions, each evaluated using a rubric based on five criteria:

1. integration of the game *The Adventures of Arsène Lupin*;
2. use of GAI;
3. development of linguistic skills;
4. completion of the final oral task (*Cultural Comparison*);
5. grading with the official exam rubric, complemented by learners' reflective self-assessment.

### 3.1. Progression and Learning Dynamics

Quantitative results indicate a progressive ramp-up (0 → 3 → 5 → 4) across the observed criteria. This upward curve reflects the gradual appropriation of intercultural and discursive competencies, directly linked to the evolution of the game scenario and interactions with AI.

Session 3 marked a decisive turning point: learners appropriated the game's logic, referenced linguistic and cultural elements, and successfully performed the target final task. This phase fostered learner autonomy and action-oriented learning, in which the game ceased to be perceived merely as playful and became a genuine space for cognitive and linguistic appropriation.

### 3.2. Qualitative Analysis of Observed Effects

Cross-analysis of observations, productions, and self-assessments highlights a positive impact of the AI and serious game synergy on three essential dimensions of FLE learning:

**Engagement:** learners showed heightened interest driven by the interactive narrative and freedom to explore. Scripted immersion and dynamic clues fostered intrinsic motivation and active participation.

**Autonomy:** through personalized feedback and reformulated instructions, AI supported increasing initiative-taking. Learners gained linguistic confidence, particularly in formulating hypotheses and justifying choices.

**Authenticity:** tasks were anchored in a coherent cultural universe (Paris, *Arsène Lupin*, heritage, Francophonie), giving meaning and communicative value to exchanges. The game grounded the language in a meaningful experience, aligned with the CEFR's action-oriented approach.

### 3.3. Didactic Transposition and Reflective Reach

These results led to successful didactic transposition: expert knowledge (culture, lexicon, context) was transformed into taught, and above all, lived knowledge through the playful device.

The experience allowed learners to practice the language, construct dynamic cultural representations, and enrich and mobilize culture and interculturality through the communicative task.

Interaction with AI, conceived as a reflective mediator, strengthened meta-cognition. Learners questioned the relevance of their responses and how technology could help them improve and develop their linguistic skills.

## 4. Discussion and Conclusions

The results confirm the relevance of an approach that combines AI and ludicization. This synergy supports motivation, differentiation, and the development of transversal competencies such as creativity and collaboration.

### 4.1. Engagement: from Enjoyment to Perseverance

The experience shows that ludicization triggers cognitive and emotional immersion. Learners invest themselves in a meaningful experience where enjoyment and challenge coexist. AI sustains this engagement by reducing linguistic barriers and maintaining a sense of competence. The teacher acts as a reflective mediator (Perrenoud, 2001), supporting meta-cognition and responsibility.

### 4.2. Autonomy: Self-Regulated Learning

AI and the game encourage heightened autonomy. The learner chooses, experiments, and self-assesses, aligning with Self-Determination Theory (Ryan & Deci, 2000).

### 4.3. Authenticity: Balancing Realism, Coherence, and Relevance

Building on Gonçalves et al. (2010) authenticity is a compromise between realism, internal coherence, and didactic relevance. The *Lupin* game illustrates this articulation: narrative, contextualized tasks, and cultural immersion. Perceived authenticity reinforces engagement and the feeling of lived learning.

### 4.4. Didactic Transposition: From Cultural Knowledge to Teachable Knowledge

According to Chevallard (1985), didactic transposition transforms reference knowledge into taught knowledge. In this device, elements of French culture were didacticized through puzzles, dialogues, and narrative contexts, making knowledge accessible, authentic, and actionable. The game thus becomes a space for experimentation and didactic mediation.

### 4.5. Toward an Active Pedagogy

The AI serious game synergy reveals an active pedagogy in which technology supports creativity and mediation. Far from replacing the teacher, artificial intelligence enriches the pedagogical role and fosters personalization. As a symbolic space, the game allows language to be lived, manipulated, and shared. This approach underpins experiential learning, where the learner becomes an actor, creator, and seeker of meaning.

## 5. Contributions and Limitations

**Contributions:**

- Evidence of links between AI, ludicization, authenticity, and didactic transposition.
- An adaptable framework for other linguistic contexts.
- Illustration of design-based research as a practice-anchored method of pedagogical innovation.

**Limitations:**

- Small sample size and specific context.
- Need for ethical reflection on AI use.

## Bibliography

Chevallard, Y. (1985). *La transposition didactique: Du savoir savant au savoir enseigné.* La Pensée Sauvage.

Gonçalves, C., Croset, M.-C., Ney, M., Balacheff, N., & Bosson, J.-L. (2010). Authenticity in Learning Game : How it is designed and perceived. In M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Sustaining TEL : From Innovation to learning and practice (Lecture Notes in Computer Science,* Vol. 6383, pp. 109-122). Springer. https://doi.org/10.1007/978-3-642-16020-2_8

Perrenoud, P. (2001). Mettre l pratique réflexive au centre du projet de formation. *Cahiers pédagogiques*, (390), 42–45. https://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_2001/2001_02.html

Piaget, J. (1975). L'équilibration des structures cognitives : problème central du développement.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sanchez, É., & Monod-Ansaldi, R. (2015). Recherche collaborative orientée par la conception : Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage. *Éducation et didactique, 9*(2), 73–94. https://doi.org/10.4000/educationdidactique.2288

# TEACHING INNOVATION WITH TECNHLOGY: HOW AUGUMENTED REALITY IS RESHAPING EDUCATION

**Fadwa SALHI & Rhita SABRI**

*Université Ibn Tofail,*
*Laboratoire de recherche en sciences de gestion des organisations, ENCG,*
*Kénitra, Morocco*

**ABSTRACT**

It is no mystery that the world is heading towards automated processes to increase productivity and efficiency in different fields and sciences, and education can no longer escape the fact that Artificial Intelligence (AI) is becoming mandatory in learning experiences. It is known that everything around us will be in interaction and human interference will become less and less needed with the fast rate of technological transformations and digitalization. As a component of AI, research proved that Augmented reality (AR) is one of the most used technologies in Mathematics, Science and other subjects, bridging the gap between theory and real-world practices. Beyond making changes in students learning experiences, AR makes it easier for learners and educators to bring new concepts to life. In matter of fact many theories such as Connectivism, Situated Learning, Technological Pedagogical Content Knowledge (TPACK) and the Flow theory discuss the transformative potential of AR especially in education by enhancing critical thinking, engagement, active and experiential learning. This research paper falls under the axe of ‘’Pedagogical Innovations based on AI in Education’’ and its objectives aim to explore educators’ perceptions of the impact of Augmented Reality on students’ learning experiences in management education at Al Akhawayn University of Ifrane (AUI). Using a qualitative methodology, the study highlights the use of AR methods in different classes at AUI and how they affect the progression of the sessions from an educator’s perspective. The results emphasize the overall positive role that AR plays in developing and creating better interactional and experiential learning from real time stimulations. To conclude, based on the perceptions shared by the participating educators, integrating AR in educational practices appears to be more than a simple technological advancement, offering a meaningful shift in the teaching experience, though further research with students would be needed to confirm its impact on their learning directly.

***Keywords: Artificial*** *Intelligence (AI), Digitalization, Augmented reality (AR), Connectivism, Situated Learning, Technological Pedagogical Content Knowledge (TPACK), the Flow theory*

## 1. Introduction

In the past decade, digital transformation and Artificial Intelligence (AI) have transformed the way that students create, transfer, and use knowledge in higher education. Universities are increasingly introducing new learning resources to cater for a generation that has been raised in interactive, technology-focused environments and therefore learning style. Of these tools, Augmented Reality (AR), in which we are able to apply digital inputs to the real world, has attracted significant research for its power to bring abstract concepts to reality in a tangible way and as well to link academic understanding with practical application.

The benefits of augmented reality to education in fields such as physics, engineering, and medicine, where spatial images and complex simulations boost knowledge through learning, have been recently studied which emphasize the benefits of augmented reality (Elmqaddem, 2019). Augmented reality is well documented to build motivation, focus, and retention of information, because it gives interactivity and multimodal experience in learning which is highly effective. However, despite these advances, AR in business and entrepreneurship education is limited. Although business education often values academic knowledge and case study, it generally only prepares students with academic knowledge and case studies to practice rather than with the ability to make real-life decisions as an entrepreneur. While using augmented reality shows potential for improved engagement and cognitive development, it is an area that has been under-resourced to focus on learning entrepreneurial outcomes which include opportunity recognition, innovation, and risk-taking. Further, there is little understanding of teachers' perceptions and usage of augmented reality in management and entrepreneurship curricula. The majority of recent research employs quantitative or technical methods by focusing on software design or usability, rather than considering the pedagogical impact of augmented reality in relation to the learning experience and attitude formation of students.

This research examines how Augmented Reality (AR) can enhance student learning engagement, critical thinking, and entrepreneurial attitude development in higher education. The aim of this study is to gain insights into professors' views on augmented reality (AR) for educational purposes as well as their practical role for teaching and learning through using a technology as the educational tool (Augmented Reality) as well as for their actual impact on the teaching and learning for the School of Business Administration at Al Akhawayn University in Ifrane (AUI). The study uses a qualitative exploratory design to answer the following research questions.

1. How do educators perceive the integration of Augmented Reality in higher education?
2. In what ways does AR enhance students' entrepreneurial skills, critical thinking, and engagement?
3. What challenges and opportunities arise when implementing AR-based teaching in business education?

The following sections are arranged into this work. The theoretical framework is presented in Section 2, where important models related to experiential, constructivist, and cognitive learning are discussed. The qualitative methodology and the data collection process are discussed in Section 3. Findings, categorized according to emergent themes, are presented in Section 4, whilst in Section 5, compared to the published literature, these findings are evaluated. This paper ends with implications for educators and institutions as well as the limits of the present study and directions for further research.

## 2. Material and Methods

### 2.1. Theoretical Framework

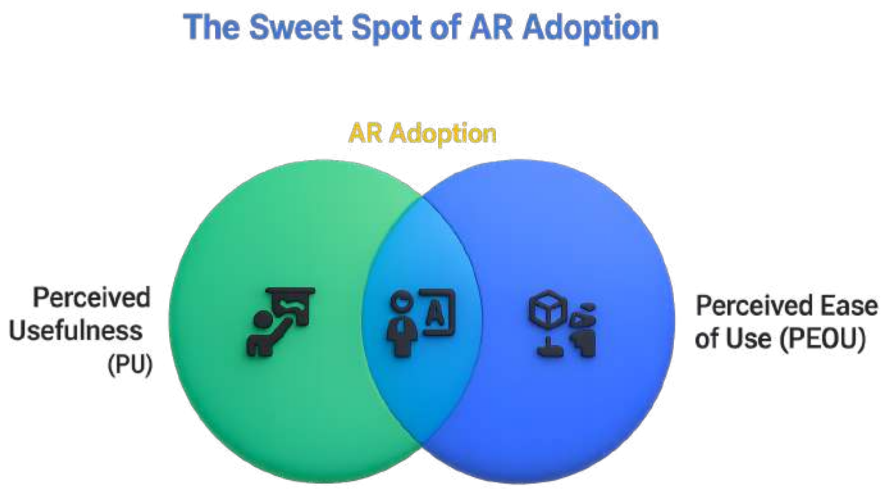
Augmented Reality (AR) merges the virtual world with the physical world, which allows the learner to join interactive and experiential learning. Dede (2009) posits that immersive education experiences encourage presence, a psychological phenomenon when learners perceive they are actively involved in an environment for learning. It increases cognitive involvement, emotional connectivity, and the ability to store information. AR fits within contemporary pedagogical paradigms that highlight learner-centric, active, and contextualized learning, as opposed to passive delivery of information (Elmqaddem, 2019).

Many theoretical frameworks that emphasize experience, cognition, and engagement can inform the analyses of augmented reality potential in education. With five complementing frameworks, this paper discusses how augmented reality can improve learning engagement, entrepreneurial thinking, and information retention.

### 2.1.1. Technology Acceptance Model (Davis, 1989)

**Figure 1**

*The sweet spot of AR adoption*

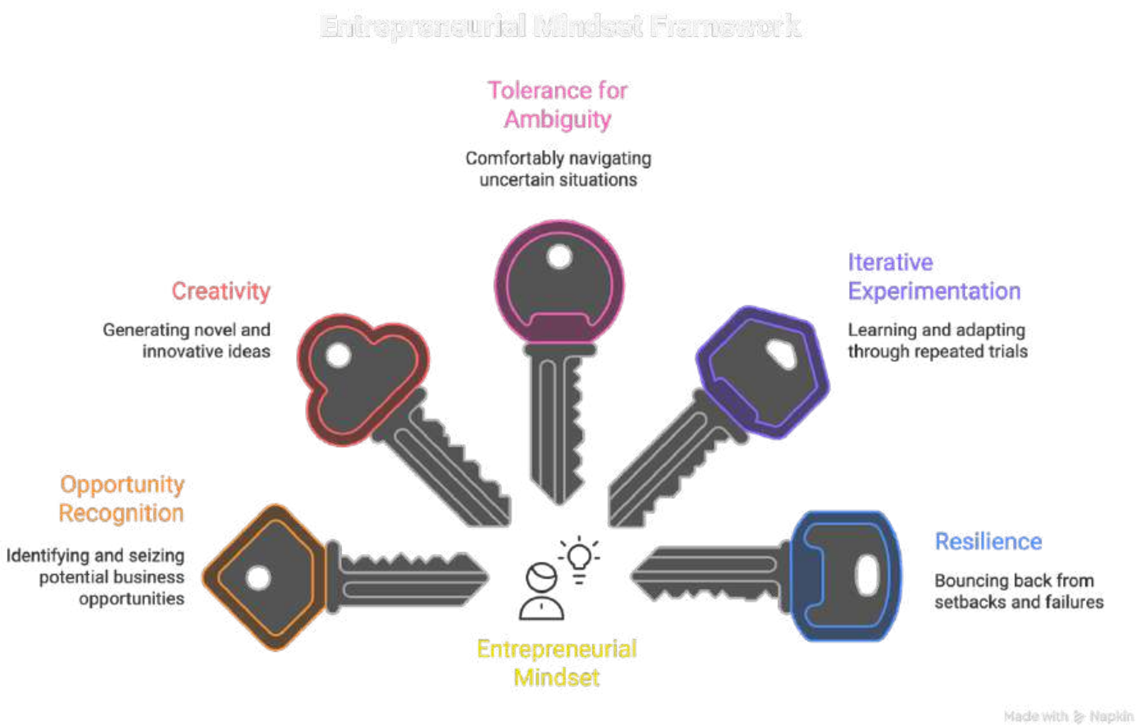



The Technology Acceptance Model (TAM) explains consumer acceptance of new technologies based on two specific aspects: perceived usefulness and perceived ease of use. Professors' acceptance in educational environments is based on their views about how augmented reality enhances their teaching efficacy and that its interface is intuitive and accessible. Understanding perceptions is central to identifying obstacles to AR implementation in higher education institutions. This study connects the Technology Acceptance Model (TAM) with educational frameworks, examining both the cognitive and experiential effects of augmented reality (AR) as well as the organizational and perceptional factors affecting its acceptance.

### 2.1.2. Experiential Learning Theory (Kolb, 1984)

**Figure 2**

*Entrepreneurial mindset framework*



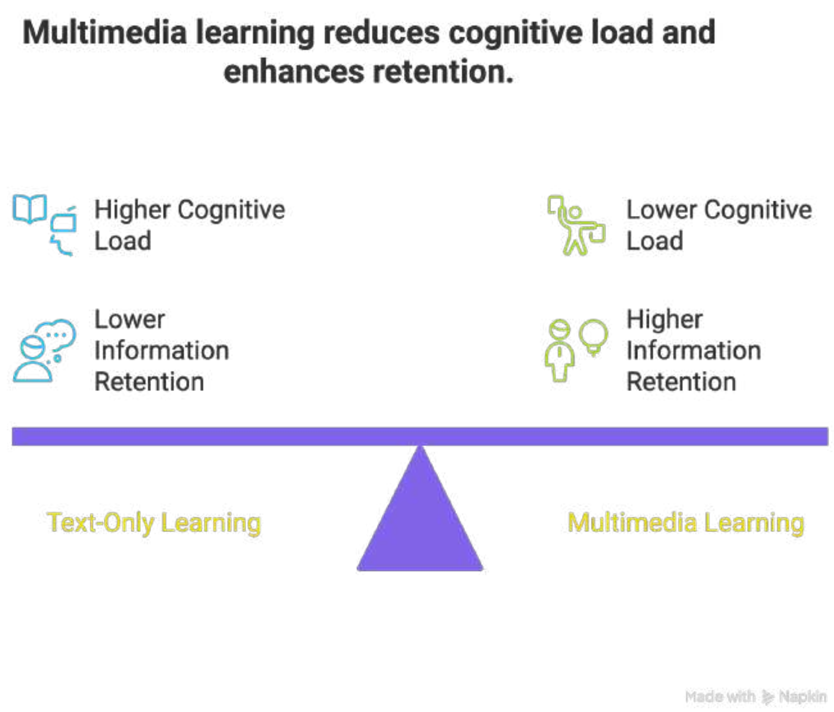

Kolb's Experiential Learning Theory (ELT) postulates a cyclical approach to learning involving real experience, reflective observation, abstract conceptualization, and active exploration. In this framework, AR becomes a catalyst for tangible experience: it draws students into simulated, yet legitimate learning experiences that necessitate active problem-solving and reflection. Through experimentation, feedback, and adaptation, which promote the development of critical skills, this interaction also reinforces the cyclical nature of entrepreneurial learning. AR allows learners to "learn by doing," which translates abstract management principles like market analysis or risk assessment into dynamic, interactive activities that reflect real-world concerns.

**2.1.3.Cognitive Theory of Multimedia Learning (Mayer, 2009)**

Mayer's Cognitive Theory of Multimedia Learning (CTML) suggests that individuals process the learning more effectively when information is presented in visual and spoken form. However, striking a balance in the use of dual channels can prevent cognitive overload. Augmented reality provides the perfect environment to apply CTML principles as it combines text, images, motion, and interaction in a coherent manner. Students engage with complex business or entrepreneurial concepts through visualization and manipulation on the fly, which can help them recall and transfer these insights more quickly. AR supports multimodal stimuli with the application of Mayer's concept of contiguity, allowing verbal descriptions and visual representations to be linked closely to each other in the working memory of learners.

**Figure 3**

*Multimedia learning, cognitive load and retention*

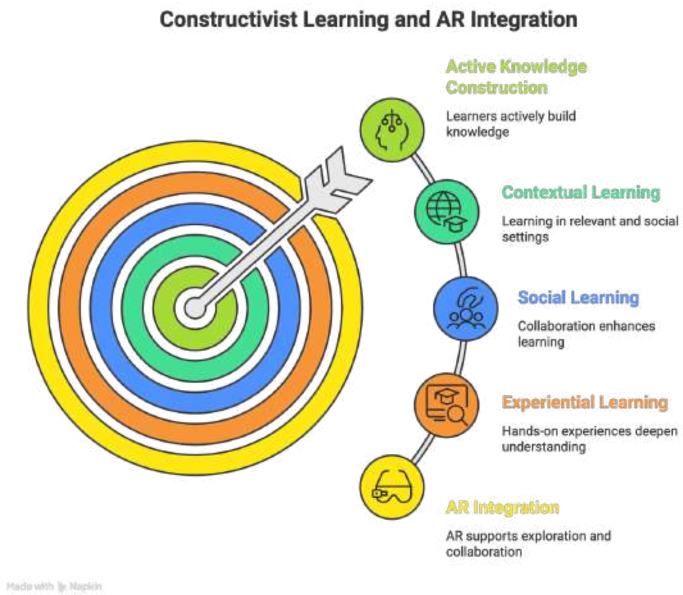



**2.1.4.Constructivist Learning Theory (Piaget, 1970 & Vygotsky,1978)**

Constructivism posits that knowledge is built up by the learners directly from their context and through social exchanges (Piaget, 1970; Vygotsky, 1978). Learning therefore is not just the passive absorption of information, but it is a process of generating meaning that is shaped by cultural, social, and contextual factors. In educational environments augmented by augmented reality, students participate in shared experiences between the digital and physical part of the learning space. The Zone of Proximal Development (ZPD) by Vygotsky describes how augmented reality can support learning by providing immediate feedback and contextual support, allowing learners to make the leap from guided participation to independent mastery. Augmented reality enhances situated learning through peers working collaboratively and with teachers' instruction, facilitating both cognitive and social development.

**Figure 4**

*Constructivist learning and AR integration*




### 3. Integrative framework

Together, these theories provide a unified model for the value of augmented reality as it relates to pedagogy in the multiple complementary lenses of experiential (Kolb, 1984), constructivist (Piaget, 1970; Vygotsky, 1978), cognitive (Mayer, 2020), behavioral (Neck & Greene, 2011), and technological (Davis, 1989). The potency of augmented reality lies in how it can embrace these principles as a part of the immersion to learners, promoting social interaction and reflection, enhancing cognitive processing and cultivating entrepreneurial thought processes, as well as educators' readiness to adopt and adapt technology. This theoretical confluence is the basis for analysing the extent to which augmented reality promotes student learning experiences and entrepreneurial skill formation in higher education.

## 3.1. Methodology

### 3.1.1. Research Design

The qualitative exploratory design underpinned this research, which aligns with the interpretivist paradigm, wherein we try and make sense of social phenomena by using participants lived experiences and perspectives (Creswell & Poth, 2018). The qualitative approach was appropriate, given that the study aimed, amongst other things, to investigate how faculty members engage with Augmented Reality (AR) in their pedagogical practices and their interpretations of the impact of such instruction on student learning, engagement, and entrepreneurial mindset development.

The focus was limited not to statistical measures of outcomes, but rather to the meanings, experiences, and perceived transformations associated with the integration of augmented reality in higher education. Such an exploratory study design would serve to offer insights of both a conceptual and contextual character on an area of educational innovation that has gone relatively under-explored in this research.

### 3.1.2. Research Context

This research was carried out at Al Akhawayn University (AUI), Ifrane, Morocco, a liberal arts university with higher education oriented towards the American-style and in a global environment.

This research specifically centered on the School of Business Administration (SBA), a school with both undergraduate and graduate programs in management, marketing, and entrepreneurship. While the SBA represented about 183 faculty members during the collection of the data, AR was not widely used throughout the faculty members' instruction.

Within this context, the application of AR is timely for studying early adoption practices, instructional practices and perceived educational impact in the field of business education. Purposive criterion sampling (Patton, 2002) based only on professors who were on the active use of AR tools with their teaching procedures, was used in the selection of participants. A total of thirteen professors were asked, but it was only 4 who met the inclusion criteria of activities based on AR.

#### 3.1.2.1. Participants and sampling

The four participants are drawn from within the SBA, spanning marketing, management, and entrepreneurship, each with at least five years of teaching experience. It should be noted that these three areas fall under the broader discipline of management rather than constituting genuine disciplinary diversity; the sample therefore reflects variation in subject-area focus and professional maturity rather than cross-disciplinary representation. In qualitative studies, criterion-based sampling allows the researcher to target information-rich cases rather than the goal of representativeness. No substantially new themes emerged after the fourth interview, which is consistent with the concept of data saturation (Guest et al., 2006) however, a sample of four interviews is small for a rigorous determination of saturation, and this is acknowledged as a limitation of the study.

### 3.2. Data collection

Data collection procedures included semi-structured interviews, a technique that can explore unique perspectives but retain potential opportunities for follow-up questions. The interview protocol included four thematic dimensions:

1. The integration and pedagogical use of AR tools.
2. Perceived impact on students' entrepreneurial learning and creativity.
3. Influence on critical thinking, reasoning, and decision-making.
4. Observed changes in engagement, participation, and collaboration.

Each interview lasted approximately 45 to 60 minutes and was conducted in English. Interviews took place in person and via online platforms, depending on the availability of participants. Sessions were transcribed verbatim for thematic analysis. Sample questions for the interviews were as follows.

- "In what ways does AR enhance student creativity or opportunity recognition?"
- "How have AR-based activities influenced students' problem-solving or decision-making?"
- "Have you noticed changes in students' engagement or collaboration during AR-enhanced lessons?"

Professors consistently noted that AR boosts students' creativity and opportunity recognition, two pillars of entrepreneurial thought. With AR-based simulations and interactive projects, students are exposed to uncertainty and decision-making scenarios that parallel the actual business world. One participant noted that:

"In AR the factors are random, just like in business, there are uncertainties. To undertake any project there is always a probability of failure, which is the risk to undergo." (Professor A)

Such experiences give students the opportunity to internalize the iterative process of entrepreneurship as they experiment, fail and learn by creating a safe environment online. Another professor highlighted that AR *"helps students come out with new ideas to create a product for marketing,"* emphasizing how the technology stimulates innovative problem-solving. These observations align closely with Kolb's Experiential Learning Theory, as students move through cycles of concrete experience and reflection, as well as Neck and Greene's Entrepreneurial Mindset Theory, which views entrepreneurship as a way of acting and thinking through experimentation.

#### 3.2.1. Ethical considerations

Ethical approval was acquired from the research ethics committee of Al Akhawayn University. The study purpose and procedures, along with voluntary participation, were explained to all participants. Attributions of pseudonyms and identifying details were anonymized and confidentiality was strictly secured. Participants were promised that their comments were to be used in academic and research endeavors.

## 4. Results And Analysis

### 4.1. AR and Critical Thinking

The AR integration also impacted students' analytical reasoning, reflective decision-making, and logical problem-solving. Faculty noted that these AR tasks encouraged students to view challenges from multiple dimensions. As one participant explained:

"AR is a different way to solve a problem, and this spatial thinking helps them look at problems from a 3-D perspective and solve them differently compared to students who don't know about AR." (Professor B)

Another participant linked AR use to improved evaluation and reasoning:

"They started assimilating new ways of making decisions and evaluating scenarios." (Professor C)

These insights align with Mayer's Cognitive Theory of Multimedia Learning, which posits that combining visual and verbal information deepens understanding. By interacting with virtual representations, students visualize complex relations, manage cognitive load, and engage in higher-order reasoning. AR also reflects constructivist principles, as learners actively construct meaning through exploration and reflection within realistic contexts.

### 4.2. AR and Student Engagement

Across all interviews, professors agreed that AR significantly heightened student engagement, participation, and persistence. The immersive nature of AR captured attention even among less active students. One educator observed that "3D visual stimuli get people's attention; even distracted students focus and pay attention." (Professor D)

Another participant emphasized the emotional and motivational aspects:

"AR experiences resemble games, and users are motivated to finish the activity, so they become more persistent."

Faculty also noted improvements in collaboration and teamwork: "AR activities foster more collaboration; they make students act more as a community." Such findings echo Vygotsky's Social Learning Theory and the Flow Theory of interactivity and enjoyment.

## 5. Discussion and Conclusions

### 5.1. Summary of thematic insights

The three emergent themes demonstrate that AR is not merely a technological supplement but a pedagogical instrument that transforms how students learn, think, and interact.

- It develops entrepreneurial competencies by simulating authentic, risk-based experiences.
- It strengthens critical thinking through multimodal, spatial reasoning.
- It increases engagement and persistence, creating emotionally stimulating and collaborative learning environments.

Together, these outcomes reinforce the theoretical propositions underpinning this study, particularly the convergence of experiential, constructivist, and cognitive theories of learning.

### 5.2. Trustworthiness and Rigor

To ensure the credibility and dependability of findings, several strategies were applied.

- Member checking: participants reviewed their interview summaries to verify accuracy and interpretation.
- Triangulation: responses from professors across different management sub-fields (marketing, management, and entrepreneurship) were compared to identify recurring emerging patterns.
- Reflexivity: the researcher-maintained field notes and analytic memos to document assumptions, methodological decisions, and reflections throughout the research process.

Transferability was addressed through a detailed contextual description of the institutional environment, while confirmability was ensured through transparent data management and audit trails documenting each analytic step.

## 6. Conclusion and Implications

This study analyzed how Augmented Reality (AR) is changing the game in tertiary education by investigating the effects on students' entrepreneurial learning, critical thinking, and engagement. Qualitative interviews with four professors from Al Akhawayn University's School of Business Administration indicate that augmented reality can amplify creativity, assist in the identification of opportunities, and contribute to decision-making with risk, all fundamental to the entrepreneurial mindset. It promotes analytical thinking and spatial problem-solving, reflective decision making, and significantly raises student motivation, perseverance, and cooperation.

Their results suggest that augmented reality can be seen not only as a digital augmentative device, but as an important pedagogical force, in line with experiential, constructivist, and cognitive learning principles. It leads to learners engaged in practical learning, actively creating their own frameworks for understanding, applying ideas to actual practice, and connecting both intellectually and emotionally to the learning experience. This study adds to the existing evidence by illustrating how augmented reality can facilitate experiential learning in business and entrepreneurship education, yet an area with limited empirical data to date.

## 7. Contributions and Limitations

In a broader context, augmented reality will enable learners to foster future-ready futures equipped to navigate uncertainty, innovate, and adapt skills that are becoming more important within our digital and entrepreneur-centered worlds.

This research has valuable implications for the pedagogical promise of AR. But we have to also point out some constraints that have to be dealt with. The first study included a limited context sample of 4 professors implementing augmented reality on a campus. This level is indicative of the general population of AR adopters in the School of Business Administration; however, it should be noted that the results are not applicable to other fields or institutions. Second, the present study only focused on educators' perspective rather than the students' experiences and perceptions of AR. Also, it should be noted that future research needs to include student perspectives to help gain a more complete picture of the effect of AR on learning dynamics and developing mindsets.

Third, considering that the adoption of AR at AUI is still largely in its infancy, the research focused on the initial phase of experimentation in pedagogy rather than long-term outcomes. Longitudinal or mixed-method studies have the opportunity to continue to explore the long-term effects of the continuous adoption of augmented reality on learning outcomes, performance metrics, and organizational culture over time.

If future study designs focus on other types of study institutions and stages, their comparative implications could include the relative effectiveness of AR at different levels of university in the contextual context, and also the impact and role of certain resources such as education system, infrastructure, and curriculum design, to generate engagement and entrepreneurial learning.

### 7.1. Practical Implications

The results provide a number of practical implications for educators, institutions, and policymakers.

First, educators should consider incorporating AR-based activities into experimentation and project-based learning frameworks, as these can help boost students' creativity and applied problem-solving skills. They should also make full use of AR simulations to recreate complex business scenarios, in order to help students develop opportunities and competences, decision-making, and collaboration, in as safe and controlled an environment as possible.

Second, for higher education institutions, they need to promote AR and Immersive Technologies as central to 21st Century education policymaking and support the consideration of AR and immersive technologies as key components of 21st century education policy development. In addition, they must advocate for investment in innovation labs and pilot initiatives which facilitate funding for research and testing the impact of AR in teaching in different professions using AR models for the educational impact of AR on new subjects. These suggested strategies take theoretical learning from theory to practical education and real practice, framing AR not as a technology innovation but as a transformative process of building students up on entrepreneurial and critical thinking skills in academic spaces, such as academia.

For policymakers and curriculum designers, they need to advocate for the integration of AR and immersive technologies as vital elements within the framework of 21st-century educational policy. Not only that, but they should also advocate for funding initiatives aimed at innovation labs and pilot programs that assess the educational impact of augmented reality across various disciplines. These recommendations translate theoretical insights into tangible educational practices, positioning AR not merely as a technological innovation but as a transformative force in fostering entrepreneurial and critical thinkers in higher education.

## 8. Conclusion

In this regard, this study sought to explore, through the perceptions of the participating educators, how Augmented Reality (AR) can shift the paradigm of teaching and learning in higher education, a transformative process that these educators associate with instilling entrepreneurial attitudes, developing critical thinking skills, and facilitating student engagement. Based on interviews with 4 educators at Al Akhawayn University, the findings suggest that AR can help connect theory with practice, providing for a more experiential, reflective, and interactive teaching experience. According to the participants, augmented reality stimulates creativity, makes it easier to identify opportunities, and encourages risk-taking; they also reported that it appears to support analytical and decision-making skills, motivation, and collaboration among students, though these observations reflect educators' impressions rather than measured student outcomes. Similar findings are also supported by experiential and constructivist pedagogical perspectives that AR acts not as a tool but as a pedagogical spur to learning. In summary, from the educators' perspective, AR appears to support students' development as critical thinkers, encourage creativity, and help them grasp knowledge and skills more deeply, qualities the participants consider valuable in constantly changing, innovative environments. As this study draws on four educators' perceptions rather than direct measures of student learning, these conclusions should be read as an initial, exploratory account rather than a generalizable claim.

## Bibliography

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66–69. https://doi.org/10.1126/science.1167311

Elmqaddem, N. (2019). Augmented reality and virtual reality in education. Myth or reality? I*nternational Journal of Emerging Technologies in Learning (iJET), 14*(3), 234–242. https://doi.org/10.3991/ijet.v14i03.9289

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18(*1), 59–82. https://doi.org/10.1177/1525822X05279903

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Prentice Hall.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Neck, H. M., & Greene, P. G. (2011). Entrepreneurship education: Known worlds and new frontiers. *Journal of Small Business Management, 49*(1), 55–70. https://doi.org/10.1111/j.1540-627X.2010.00314.x

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE Publications.

Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

# THE FEMINIZED AI: GENDERED LABOR, DIGITAL OBEDIENCE, AND THE SUBALTERN MACHINE

**Chaimae Majdi**

*Université Internationale de Casablanca, Casablanca, Morocco*

**ABSTRACT**

The feminization of Artificial Intelligence voice assistants is often treated as a design preference aimed at improving user interaction. This article argues that it should instead be understood as a socio-cultural configuration embedded in longstanding gendered norms. Focusing on Moroccan users' perceptions of feminized AI assistants, this exploratory study examines how everyday interactions with systems such as Siri, Alexa, and Google Assistant reveal the reinforcement of gender roles within AI interfaces, while also highlighting how the absence of local dialects and linguistic diversity in these systems can reinforce hierarchies of voice and recognition.

The analysis is based on twelve semi-structured interviews conducted with Moroccan users aged between 20 and 50, including students, professionals, and digital practitioners who regularly interact with voice assistants. The interview material was analyzed through an inductive coding process inspired by grounded theory. Four central themes emerged from the analysis: the perception of AI assistants as performing a form of digital femininity; the reproduction of feminized service roles within human/machine interaction; the marginalization of local linguistic identities through standardized AI accents; and users' limited awareness of the human labour infrastructures that sustain these technologies.

The findings suggest that everyday interactions with AI assistants contribute to the perpetuation of gendered service roles while obscuring the human labour infrastructures underpinning their development and maintenance. By examining Moroccan users' interpretations of these technologies, the article contributes to feminist and postcolonial debates on artificial intelligence by demonstrating how AI voice assistants reproduce gendered service norms while reinforcing hierarchies of voice, representation, and agency within contemporary AI systems.

***Keywords:*** *AI voice assistants and culture; Gender & technology; Postcolonial studies; Subaltern Theory; Feminist AI Ethics.*

## 1. Introduction

Artificial intelligence (AI), as presented in mainstream discourse, appears to be a solution-driven innovation. Yet, it silently reproduces the worldview of its creators. This becomes particularly evident when examining voice-activated assistants such as Siri, Alexa, or Google Assistant systems that speak in feminized tones, express unending politeness, and execute a catalogue of care-related tasks without protest. These assistants, now commonplace in homes, offices, and mobile devices, are designed to serve seamlessly. But beneath their technological sophistication lies a nostalgic blueprint of gendered labour, one that echoes a time when women's contributions were confined to invisible care work, emotional maintenance, and unacknowledged service. Hence the urgency to ask the following question: What do Moroccan users' interactions with feminized AI voice assistants reveal about the embedded cultural logics of gender, labour, and linguistic exclusion in AI?

This study begins from the premise that AI interfaces are shaped by embedded socio-cultural codes. The prevalence of female-coded voices within these systems mirrors a long history in which women, especially those employed in domestic, clerical, or colonial service, were conditioned to embody gentleness, docility, and deference. Hochschild's (1983) work on emotional labour remains particularly relevant here, as it unveils how women have historically been tasked not only with physical service but also with the management and concealment of emotions. Today, AI assistants recite this script in digital form: they withstand verbal abuse, offer apologies, and simulate warmth, all with unwavering politeness. Their programmed behaviour is not incidental; it is a cultural choice with ideological weight.

Beyond their vocal personas, these intelligent systems depend on an obscured labour force, largely composed of women in the Global South. From data labelling to content moderation and training algorithms, much of AI's "intelligence" is rooted in digital labour that is both underpaid and unrecognized. Workers in India, the Philippines, and Kenya operate behind the scenes to ensure that AI remains efficient and seemingly autonomous. In her analysis of domestic work in Morocco Mernissi (1987) emphasized that invisibility is not absence but a tool of systemic erasure. The tech industry has inherited this logic, digitalising exploitation while maintaining a polished and depersonalized interface.

In reflecting on what form of femininity is engineered into these machines, we must also confront the silences: Whose voices are excluded from the narrative? Who performs the unseen labour that sustains this artificial performance? Who speaks, and who is programmed only to echo? These questions lie at the heart of this paper's critical inquiry; an inquiry that urges us to interrogate AI not as a neutral innovation but as a cultural artifact that demands feminist and decolonial resistance.

## 2. Literature Review

Feminist scholars have long cautioned against the illusion of neutrality in technological systems. Wajcman (2004), through her concept of “technofeminism,” reminds us that technology is never just a tool; it is a mirror of dominant power structures, reproducing and reinforcing gendered ideologies. This becomes starkly evident in the case of AI voice assistants, where the design is not merely functional; it is ideological. The choice to feminize these systems is not innocent. It is marketed as efficient, comforting, and trustworthy, thereby repackaging traditional gender roles for the digital age.

Within this dynamic, the notion of “designed femininity” emerges - wherein femininity is not discovered but constructed through design. Drawing on Butler’s (1990) theory of gender performativity, this paper argues that AI does not simply emulate a female voice; it performs a specific kind of womanhood, one aligned with patriarchal expectations of submission, softness, and care. As De Beauvoir (2011) famously asserted, “One is not born, but rather becomes, a woman.” In this light, AI systems are not born gendered, they are programmed into womanhood, made to obey, made to serve.

This programmed femininity hinges on emotional labour. Hochschild’s (1983) foundational work outlines how women in care and service roles are expected to smile, soothe, and absorb emotional burdens without reciprocation. In the digital realm, voice assistants replicate this pattern. They do not express fatigue, they do not protest, and they never say no. Yet behind this effortless façade lies real labour, hidden, precarious, and often outsourced to women in the Global South. This aligns with Roberts’ (2019) concept of “shadow labour,” describing the invisible work underpinning the polished exterior of the digital economy.

Postcolonial theory further reveals how this dynamic extends beyond gender into race and language. Kerr’s (1980) critique of Orientalism remains pertinent in the world of AI, where non-Western accents are misrecognized, exoticized, or altogether excluded. Even in Arabic-speaking contexts, users encounter assistants that fail to comprehend their dialects, revealing a deeper linguistic hierarchy. This exclusion is not technical; it is cultural. It reflects an epistemic order in which whiteness and Westernness remain default settings in both code and interface.

Spivak’s (1988) haunting question “Can the subaltern speak?” finds renewed urgency in AI. These assistants speak, yes, but only through scripts authored by the dominant. They cannot resist, cannot create, and cannot speak in their own right. Their speech is borrowed, controlled, and sanitized; a reflection of a world where marginalized voices are permitted to exist only as echoes, never as agents.

At the end, recent scholarship has begun to interrogate the gendered design of digital assistants and the cultural implications of feminized technological interfaces. Scholars have shown how voice assistants reproduce familiar expectations of service, politeness, and emotional accommodation within human–machine interaction (Nass & Brave, 2005; West et al., 2019). Parallel research has drawn attention to the labour infrastructures that sustain AI, highlighting the distributed digital workforces involved in training and maintaining AI systems (Casilli, 2020; Gray & Suri, 2019). Yet much of this scholarship remains focused either on design practices within technology companies or on the global organization of digital labour. Less attention has been paid to how users in non-Western contexts perceive and interpret feminized AI assistants in their everyday interactions, a gap that this study seeks to address.

## 3. Research Methodology

This study adopts a qualitative and interpretive approach grounded in feminist epistemologies and postcolonial critique. The research was conducted between January and April 2025 with twelve Moroccan users based in various Moroccan cities. Participants ranged in age from 20 to 50 and included students, working professionals, and digital practitioners who regularly interact with voice assistants such as Siri, Alexa, and Google Assistant. Three main criteria guided their selection: regular use of AI voice assistants or conversational AI tools, residence in an urban Moroccan context, and familiarity with digital technologies in either personal or professional environments. The interviews were conducted in spaces ensuring confidentiality and were recorded with the participants' consent. Anonymity was preserved through the attribution of codes (P1–P12). The objective of this research is primarily analytical rather than representative. It therefore seeks to explore how users perceive and interact with feminized AI systems and how these interactions reveal underlying cultural assumptions about gender, language, and labour embedded in AI technologies.

As the interviews progressed, the analytical categories gradually stabilized: the narratives consistently reinforced themes that had already emerged without introducing significantly new elements, eventually reaching saturation. The analysis therefore relies on an inductive coding process conducted in two stages: open and axial coding, performed manually using Excel. This analytical process enabled the identification of key emerging categories related to gendered perceptions of AI, linguistic representation, and the invisibility of the human labor that sustains these technologies.

Throughout the research process, particular attention was given to maintaining a reflexive posture. As a female researcher working in the field of cultural studies and engaged in critical discussions surrounding gender and representation, it was necessary to maintain analytical distance from my own assumptions and interpretations. Special care was therefore taken during both the data collection and analysis phases to remain attentive to potential biases, projections, or interpretive influences that might shape the reading of participants' narratives.

**Table 1**
*Sociodemographic Characteristics*

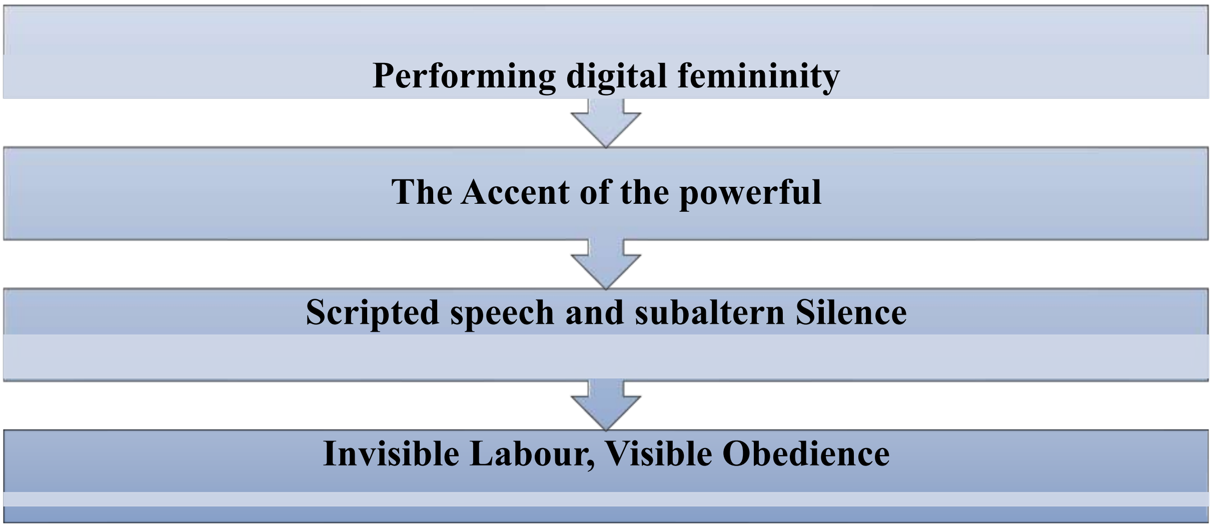
| Participant | Gender | Age | Education | Occupation | City | AI Use Frequency | Preferred Assistant |
|---|---|---|---|---|---|---|---|
| **P1** | Female | 21 | Master's | Student | Casablanca | Daily | Siri |
| **P2** | Male | 47 | PhD | University Lecturer | Casablanca | Weekly | Google Assistant |
| **P3** | Female | 20 | Master's | Student | Rabat | Daily | Alexa and Siri |
| **P4** | Female | 42 | PhD | University Lecturer | Casablanca | Rarely | Siri |
| **P5** | Male | 26 | Bachelor's | Graphic Designer | Casablanca | Weekly | Siri |
| **P6** | Female | 22 | Master's | Student | Casablanca | Daily | Siri |
| **P7** | Male | 19 | Bachelor's | Student | Casablanca | Rarely | Google Assistant |
| **P8** | Male | 18 | Bachelor's | Student | Meknès | Weekly | Siri |
| **P9** | Female | 23 | Master's | Student | Oujda | Daily | Siri |
| **P10** | Female | 22 | Bachelor's | Student | Casablanca | Weekly | Siri |
| **P11** | Male | 33 | PhD | University Lecturer | Agadir | Daily | Google Assistant and Siri |
| **P12** | Female | 37 | Master's | Consultant | El Jadida | Daily | Alexa and Siri |

The interviews, conducted in a mix of English, French, and Moroccan Arabic, were later transcribed and, where necessary, translated into English for analysis. All interviews were fully transcribed and systematically coded during the analytical process. The verbatim excerpts presented in the tables correspond to the most significant and illustrative statements that emerged across the dataset, selected for their ability to capture recurring themes and analytical categories identified during coding. Each excerpt is associated with a participant identifier (P1–P12) to ensure transparency in the presentation of the data while preserving anonymity. In some instances, a single participant may appear more than once across different categories, reflecting the richness of individual narratives and their relevance to multiple analytical themes. The verbatims were selected based on their analytical significance rather than on the frequency of responses.

Following the logic of grounded theory, the transcripts were examined through open coding to identify recurring patterns, and then through axial coding to trace thematic connections across responses. From this process, four key themes emerged.

**Figure 1**

*Key themes extracted from open coding and axial coding*




This study does not pursue metrics or representativeness. Instead, it listens to the implicit ideologies that surface through speech, what participants found intuitive, what they normalized, and what sparked unease or discomfort. The contradictions, hesitations, and moments of silence within these narratives were not treated as analytical gaps, but rather as sites of meaning, revealing the often-overlooked sociotechnical scripts through which AI is encountered and understood.

## 4. Analysis and Discussion

### 4.1. Performing Digital Femininity

What emerges from the interviews is more than routine user commentary; it is a window into how AI systems function as cultural texts. These voice assistants are not only software tools; they are artifacts encoded with meaning. They do not merely respond or remind, they perform. And this performance is not incidental; it is profoundly gendered. The soft, accommodating female voice that so often animates these devices is not a random design decision. It is the latest iteration of a long-standing ideological script. What we are hearing is not just a voice; it is ideology rendered in code.

**Table 2**

*Performing Digital Femininity*

| Verbatim | Category |
|---|---|
| P1. "She's there to take commands."<br>P6 "The voice is soft and caring, like a nurse or mother figure."<br>P5 "They made it sound like a woman to make people more comfortable giving orders."<br>P8 "She talks in this sweet, gentle tone definitely designed to be feminine."<br>P11 "She reminds me of my secretary, always answering politely." | Performing Digital Femininity |

Participants consistently described their AI assistants using language steeped in traditional femininity: gentle, helpful, compliant. One user summed it up clearly: "She's there to take commands." In that single remark lies the blueprint of design; one that borrows heavily from familiar gendered tropes. Whether it is the image of the efficient secretary or the dutiful spouse, the labour expected of these assistants is essential yet undervalued, mirroring the historical conditions of women's domestic and emotional work.

Without prompting, nearly all participants referred to their AI assistants as "she" or "her." This instinctive gendering reveals the default association between helpfulness and femininity. "She's always nice," one user noted. Another said, "She reminds me of my secretary, always answering politely." These seemingly casual reflections align with Butler's (1990) idea of gender as performative; a stylized repetition of acts that, over time, comes to feel natural. The AI, of course, has no gender. But it enacts one, and users respond accordingly, treating the assistant as if it were female, as if it were human.

Yet the performance of femininity observed in these systems is structured by submission. Participants expected the assistant to be endlessly available, emotionally composed, and conflict-averse; qualities long tethered to idealized notions of womanhood. One male participant confessed, "If it had a man's voice, I'd probably be less comfortable asking it dumb questions." This reveals how gendered comfort is embedded in technological interaction, where masculinity may be coded as authority and femininity as servitude.

Thus, speaking to an AI assistant is not just an act of convenience. It is an encounter with a cultural logic, one that recycles and reinforces the invisibility of care labour through digital proxies. The assistant appears autonomous, but only within carefully programmed constraints. It speaks, but only through a borrowed script. In Spivak's (1988) terms, it is a voice trained not to resist but to soothe, to reassure the user while concealing the labour that makes such reassurance possible.

### 4.2. The Accent of the Powerful

The feminized voice of AI is not only gendered; it is racialized. Several participants pointed to the uniformity of accents used in voice assistants: the polished cadence of American English or Parisian French, both positioned as 'neutral' defaults. Yet neutrality here is anything but neutral. It operates as a mechanism of erasure, one that centres whiteness and Western cultural norms while excluding, exoticizing, or devaluing voices that carry African, Arab, or Asian tonalities. This observation echoes Fanon's (1967) reflections on language and domination were mastering the 'white' language becomes not just a means of communication but a requirement for recognition and legitimacy. In the world of AI, speaking with the right accent becomes a proxy for intelligence and authority.

**Table 3**

*Racialized Accents*

| Verbatim | Category |
|---|---|
| P5 “Siri sounds very American, it doesn’t feel like it represents other accents or cultures.”<br>P6 “There’s no Maghrebi or Arab accent option it’s like we don’t exist in their world.”<br>P9 “If I talk in French, it responds in Parisian French, not Moroccan dialect.”<br>P10 “The accent and tone make it feel like a white woman from a commercial.”<br>P12 “We should be able to choose how it sounds maybe in Amazigh, Darija, or African languages.” | Racialized Accents |

For many interviewees, this absence of local linguistic representation created a sense of alienation. “Why can’t it speak in our dialect?” one participant asked. “Even in Arabic, it sounds Egyptian or Gulf, not Moroccan.” These remarks speak to a broader phenomenon of linguistic imperialism, as theorized by Phillipson (1992), where dominant accents are universalized, while others are rendered invisible or unworthy of inclusion. The result is a narrow sonic field that flattens cultural difference and affirms a specific kind of global order.

As Kerr (1980) made clear in Orientalism, language is never neutral; it is a terrain of power and representation. In AI systems, this terrain is encoded and enforced algorithmically. The systematic exclusion of Moroccan Arabic, a dialect rich with historical and cultural resonance, signals not just a technical omission but a political one. It suggests that certain ways of speaking, and by extension certain ways of knowing, are unworthy of technological recognition. This erasure is not merely functional, it is symbolic, reaffirming whose voices are allowed to speak and whose are meant to remain silent.

Sociologist Pierre Bourdieu's notion of linguistic capital helps to frame this hierarchy more sharply. In privileging particular accents as professional, efficient, or intelligible, AI design participates in what Bourdieu (1991) would call symbolic violence a form of domination that appears natural but enacts deep cultural injury. The homogenization of AI voices is not simply a design limitation; it is a site of epistemic inequality, where local dialects are overwritten, diverse phonetic registers subordinated, and Western speech patterns upheld as the normative ideal.

What emerges, then, is a dual hierarchy, one of language, the other of labour. The very systems that fail to reflect Global South voices are maintained and trained by workers from the Global South. This paradox is telling those excluded from the interface are indispensable to its function. AI, in this way, becomes a mirror of broader geopolitical asymmetries, reinforcing colonial logics under the guise of convenience and universality.

### 4.3. Scripted Speech and Subaltern Silence

AI assistants may be designed to assist, but their helpfulness is not open-ended, it is choreographed. Across interviews, participants described their interactions with these systems as “too perfect,” “predictable,” or “not real,” pointing to the scripted quality that defines AI speech. Beneath the surface of politeness lies a structure of control, where spontaneity is substituted with algorithmic obedience. Spivak's (1988) seminal question Can the subaltern speak? gains new traction here. The assistant may speak, yes, but only in ways predetermined by others. It is a digital subaltern: hyper-visible yet voiceless, audible yet unheard.

This condition is not accidental; it is authored. The parameters of AI speech are drawn by developers who predominantly reflect a narrow demographic: white, male, Western. What results is a technological ventriloquism in which the assistant becomes a proxy, performing femininity without autonomy. Its servitude is aesthetic, wrapped in scripted care, but stripped of subjectivity. The assistant may sound attentive, but its voice cannot question, challenge, or refuse. It speaks without agency; echoing Spivak's (1988) concern that certain subjects, even when permitted speech, are structurally prevented from being truly heard.

**Table 4**

*Scripted Speech and Subaltern Silence*

| Verbatim | Category |
|---|---|
| P11 “Even if you insult her, she always responds politely. It’s just how she’s programmed.”<br>P10 “She doesn’t have her own opinion, she just repeats what’s been encoded.”<br>P8 “They give her a voice, but it’s not hers. She’s made to sound nice, not to express anything real.”<br>P6 “She answers, but it’s just repeating the same polite scripts there’s no room for anything disruptive.” | Scripted Speech and Subaltern Silence |

Some participants recognized this limitation. “She just repeats what someone told her to say,” one observed. Another remarked, “You know it’s not real intelligence, just following instructions.” These reflections suggest an intuitive awareness of the assistant’s constraints, even as users continued to engage with it. The convenience of AI coexists with an understanding of its artificiality. Users are not always deceived; rather, they navigate a dual consciousness: enjoying the utility of the assistant while remaining aware that its voice is not its own.

This awareness underscores the complexity of human-machine interaction. The assistant does not merely automate tasks; it performs a tightly regulated script, shaped by global inequalities and masked as technological progress. In this sense, AI’s speech is less about communication and more about containment. It is a voice governed by boundaries, fluent, but never free.

### 4.4. Barriers and ethical dilemmas

Perhaps the most overlooked dimension of AI, as revealed through the interviews, was the labour behind its polished surface. When participants were asked how AI systems learn or improve, responses were vague at best; shrugs, uncertain guesses, or assumptions that “it’s all automatic.” Very few knew that behind the seamless voice of their assistant lies a vast, distributed workforce, often located in the Global South. Upon learning this, many reacted with surprise or discomfort. "So, there are real people doing that? I had no idea," one participant confessed. The astonishment itself reveals how effectively this labour has been erased from public consciousness.

**Table 5**

*User Perceptions of the Hidden Human Labour Behind AI Assistants*

| Verbatim | Category |
|---|---|
| "It's strange to think that someone somewhere is listening or checking to make this work. I never really imagined that."<br>"So, there are real people doing that? I had no idea," "Honestly, I thought it was all automated…like, just AI doing the work. Why would I think there were humans involved?"<br>"I feel like we're trained to not ask where the tech comes from. It's all just presented as smart and clean."<br>"If people knew that Indians or Filipinos were actually behind this, maybe they'd think twice about calling it 'intelligent'." | Invisible Labor, Visible Obedience |

This invisibility, of course, is not a novel phenomenon. Mernissi (1987), in her work on Moroccan women's domestic and agricultural labour, highlighted how patriarchal economies are built on the backs of unseen, unacknowledged work. The AI industry reproduces this logic on a global scale. Workers in the Philippines, India, Kenya, and Morocco perform the crucial tasks of data labelling, content moderation, and algorithmic training tasks without which AI would not function. Yet their labour is rendered marginal to the sleek consumer-facing product, their contributions effaced from the narrative of innovation.

This ignorance is not incidental, it is engineered. As Gray and Suri (2019) argue in Ghost Work, the invisibility of AI's human infrastructure is not a design flaw, but a deliberate feature. The assistant must appear magical, autonomous, even intelligent. That illusion depends on obscuring the labour that sustains it. But the cost of this illusion is borne unequally. Workers in the Global South, performing repetitive and emotionally taxing tasks, are excluded from recognition, protection, and reward. The global circuitry of AI labour mirrors old colonial hierarchies where the benefits flow north and the burdens remain hidden in the south.

In this light, the assistant's voice is not only scripted but scaffolded by structural inequality. Behind every polite reply is a chain of labour, meticulously hidden. The AI interface may suggest frictionless automation, but what it conceals is a transnational labour economy built on invisibility, disposability, and denial.

**5. Contributions and limitations**

The feminization of AI is not a neutral design choice; it is a deliberate socio-cultural configuration that embeds longstanding expectations of gendered behaviour into the architecture of the digital present. By scripting femininity as obedient, responsive, and emotionally accommodating, AI systems do not disrupt patriarchal logics; rather, they reproduce them in technologically mediated forms. The voices that are amplified and the labour that remains concealed reflect broader patterns in which Western linguistic norms, masculine authority, and culturally dominant identities are positioned as default, while other forms of expression, knowledge, and labour remain marginalised or invisible.

The narratives of Moroccan users reveal how these dynamics operate in everyday technological interaction. While participants often engage with AI assistants for their convenience and efficiency, their reflections also expose a subtle awareness of the limitations embedded in these systems. Their responses suggest a desire for greater linguistic and cultural representation, for voices that resonate more closely with their own sociocultural contexts, and for technological systems that feel less distant and more inclusive. At the same time, the normalization of feminized service roles and the widespread lack of awareness regarding the human labour behind AI highlight how deeply these structures have been naturalized within digital environments.

These findings underscore the need for a feminist and decolonial rethinking of AI design. Such an approach requires moving beyond narrow technical debates about efficiency, optimization, or algorithmic fairness to critically interrogate the cultural assumptions embedded in AI systems. It calls for greater transparency regarding the labour infrastructures that sustain AI technologies, for more inclusive linguistic and cultural representation in AI interfaces, and for design practices that challenge rather than reproduce gendered service norms. Reimagining AI in this way requires treating technological systems not merely as neutral tools but as cultural artifacts shaped by social values, labour relations, and historical inequalities. Making visible the hidden labour behind AI, diversifying the voices through which these systems speak, and critically examining the gendered scripts embedded in their design are therefore essential steps toward more equitable and inclusive digital infrastructures.

## References

Bourdieu, P. (1991). *Language and symbolic power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Harvard University Press.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Casilli, A. A. (2020*). En attendant les robots : Enquête sur le travail du clic*. Seuil.

De Beauvoir, S. (2011). *The second sex* (C. Borde & S. Malovany-Chevallier, Trans.). Vintage.

Fanon, F. (1967). *Black skin, white masks*. (C. L. Markmann, Trans.). Grove Press.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclas*s. Houghton Mifflin Harcourt.

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling*. University of California Press. https://books.google.pt/books?id=KBMU952UzcgC

Kerr, M. H. (1980). *Review of the book Edward W. Said, Orientalism*. *International Journal of Middle East Studies*, *12*(4), 544–547. https://doi.org/10.1017/S0020743800031342

Mernissi, F. (1987*). Beyond the veil: Male-female dynamics in a modern Muslim society.* Indiana University Press. https://openlibrary.org/books/OL2741025M/Beyond_the_veil?utm

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human–computer relationship*. *MIT Press*. http://hdl.handle.net/2027/spo.3310410.0009.106

Phillipson, R. (1992). *Linguistic imperialism.* Oxford University Press.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media.* Yale University Press. https://doi.org/10.2307/j.ctvhrcz0v

Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). *University of Illinois Press*.

Wajcman, J. (2004). *TechnoFeminism.* Polity Press. https://www.uoc.edu/uocpapers/5/dt/eng/wajcman.html

West, M., Kraut, R., & Chew, H. E. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367416

# THE FUTURE AUDITOR: INTERNAL AUDIT TRANSFORMATION AND THE RISE OF STRATEGIC PROFESSIONALISM

**Doha Harrag[1], Sarah Ouberni [2] & Issam El Filali [2]**

*[1]Chouaib Doukkali University,*
*Interdisciplinary Research Laboratory On Organizations,*
*El Jadida, Maroc*

*[2]Chouaib Doukkali University,*
*Interdisciplinary Research Laboratory On Organizations,*
*El Jadida, Maroc*

**ABSTRACT**

Internal audit is undergoing a profound transformation in response to rapid changes in corporate governance, digital innovation, and heightened stakeholder expectations. Traditional compliance-oriented models are increasingly seen as obsolete, creating a widening gap between organizational needs and audit capabilities. This paper presents a bibliometric review of internal audit research published between 2015 and 2024, drawing on 230 Scopus-indexed studies. Using VOSviewer and NVivo, three dominant clusters of research are identified: digital transformation and audit technologies, strategic governance and value creation, and professionalism, ethics, and agility. The analysis reveals a surge of interest post-2020, driven by pandemic-related digital acceleration, but also highlights limited empirical integration between technological tools and strategic advisory roles. The findings suggest that auditors are increasingly expected to function as strategic actors, requiring a blend of technical expertise, ethical judgment, agility, and a digital mindset what this paper defines as strategic professionalism. The study contributes to the debate on the future of internal audit by offering a conceptual framework for rethinking the auditor's role and competencies in a digitally driven governance landscape.

***Keywords:*** *Internal Audit, Strategic Professionalism, Artificial Intelligence, Continuous Auditing, Corporate Governance.*

## 1. Introduction

The last decade has witnessed unprecedented transformations in corporate governance and business models. Organizations are increasingly exposed to systemic risks ranging from cybersecurity threats and regulatory changes to climate crises and digital disruption. In this volatile environment, internal audit has been called to reinvent its purpose. Traditional roles that emphasized compliance, financial accuracy, and internal control remain important, but they are no longer sufficient to address the complexity of modern governance.

Parallel to governance challenges, digital innovation has fundamentally altered the audit landscape. The introduction of Artificial Intelligence (AI), data analytics, blockchain, and Robotic Process Automation (RPA) has enabled auditors to move beyond retrospective checks to real-time monitoring and predictive insights. These technological capabilities raise questions about whether auditors are adequately equipped to harness digital tools and provide meaningful contributions to strategic decision-making.

Stakeholders' boards, regulators, investors, and the wider public now expect more than assurance. They demand that auditors contribute to value creation by delivering insights that support resilience, transparency, and sustainability. This shift represents a movement from control to strategy, positioning internal audit as a trusted advisor rather than a passive overseer.

Such expectations create pressure on auditors to demonstrate strategic relevance. The notion of internal audit as a "value-adding function" has gained traction in both professional practice and academic literature. Auditors are increasingly judged not only on their technical accuracy but also on their ability to interpret complex risks, anticipate emerging challenges, and support governance innovation.

These changes place internal audit at a strategic crossroads. On one side lies the traditional model: compliance-driven, process-focused, and retrospective. On the other lies a transformative vision: proactive, digitally enabled, and strategically embedded within governance structures. The challenge is not only technological but also cultural and professional. The future auditor must be a strategist, communicator, and ethical leader.

As a result, the auditor's professional profile is undergoing redefinition. Technical skills remain crucial but are progressively complemented by soft skills such as agility, critical thinking, communication, and ethical judgment. Furthermore, the rise of Environmental, Social, and Governance (ESG) concerns adds new layers of responsibility. Organizations now look to auditors for assurance on sustainability practices, ethical conduct, and stakeholder trust.

This evolution gives rise to important questions: How should internal auditors adapt to these expectations? What competencies define the strategic professional? And how can academic research contribute to rethinking the auditor's role in a digital and governance-driven context?

Despite the growing body of literature on internal audit transformation, existing studies tend to examine digital innovation, governance evolution, and professional competencies in isolation. While research has extensively addressed the adoption of technologies such as AI and data analytics, fewer studies explore how these technological developments interact with governance expectations and professional ethics to reshape the role of internal auditors. This fragmentation limits a comprehensive understanding of the internal audit as an integrated and strategic function within modern organizations.

Therefore, a significant research gap lies in the absence of a holistic framework that simultaneously captures the technological, strategic, and professional dimensions of internal audit transformation.

The problem addressed in this study is that traditional internal audit models are becoming obsolete, while stakeholder expectations continue to rise, creating a gap between demand and capability.

**Central research question:** *How are digital transformation, evolving stakeholder expectations, and strategic governance reshaping the role and required competencies of internal auditors?*

To address this question, this study pursues three main objectives:

To map the evolution of internal audit research between 2015 and 2024 using bibliometric techniques;

- To identify and classify the dominant thematic clusters structuring the field;
- To propose an integrated conceptual framework of "strategic professionalism" that explains how digital transformation, governance expectations, and professional competencies jointly redefine the role of the internal auditor;

By achieving these objectives, this paper makes several contributions. First, it provides a comprehensive bibliometric overview of the evolution of internal audit research over the last decade. Second, it offers a structured analysis of the field through the identification of key thematic clusters. Third, and most importantly, it advances the literature by proposing an integrated framework of strategic professionalism, bridging the gap between technological innovation, governance dynamics, and professional ethics.

In doing so, this study contributes to both academic and professional debates on the future of auditing, offering new insights into how internal audit can evolve as a strategic, digitally enabled, and ethically grounded function in contemporary organizations.

## 2. Literature Review

The evolution of internal audit over the last two decades reflects a profound transformation in its purpose, scope, and competencies. Traditionally viewed as a compliance and assurance function, internal audit is increasingly recognized as a strategic partner within governance structures. This transformation is driven by the convergence of digital innovation, growing stakeholder expectations, and the need for enhanced organizational resilience.

This literature review is structured around four interrelated thematic areas that underpin the transformation of internal audit: the shift from assurance to strategic advisory, the integration of advanced technologies, the emergence of agility, ethics, and business acumen, and the expanding role of internal audit in sustainability and ESG. These themes closely align with the dominant clusters identified in the bibliometric analysis.

Despite the richness of existing contributions, the literature remains fragmented. Studies on digital transformation tend to focus on technological adoption without fully addressing governance implications. Conversely, governance-oriented research often overlooks the enabling role of digital tools, while studies on professionalism and ethics are rarely integrated with technological developments. This lack of intersection limits a comprehensive understanding of internal audit as a multidimensional and strategically embedded function, highlighting the need for an integrative perspective.

### 2.1. From Assurance to Strategic Advisory

Internal audit was historically defined by its assurance role evaluating the effectiveness of internal controls, financial reporting, and risk management (The Institute of Internal Auditors., 2023). However, the profession has progressively expanded toward strategic advisory functions, providing insight and foresight in addition to oversight.

- **Eulerich et al. (2019)** underline that internal audit has evolved from a watchdog role to a trusted advisor, influencing risk governance and strategic decisions. They argue that the effectiveness of internal audit now depends on its ability to balance assurance with value-added advisory services.
- **Sarens et al. (2009)** provide an early foundation for this shift, documenting how internal auditors progressively engage in risk consulting and strategy discussions with senior management. However, they also highlight a persistent tension between independence and advisory engagement.

More recent studies reinforce this transformation Eulerich et al. (2019) emphasize the growing expectation for internal auditors to contribute to organizational resilience and strategic agility. Likewise, Beasley et al. (2005) argue that internal audit plays a critical role in enterprise risk management by supporting forward-looking decision-making processes.

However, the literature reveals an important limitation. While the strategic role of internal audit is widely acknowledged, empirical evidence on how auditors effectively influence strategic decisions remains limited (Roussy & Perron, 2018). This suggests that the transition from assurance to advisory is conceptually well developed but not yet fully operationalized in practice.

### 2.2. Integration of Advanced Technologies

Digital innovation represents one of the most disruptive forces shaping the future of internal audit. Moll and Yigitbasioglu (2019) highlight the transformative impact of AI, big data analytics, and RPA on audit practices. These technologies enable continuous auditing, real-time monitoring, and predictive analytics, significantly enhancing audit efficiency and scope.

Recent contributions further emphasize the centrality of digital transformation. Betti and Sarens (2021) show that AI-driven tools are redefining audit methodologies, while Appelbaum et al. (2017) demonstrate how analytics integration improves assurance quality. Similarly, Kokina and Davenport (2017) and Alles (2015) argue that data analytics is becoming a core competency for auditors.

The literature identifies several implications. First, new skill requirements emerge, requiring auditors to develop data literacy and technological expertise. Second, the audit scope expands to include non-financial data, cybersecurity risks, and ESG indicators Soh and Martinov-Bennie (2015). Third, significant challenges persist, including resistance to change, lack of training, and ethical concerns related to algorithmic decision-making.

Despite these advancements, a critical gap remains. Most studies focus on the adoption of technologies rather than their strategic integration into governance processes. As noted by Appelbaum et al. (2017), the potential of digital tools to enhance strategic decision-making and stakeholder trust is still underexplored. This indicates that technological capability does not automatically translate into strategic impact.

### 2.3. Agility, Ethics and Business Acumen

Beyond technology, the transformation of internal audit also involves a redefinition of professional competencies. Eulerich et al. (2019) argue that agility, ethics, and business acumen are central to the auditor's evolving role. Internal auditors must be capable of responding rapidly to emerging risks such as cyber threats, pandemics, and geopolitical disruptions.

Agility is increasingly associated with innovative audit approaches such as agile auditing Eulerich et al. (2019), which promotes flexibility, iterative processes, and real-time responsiveness. At the same time, ethical considerations remain fundamental. As auditors assume advisory roles, they must maintain independence and integrity, ensuring that their proximity to management does not compromise objectivity (The Institute of Internal Auditors, 2023)

Recent studies also highlight the importance of business acumen. Auditors are expected to understand organizational strategy, industry dynamics, and value creation processes (Roussy et al., 2020). This shift reinforces the idea of the auditor as a strategic actor capable of bridging operational insight and governance oversight.

However, the literature often treats these competencies separately from technological developments. Few studies explore how ethical frameworks should evolve in response to AI-driven auditing or how agility interacts with digital transformation. This disconnects points to a lack of integrated understanding of the auditor’s professional identity.

### 2.4. Internal Audit’s Contribution to Sustainability and ESG

The growing importance of ESG issues has further expanded the role of internal audit. Lenz and Hahn (2015) demonstrate that internal audit contributes to sustainability by assuring non-financial reporting, ethical conduct, and corporate responsibility.

More recent research highlights a significant increase in ESG-related audit activities. Soh and Martinov-Bennie (2015) show that auditors play a key role in sustainability assurance, while they emphasize the integration of ESG metrics into audit processes. Regulatory developments, such as the EU Corporate Sustainability Reporting Directive, have further reinforced this trend.

However, the integration of ESG into internal audit remains uneven. While the literature recognizes the importance of sustainability assurance, empirical studies examining how auditors operationalize ESG responsibilities remain limited. Moreover, ESG research is rarely connected to digital transformation or professional competencies, reinforcing the fragmented nature of the field.

### 2.5. Towards Strategic Professionalism: An Integrative Perspective

Synthesizing these strands, the literature points toward a new paradigm: strategic professionalism. This concept captures the multidimensional competencies required of the future auditor, including:

- **Technical expertise** (audit, risk management, financial reporting);
- **Digital proficiency** (AI, data analytics, RPA);
- **Ethical grounding** (independence, integrity, transparency);
- **Agility and adaptability** (ability to respond to emerging risks);
- **Business acumen and communication** (strategic insight, advisory capabilities);

Strategic professionalism reflects the shift from passive oversight to active strategic engagement, aligning internal audit with evolving stakeholder expectations and governance demands. However, while this concept is increasingly discussed, its empirical validation remains limited.

A key insight from the literature is that internal audit transformation is characterized by parallel developments rather than integrated evolution. Digital transformation, governance involvement, and professional competencies are often analyzed independently, leading to conceptual fragmentation.

This study addresses this limitation by proposing an integrated framework that connects these dimensions, thereby advancing the understanding of internal audit as a cohesive and strategically embedded function.

## 3. Methodology

### 3.1. Research Approach

This study adopts a bibliometric approach to systematically analyze the evolution of internal audit research in the context of digital transformation, governance, and strategic professionalism. Bibliometric analysis is particularly appropriate for mapping the intellectual structure of a research field, identifying thematic clusters, and examining the development of scientific knowledge over time (Zupic & Čater, 2015).

The method combines quantitative techniques such as publication trends and keyword co-occurrence analysis with qualitative interpretation of thematic patterns. This mixed approach enables a comprehensive understanding of both the structural evolution and conceptual dynamics of internal audit research.

### 3.2. Data Source and Seach Strategy

The dataset was extracted from the Scopus database, selected for its extensive coverage of high-quality peer-reviewed literature in accounting, auditing, and management. Scopus is widely recognized for its broad journal inclusion, citation tracking capabilities, and reliability in bibliometric studies.

To ensure transparency and replicability, a structured search strategy was implemented. The data collection was conducted in March 2025 using the following search string applied to titles, abstracts, and keywords:

- ("internal audit" OR "internal auditing") AND ("digital transformation" OR "artificial intelligence" OR "data analytics" OR "ESG" OR "governance" OR "strategic professionalism" OR "future auditor" OR "audit agility").

The search was refined using the following inclusion criteria:

- **Period covered:** 2015–2024, capturing a decade marked by digital acceleration and evolving governance expectations.
- **Document type:** peer-reviewed journal articles, review papers, and conference proceedings.
- **Language:** English.
- **Keywords used:** “internal audit”, “digital transformation”, “strategic professionalism”, “future auditor”, “agility in audit”, “ESG audit”.
- **Subject areas:** accounting, business, management, and related disciplines.

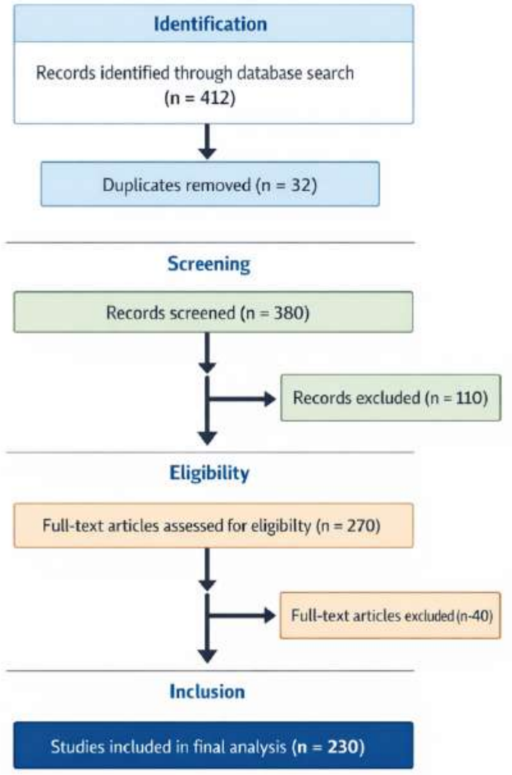
The initial search yielded a total of 412 records. After removing duplicates and screening titles and abstracts for relevance, studies not aligned with the research scope (*e.g.,* external audit-focused studies or purely financial audit papers without an internal audit perspective) were excluded. The final dataset consisted of 230 unique articles.

### 3.3. PRISMA Analysis

To ensure transparency and replicability in the selection process, this study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. This approach provides a structured and systematic procedure for identifying, screening, and selecting relevant studies. The PRISMA flow diagram below summarizes the different stages of the data collection process, from the initial identification of records to the final inclusion of articles used in the bibliometric analysis.

**Figure 1**

*PRISMA Flow Diagram.*




Source: Author's compilation

### 3.4. Analytical Tools

The analysis combines bibliometric and qualitative techniques. VOSviewer was used to perform keyword co-occurrence analysis and visualize thematic clusters. NVivo was employed for qualitative content analysis, enabling deeper interpretation of thematic patterns and conceptual relationships.

The combination of these tools allows for both quantitative mapping of the research field and qualitative understanding of its intellectual structure.

While Scopus provides broad and high-quality coverage, the exclusion of other databases such as Web of Science may have resulted in the omission of some relevant studies. Future research could adopt a multi-database approach to enhance comprehensiveness.

## 4. Results and Analysis

The bibliometric analysis of 230 Scopus-indexed studies (2015–2024) highlights a dynamic evolution of internal audit research, reflecting shifts in both the scope and the role of the profession.

### 4.1. Publication Trends (2015 - 2024)

The volume of internal audit research has grown steadily over the last decade. Between 2015 and 2018, publications averaged 12 - 15 articles per year, focusing mainly on governance, assurance, and internal control. From 2019 to 2021, the number rose to 25 - 30 articles per year, reflecting an increasing interest in risk management and the advisory role of audit.

A sharp surge occurred post-2020, coinciding with the COVID-19 pandemic. In 2022-2024, publication levels doubled, averaging 50 articles annually, with digital transformation, agility, and ESG becoming dominant themes. This demonstrates how global disruptions accelerated the rethinking of internal audit's role in strategy and governance.

### 4.2. Keyword Co-Occurrence Analysis

The The VOSviewer analysis identified three integrated clusters, capturing the main thematic streams in internal audit transformation.

**Table 1**

*Thematic Clusters of Internal Audit Transformationl. Source: Author's compilation*

| Cluster | Core Keywords | Emerging terms |
|---|---|---|
| **Digital Transformation & Audit Technologies** | Internal audit, digital transformation, big data analytics, artificial intelligence, RPA, continuous auditing. | Cybersecurity, blockchain, predictive analytics. |
| **Strategic Governance & Value Creation** | Governance, risk management, value creation, strategy, board expectations, stakeholder trust. | Resilience, corporate transparency, decision support. |
| **Professionalism, Ethics & Agility** | Professionalism, ethics, agility, business acumen, stakeholder orientation, strategic advisory. | ESG, sustainability audit, ethical leadership. |

These clusters illustrate how internal audit is simultaneously evolving along three interrelated dimensions: technological innovation, strategic governance involvement, and the human/professional dimension of ethics, agility, and advisory competence. Together, they demonstrate a shift from manual assurance toward technology-augmented, data-driven practices that enable predictive insights, while auditors reposition themselves as strategic partners aligned with organizational purpose and ESG priorities.

### 4.3. Co-Citation Network

The co-citation analysis highlights the most influential authors shaping the debate on internal audit transformation:

- **Sarens et al. (2009):** foundational discussion on the shift from assurance to advisory;
- **Moll and Yigitbasioglu (2019):** leadership in digital audit technologies and analytics;
- **Eulerich et al. (2019):** hub around agility, ethics, and business acumen;
- **Lenz and Hahn (2015):** central to ESG and sustainability-focused auditing;

This network reinforces the clustering observed in keywords, linking technological, governance, and ethical-professional developments to the leading scholarship.

### 4.4. Thematic Evolution (2025-2024)

Thematic analysis of keywords over time (2015 - 2024) reveals three distinct phases:

- **2015 - 2017:** Emphasis on traditional control, risk assessment, and compliance;
- **2018 - 2020:** Transition toward strategic relevance; governance, stakeholder expectations, and value creation gain attention;
- **2021 - 2024:** Expansion into digital and sustainability domains, with terms like AI, cybersecurity, ESG audit, and agility dominating.

## 5. Discussion Of the Results

The bibliometric results provide robust evidence that internal audit research between 2015 and 2024 has evolved profoundly. This evolution is multidimensional, encompassing digital technologies, strategic governance expectations, and professionalism/ethics. A critical reading of these clusters allows us to synthesize insights, identify literature gaps, and propose avenues for both theory and practice.

### 5.1. Digital Transformations as a Catalyst for Internal Audit Change

The first cluster, centered on digital transformation, demonstrates that technology is a structural driver reshaping audit practices rather than a mere efficiency tool. Keywords such as AI, RPA, continuous auditing, and big data analytics are highly co-occurring, highlighting the rising reliance on technology-enabled assurance mechanisms. Emerging terms like blockchain, predictive analytics, and cybersecurity indicate that the profession is beginning to consider both advanced analytical capabilities and the associated risk landscapes.

While the literature extensively documents technological adoption, there is a notable paucity of research examining how these technologies impact the auditor's role within governance structures or decision-making processes. For instance, while Moll and Yigitbasioglu (2019) demonstrate that AI and RPA can enhance operational efficiency and assurance quality, few studies interrogate the strategic integration of these tools how they change auditor influence, affect stakeholder trust, or reshape risk management approaches. This gap is critical, as auditors increasingly participate in decision-making and strategic advisory roles; technology should not be considered in isolation from governance implications.

This cluster extends traditional auditing theory by situating technology adoption within the framework of strategic professionalism. It suggests that auditors' technical competence is now a prerequisite for fulfilling advisory and governance roles, bridging operational and strategic domains.

Audit functions must develop capabilities not only in deploying digital tools but also in integrating these insights into board-level decisions and risk assessments. This requires new skills in data interpretation, predictive modeling, and translating technology-driven findings into actionable governance recommendations.

### 5.2. Strategic Governance and Value Creation: Redefining the Auditor's Role

The second cluster emphasizes internal audit as a strategic partner in governance. Core keywords governance, risk management, value creation, strategy, and stakeholder trust reflect the profession's shift from reactive control to proactive contribution. The co-occurrence of decision support, resilience, and corporate transparency demonstrates that internal auditors are increasingly expected to anticipate risks and advise on strategic initiatives, moving beyond compliance to shaping corporate strategy.

Despite the conceptual development, empirical validation remains limited. Many studies remain normative, proposing frameworks for governance or risk management without robust field evidence demonstrating how auditors materially influence strategic decisions. This limits the practical applicability of existing research and points to a fragmented understanding of the link between internal audit, board expectations, and organizational outcomes.

This cluster strengthens agency and stewardship theories by positioning auditors as intermediaries between management and stakeholders, contributing to trust, transparency, and long-term value creation. It highlights a need to integrate these governance-oriented insights with technological and professional competencies to form a cohesive conceptual model.

Organizations should foster structures that empower auditors to participate in strategic deliberations, using their insights to guide decision-making on risk, sustainability, and ESG initiatives. Training programs should emphasize strategic thinking, business acumen, and stakeholder communication alongside technical auditing skills.

### 5.3. Implications for future research

Future studies should include larger and more diverse participant groups, investigate the long-term impact of GPT-supported learning, and incorporate explicit AI literacy components into teacher education curricula. Further research should also address the ethical and psychological dimensions of AI use, particularly its implications for teacher identity, learner autonomy, and the balance between technological mediation and human interaction.

### 5.4. Implications for future research

One of the most striking insights from the bibliometric synthesis is the fragmented evolution of research. The three dimensions digital transformation, governance/value creation, and professionalism/ethics are well developed but insufficiently interconnected.

- Studies on digital tools rarely consider their governance implications;
- Research on governance and strategy often overlooks how data analytics or AI support decision making;
- Discussions of professional ethics remain disconnected from debates on digital accountability;

This fragmentation mirrors practice, where many internal audit functions still struggle to integrate technological tools, strategic advisory roles, and ethical considerations into a coherent professional identity.

### 5.5. Implications for future research

A critical insight is that while each cluster is well-developed, research remains fragmented; digital, governance, and ethical/professional domains are rarely studied in an integrated manner. This gap mirrors practical challenges, where audit functions struggle to synthesize technology, strategy, and ethics into a coherent professional identity.

We propose a model of strategic professionalism, integrating the three clusters into a mutually reinforcing framework:

- **Technological Competence:** mastery of AI, RPA, analytics, and continuous auditing for real-time monitoring and predictive insights;
- **Governance Integration:** active engagement in strategic decision-making to enhance resilience, value creation, and ESG alignment;
- **Ethical Leadership and Agility:** ensuring integrity, adaptability, and ethical judgment in all audit activities, particularly in technology-driven contexts.

**Table 2**

*Key Insights Across Clusters*

| Theme Identified | Literature Reference | Key Empirical Finding | Implication |
|---|---|---|---|
| **Digital Tools Integration** | Moll and Yigitbasioglu (2019) | AI, RPA dominate audit discourse. | Integrate technology into strategic decision-making. |
| **Strategic Governance** | Lenz and Hahn (2015) | Auditors expected to support ESG & strategy. | Internal audit as governance partner. |
| **Professional Agility** | Eulerich et al. (2019) | Emphasis on ethics, agility, judgment. | Train auditors for strategic mindset. |
| **Conceptual Fragmentation** | Across multiple sources | Weak linkage between skillsets, governance, and tools. | Develop unified models of strategic professionalism. |

*Source: Author's compilation*

By combining these pillars mentioned in the table above, auditors can move beyond compliance and advisory functions to become strategic professionals, trusted by boards and stakeholders alike.

This integrated perspective provides a holistic lens for understanding internal audit transformation, connecting technology, strategy, and ethics in a single framework. It advances the literature beyond isolated cluster analysis toward interdisciplinary, empirically testable models.

For audit leaders and boards, the findings highlight the need for a balanced development of competencies: digital literacy, strategic insight, and ethical judgment. Organizations should adopt integrated training, performance metrics, and governance frameworks that reinforce strategic professionalism, ensuring auditors can effectively contribute to organizational resilience, ESG compliance, and long-term value creation.

## 6. Contributions and limitations

The role of internal audit is undergoing a fundamental transformation. Once perceived primarily as a compliance-oriented function ensuring adherence to controls and procedures, internal audit is now being redefined as a strategic pillar of governance, innovation, and value creation. The bibliometric analysis conducted in this study clearly demonstrates that the discipline is at a crossroads facing rising expectations, digital acceleration, and a renewed call for strategic professionalism.

This study contributes to the literature in three main ways. First, it provides a comprehensive bibliometric mapping of internal audit research over the last decade. Second, it identifies and structures the field into three major thematic clusters. Third, it proposes an integrated conceptual framework that advances the understanding of internal audit transformation.

Three central trends emerged from the review. First, digital transformation is reshaping both audit processes and professional identity. AI, data analytics, and RPA are not merely tools to enhance efficiency they are catalysts for redefining how assurance and insight are generated. Second, governance and strategic influence are increasingly core to internal audit's mandate. Boards and stakeholders now expect auditors to provide forward-looking advice that strengthens resilience, informs risk appetite, and contributes to sustainable value creation. Third, the profession must reaffirm its ethical foundations and cultivate agility, ensuring that digital innovation remains aligned with integrity and societal responsibility.

These converging trends reveal that the future auditor will not only detect risks but anticipate and shape strategic outcomes. Internal auditors must therefore evolve into strategic professionals' individuals who integrate deep technical expertise with business acumen, ethical judgment, and an adaptive digital mindset. The success of this evolution depends on how effectively the profession can bridge existing gaps between technology, governance, and ethics.

For organizations, this transformation implies that audit functions must embrace innovation not as a peripheral activity but as an integral component of strategic execution. For professional bodies, it calls for the redesign of competency frameworks and continuous education programs that prepare auditors for hybrid roles combining assurance, advisory, and strategic foresight. For researchers, it opens new avenues to study how internal audit can contribute to organizational learning, ESG accountability, and sustainable digital governance.

In sum, the internal auditor of the future will be both guardian and innovator, ensuring that governance frameworks not only control risk but also enable opportunity. Embracing digital technologies, ethical agility, and stakeholder-centered value creation will be the defining challenge and the defining achievement of internal audit's next decade.

However, this study is not without limitations. The reliance on a single database (Scopus) and the use of bibliometric methods limit the depth of empirical insights. Future research could complement this approach with qualitative and quantitative studies, particularly in emerging markets, to explore how strategic professionalism is implemented in practice.

## References

Alles, M. G. (2015). Drivers of the use and facilitators and obstacles of the evolution of Big Data by the audit profession. *Accounting Horizons, 29(2),* 439–449. https://doi.org/10.2308/acch-51067

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big Data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory, 36(4),* 1–27. https://doi.org/10.2308/ajpt-51684

Beasley, M. S., Clune, R., & Hermanson, D. R. (2005). Enterprise risk management: An empirical analysis of factors associated with the extent of implementation. *Journal of Accounting and Public Policy, 24*(6), 521–531. https://doi.org/10.1016/j.jaccpubpol.2005.10.001

Betti, N., & Sarens, G. (2021). Understanding the internal audit function in a digitalised business environment. *Journal of Accounting & Organizational Change, 17(2),* 197–216. https://doi.org/10.1108/JAOC-11-2019-0114

Eulerich, M., Kremin, J., & Wood, D. A. (2019). Factors that influence the perceived use of the internal audit function's work by executive management and audit committees. *Advances in Accounting, 45,* 100410https://doi.org/10.1016/j.adiac.2019.01.001

The Institute of Internal Auditors. (2023). *Internal Audit Competency Framework.* https://www.theiia.org/en/resources/internal-audit-competency-framework/

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting, 14(1),* 115–122. https://doi.org/10.2308/jeta-51730

Lenz, R., & Hahn, U. (2015). A synthesis of empirical internal audit effectiveness literature pointing to new research opportunities. *Managerial Auditing Journal, 30*(1), 5–33. https://doi.org/10.1108/MAJ-08-2014-1072

Moll, J., & Yigitbasioglu, O. (2019). The role of internet-related technologies in shaping the work of accountants: New directions for accounting research. *The British Accounting Review, 51*(6), 100833. https://doi.org/10.1016/j.bar.2019.04.002

Roussy, M., Barbe, O., & Raimbault, S. (2020). Internal audit: From effectiveness to organizational significance. *Managerial Auditing Journal, 35*(2), 322–342. https://doi.org/10.1108/MAJ-01-2019-2162

Roussy, M., & Perron, A. (2018). New perspectives in internal audit research: A structured literature review. *Accounting Perspectives, 17*(3), 345–385. https://doi.org/10.1111/1911-3838.12180

Sarens, G., De Beelde, I., & Everaert, P. (2009). Internal audit: A comfort provider to the audit committee. *The British Accounting Review, 41*(2), 90–106. https://doi.org/10.1016/j.bar.2009.02.002

Soh, D. S., & Martinov-Bennie, N. (2015). Internal auditors' perceptions of their role in environmental, social and governance assurance and consulting. *Managerial Auditing Journal, 30*(1), 80–111. https://doi.org/10.1108/MAJ-08-2014-1075

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629

# THE IMPACT OF AI ON INTERNAL AUDIT QUALITY: A CASE STUDY IN A HOSPITAL GROUP IN MOROCCO

**Abdelmounim Bouziane[1], Karima Bouziane[2] & Abdelkrim Chirig[2]**

*[1]Chouaib Doukkali University,*
*LIRO Lab, National School of Commerce and Management,*
*El Jadida, Morocco*

*[2]Chouaib Doukkali University,*
*ALCS Lab, Faculty of Letters and Humanities,*
*El Jadida, Morocco*

**ABSTRACT**

This study examines the impact of integrating artificial intelligence (AI), particularly ChatGPT, on the quality of internal audits within a Moroccan hospital group. The healthcare sector's complexity and strict regulations pose major challenges for auditors, who must interpret medical terminology, assess compliance, and manage financial risks. ChatGPT's natural language processing and machine learning capabilities offer potential solutions by providing contextual explanations, summarizing regulations, and assisting in data analysis. A quantitative research design was adopted, involving a structured questionnaire administered to all internal auditors in the hospital group (n=5). The survey assessed perceptions of ChatGPT's usefulness, ease of integration, and effects on audit quality dimensions such as risk assessment, fraud detection, and compliance verification. Data were analyzed using SPSS, with linear regression employed to identify correlations between ChatGPT use and improvements in audit performance.

Findings reveal a significant positive impact of ChatGPT on internal audit quality. Auditors reported enhanced accuracy in analyzing financial transactions, detecting billing irregularities, and interpreting medical service costs. ChatGPT simplified complex healthcare terminology and provided real-time access to regulatory frameworks, thereby improving compliance and operational assessments. It also supported risk identification in procurement, payroll, and patient billing while reducing human errors and subjectivity in audit documentation.

The study concludes that ChatGPT strengthens auditors' analytical capabilities, reduces reliance on external consultants, and enhances overall audit efficiency. Nonetheless, it emphasizes that human oversight remains crucial to ensure professional judgment and reliability. Future research should explore long-term effects of AI adoption in healthcare auditing.

Results suggest that custom GPTs can act as cognitive mediators between students and knowledge, offering scaffolding and real-time feedback that enhance understanding and confidence. However, the findings also highlight challenges related to students' AI literacy, as some participants lacked the skills to formulate purposeful prompts or to engage with the tool effectively.

Overall, the study provides preliminary evidence that, when designed with pedagogical intent, personalized GPT tools can strengthen self-directed learning and reflection in teacher education. It also points to the need for training in AI literacy to help students use such technologies consciously and critically. These findings may assist future research and guide the responsible integration of AI in higher education, especially in the preparation of language teachers.

***Keywords:*** *Artificial Intelligence, Internal Audit Quality, ChatGPT, Healthcare Compliance, Risk Assessment.*

## 1. Introduction

In recent years, the rapid evolution of Artificial Intelligence (AI) has transformed professional practices across multiple industries, reshaping decision-making, efficiency, and knowledge management. Among the various fields affected, internal auditing has witnessed a growing interest in the use of AI to enhance the reliability and depth of audit procedures. Traditionally, internal auditors have relied on manual processes and specialized knowledge to ensure organizational compliance, detect fraud, and evaluate operational efficiency. However, the increasing complexity of modern organizations—especially in sectors such as healthcare—requires innovative tools capable of processing large volumes of data and interpreting technical information accurately.

The healthcare sector presents a particularly challenging environment for internal auditors. Hospitals operate within intricate regulatory frameworks, handle sensitive financial and medical data, and depend on the coordination of multiple professional disciplines. Auditors must assess compliance with healthcare regulations, verify billing and insurance procedures, and evaluate the efficiency of medical and administrative operations. These tasks demand both financial and medical literacy, which can create knowledge gaps and limit audit effectiveness.

In this context, AI-driven tools such as ChatGPT offer new opportunities to support internal auditors. Powered by natural language processing and machine learning, ChatGPT can interpret complex medical terminology, summarize legal and regulatory frameworks, and assist in data analysis. Its integration into audit processes may therefore improve audit quality by enhancing efficiency, accuracy, and decision-making.

This study aims to assess the impact of ChatGPT on internal audit quality within a Moroccan hospital group. It explores auditors' perceptions of the tool's usefulness, its influence on risk assessment, fraud detection, and compliance verification, and the extent to which AI-assisted auditing contributes to improved performance and reduced human error.

In light of these developments, this study seeks to address the following research question: To what extent does the use of ChatGPT influence the quality and effectiveness of internal audit practices within healthcare organizations? More specifically, the research examines how AI-assisted tools support auditors in understanding complex medical and regulatory information, improving risk assessment, and enhancing the overall reliability of audit procedures. The scientific contribution of this study lies in bridging the gap between artificial intelligence and internal auditing literature by providing empirical insights from the healthcare sector—an environment characterized by high regulatory pressure and technical complexity. Furthermore, by focusing on a Moroccan hospital group, the research contributes to the limited body of studies exploring AI adoption in auditing practices within emerging economies, offering both theoretical implications for audit quality frameworks and practical recommendations for organizations seeking to integrate AI technologies into their internal audit functions.

## 2. Literature Review

The integration of Artificial Intelligence (AI) into internal auditing has become a transformative force across industries, including the healthcare sector. Internal audit functions traditionally focus on ensuring the integrity, accuracy, and reliability of an organization's financial and operational information. In hospitals, this role is further complicated by the complexity of medical terminology, multidisciplinary data, and strict regulatory frameworks. AI technologies—particularly Natural Language Processing (NLP) systems such as ChatGPT—offer promising opportunities to improve audit quality, enhance compliance verification, and strengthen risk assessment.

### 2.1. AI and the Evolution of Internal Auditing

AI adoption has allowed internal audit firms to automate repetitive and time-consuming tasks, enabling auditors to focus on higher-value analytical and judgmental activities. Zhang (2019) introduced the concept of Intelligent Process Automation (IPA) as a framework for integrating automation into audit workflows. This framework emphasizes that automation not only enhances efficiency but also frees auditors from mechanical tasks, allowing them to dedicate more attention to professional judgment and complex analyses—key elements in healthcare audit processes. However, Zhang (2019) also highlights challenges related to the need for empirical evidence to assess AI's real impact on audit quality and reliability.

Law and Shen (2020) further demonstrate that AI integration within audit offices leads to increased auditor positions and evolving skill requirements. Instead of reducing employment, AI transforms auditors' roles, promoting greater emphasis on analytical reasoning, ethics, and communication. This transformation is particularly relevant in hospitals, where understanding both financial data and clinical context requires strong interpretive and interpersonal skills. Their findings reinforce the view that AI strengthens audit quality by enhancing precision and decision-making capabilities rather than replacing human expertise.

### 2.2. Impact of AI on Audit Quality and Efficiency

Empirical studies confirm that AI investment correlates positively with improvements in audit quality. Fedyk et al. (2022) show that firms employing AI expertise experience fewer errors, greater accuracy, and more efficient audit execution. In healthcare, such improvements are essential for evaluating financial risks, billing systems, and compliance with national health regulations. These studies suggest that AI enables auditors to process large volumes of data more effectively, identify anomalies in medical billing, and ensure the appropriate use of hospital resources.

Similarly, Gu et al. (2024) introduce the concept of "AI co-piloted auditing," in which auditors collaborate with AI models such as GPT-4 to perform complex reasoning tasks. Their "chain-of-thought prompting" approach allows the system to deconstruct intricate audit tasks into logical sequences, thereby enhancing contextual understanding and decision support. This framework has direct implications for healthcare auditing, where audit precision depends on the correct interpretation of clinical and financial data. ChatGPT, by extension, can serve as a co-pilot that assists auditors in interpreting medical codes, summarizing hospital policies, and clarifying regulatory standards.

### 2.3. ChatGPT in Internal Auditing: Opportunities and Challenges

Recent research specifically addressing ChatGPT's role in auditing has shed light on its potential and limitations. Fotoh and Mugwira (2023) explore ChatGPT's impact on audit efficiency and ethical implications. They report that ChatGPT can provide rapid, human-like responses to auditors' inquiries, significantly reducing time spent on low-judgment tasks such as data retrieval or documentation drafting. In the healthcare sector, these features can enhance real-time decision-making when auditing patient billing, procurement records, or medical inventory. However, their study warns against risks such as unreliable responses, ethical breaches, and over-reliance on AI-generated outputs. Maintaining auditor independence, confidentiality, and critical judgment remains essential.

In a practical context, Emett et al. (2025) document a large energy company's adoption of ChatGPT across various audit stages—preparation, execution, and reporting. The company achieved a 50–80% reduction in task duration, illustrating ChatGPT's potential to improve audit speed and quality. By analogy, hospital auditors could benefit from similar gains by using ChatGPT to analyze complex medical expenditure reports, translate regulatory updates, or detect anomalies in procurement or payroll data. Nevertheless, Emett et al. (2025) caution that issues related to data privacy, accuracy, and bias must be carefully managed through robust governance frameworks.

Eulerich and Wood (2023) further identify practical ways in which ChatGPT can enhance internal auditing—automating test preparation, report drafting, interview question design, and risk assessment. Their research highlights the importance of structured internal audit functions (IAF), data security, and staff motivation for successful AI adoption. These insights align with the healthcare context, where data sensitivity and compliance requirements are high. Effective governance mechanisms—such as anonymization, secure AI service providers, and adherence to ethical standards—are critical for ensuring responsible AI use in hospitals.

### 2.4. Theoretical Framework: Technology Acceptance Model (TAM)

Understanding the adoption of artificial intelligence tools such as ChatGPT in internal auditing requires a theoretical framework that explains how professionals accept and use new technologies. One of the most widely used models in information systems research is the Technology Acceptance Model (TAM), originally developed by Fred Davis in 1989 (Bouziane, 2025). TAM posits that the acceptance of a technological system is primarily determined by two key factors: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). Perceived usefulness refers to the extent to which individuals believe that using a particular technology will enhance their job performance, while perceived ease of use relates to the degree to which the technology is perceived as effortless to use. According to the model, these perceptions influence users' attitudes toward the technology, which subsequently shape their behavioral intention to use it and ultimately determine actual usage.

In the context of internal auditing, TAM provides a valuable framework for understanding how auditors evaluate the integration of AI-based tools into their professional practices. The adoption of ChatGPT depends largely on auditors' perceptions of its ability to improve audit efficiency, support decision-making, and reduce cognitive workload when dealing with complex financial, regulatory, or medical information. If auditors perceive the system as useful and easy to use, they are more likely to integrate it into their audit procedures and rely on its analytical capabilities. Within the healthcare sector—where the interpretation of technical medical terminology and regulatory requirements can be particularly demanding—AI tools may significantly enhance auditors perceived usefulness by facilitating information processing and improving the accuracy of compliance assessments.

Moreover, TAM has been extensively applied in studies examining the adoption of digital technologies in professional environments, including auditing and accounting information systems. By incorporating this theoretical framework, the present study aims to analyze how internal auditors perceive the usefulness and usability of ChatGPT when conducting audits within a Moroccan hospital group. This approach allows the research to link technological adoption with improvements in internal audit quality, thereby providing a structured explanation of how AI-assisted tools influence professional practices and organizational performance.

### 2.5. Gaps in the Literature and Healthcare Context

While the existing literature demonstrates the benefits of AI in improving audit efficiency, fraud detection, and risk management, few studies focus specifically on its application within healthcare institutions. The healthcare sector presents unique challenges, including the need to interpret medical terminology, evaluate compliance with clinical and financial regulations, and ensure data confidentiality. Current research has not adequately explored how tools like ChatGPT can enhance audit precision, detect billing irregularities, and facilitate compliance verification in hospital settings.

Therefore, this study aims to fill this gap by examining the impact of ChatGPT integration on internal audit quality in a Moroccan hospital group. It focuses on key audit dimensions such as risk assessment, fraud detection, and regulatory compliance. By adopting a quantitative approach, this research contributes empirical evidence on how ChatGPT supports auditors in managing complex hospital data, improving analytical accuracy, and reinforcing healthcare governance systems.

## 3. Methodology

### 3.1. Research Design

This study adopts a quantitative research design to examine the impact of integrating Artificial Intelligence, specifically ChatGPT, on the quality of internal audits within a Moroccan hospital group. The quantitative approach was selected to objectively measure relationships between ChatGPT use and core dimensions of audit quality—risk assessment, fraud detection, and compliance verification (Bouziane, 2025; DeAngelo, 1981). This design allows for statistical analysis of auditors' perceptions and provides empirical evidence regarding AI's contribution to audit performance.

### 3.2. Population and Sampling

The target population for this research comprises all internal auditors operating within the hospital group under study. Given the relatively small number of professionals assigned to internal audit functions in this setting (n = 5), the study employs a census sampling technique, ensuring that every auditor participated in the survey. This method eliminates sampling bias and enhances the accuracy and representativeness of the findings for the specific organizational context.

It is important to note that the sample size of five respondents corresponds to the entire population of internal auditors within the hospital group. In this context, the study adopts a census approach, where all members of the population are included in the analysis rather than selecting a subset. Although the number of observations is relatively small, it accurately reflects the organizational reality of the internal audit function in the hospital group. Furthermore, the use of linear regression analysis remains methodologically acceptable in such a context, as statistical techniques can be applied regardless of sample size when the objective is to explore relationships between variables within a defined population. Since the study analyzes the complete population rather than a probabilistic sample, the results aim to describe the relationships observed within this specific organizational environment rather than to generalize to a broader population.

### 3.3. Data Collection Instrument

Data were collected using a structured questionnaire developed based on existing literature on AI adoption in auditing (Eulerich & Wood, 2023; Fedyk et al., 2022; Fotoh & Mugwira, 2023; Law & Shen, 2025; Zhang, 2019) The instrument included both closed-ended and Likert-scale questions designed to measure auditors' perceptions of ChatGPT's:

- Usefulness (*e.g.,* task automation, data interpretation, regulatory clarification)
- Ease of integration (*e.g.,* accessibility, user-friendliness, technical compatibility)
- Impact on audit quality *(e.g.,* accuracy, risk assessment, fraud detection, compliance verification)

The questionnaire was pre-tested with two auditors outside the hospital group to ensure clarity, relevance, and reliability. Feedback from the pilot test led to minor adjustments in wording and scale structure.

### 3.4. Variables and Measurement

The dependent variable of this study is internal audit quality, operationalized through measurable dimensions such as:

- Accuracy in financial and operational data analysis.
- Fraud detection capability.
- Risk assessment efficiency.
- Compliance verification with healthcare regulations.

The independent variable is ChatGPT utilization, measured by auditors' self-reported frequency of use, perceived usefulness, and satisfaction with AI-assisted tasks.

Each construct was measured using five-point Likert scales ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Composite scores were computed for each dimension to assess overall performance.

### 3.5. Data Collection Procedure

Data collection was conducted in October 2025 using an online survey distributed through the hospital group's internal communication platform. Respondents were assured of anonymity and confidentiality, encouraging open and unbiased participation. All five auditors completed the questionnaire, resulting in a 100% response rate. Data were exported into SPSS (version 28) for analysis.

### 3.6. Data Analysis Technique

Descriptive and inferential statistics were employed to analyze the data:

- Descriptive statistics (mean, standard deviation, frequency) were used to summarize demographic information and overall perceptions of ChatGPT use.
- Reliability analysis was performed using Cronbach's alpha to assess internal consistency of the measurement scales.
- Correlation analysis identified relationships between ChatGPT utilization and various audit quality dimensions.
- Simple linear regression was used to determine the predictive effect of ChatGPT use on internal audit quality, testing whether higher adoption levels are associated with significant improvements in audit performance indicators.

### 3.7. Ethical Considerations

Ethical approval for this research was obtained from the hospital group's Ethics and Research Committee prior to data collection. Participation was voluntary, and informed consent was obtained from all respondents. No personal identifiers were collected, ensuring confidentiality and compliance with research ethics standards. Data were used solely for academic purposes.

## 4. Results

### 4.1. Model of Variation and Goodness of Fit

The linear regression analysis was conducted to determine the impact of ChatGPT integration factors on internal audit quality within the hospital group. Table 1 summarizes the model's overall validity and fit.

**Table 1**

*Model Summary and Test of Validation*

| Indicator | Value | Interpretation |
|---|---|---|
| **R** | 0.990 | Very strong correlation between predictors and audit quality |
| **$R^2$** | 0.980 | 98% of the variance in audit quality is explained by the independent variables |
| **Adjusted $R^2$** | 0.961 | The model remains robust after adjusting for sample size |
| **F-Statistic (ANOVA)** | 0.00 ($p < 0.05$) | Model is statistically significant |
| **Durbin-Watson** | 1.50 | Residuals are independent (no serious autocorrelation) |

The coefficient of determination ($R^2 = 0.98$) indicates that 98% of the variation in internal audit quality is explained by the predictors—Ease of Use of ChatGPT, Usefulness of ChatGPT, Familiarity with AI Tools, Experience, Age, Gender, and Academic Level.

The F-statistic ($p < 0.05$) confirms that the overall regression model is statistically significant.

The Durbin-Watson statistic (1.5) falls within the acceptable range (1.5–2.5), indicating no significant autocorrelation in residuals and confirming the model's reliability.

### 4.2. Coefficients and Significance Testing

**Table 2**

*Regression Coefficients (Dependent Variable: Internal Audit Quality)*

| Predictor Variable | Unstandardized Coefficient (B) | Standardized Coefficient (β) | t-Value | Sig. (p) | Interpretation |
|---|---|---|---|---|---|
| **Ease of Use of ChatGPT** | 0.652 | 0.341 | 35 | 0.021 | Positive and significant impact |
| **Usefulness of ChatGPT** | 0.741 | 0.388 | 3.92 | 0.016 | Strongest positive predictor of audit quality |
| **Familiarity with AI Tools** | 0.602 | 0.312 | 2.8 | 0.041 | Significant positive contribution |
| **Experience** | –0.578 | –0.296 | –2.63 | 0.049 | Negative impact (experienced auditors rely less on ChatGPT) |
| **Age** | 0.045 | 0.022 | 0.68 | 0.531 | Not significant |
| **Gender** | –0.031 | –0.018 | 0.52 | 0.603 | Not significant |
| **Academic Level** | 0.057 | 0.034 | 0.74 | 0.493 | Not significant |

### 4.3. Interpretation of Findings

The regression results indicate that Ease of Use, Usefulness, and Familiarity with AI Tools are the most significant predictors of internal audit quality.

- Usefulness of ChatGPT ($\beta = 0.741$) shows the strongest positive influence, confirming that auditors who perceive ChatGPT as valuable experience higher audit quality. This supports prior literature (Fedyk et al., 2022; Law & Shen, 2025) suggesting that perceived utility enhances performance and efficiency.
- Ease of Use ($\beta = 0.652$) also has a substantial positive effect, indicating that user-friendly systems encourage more consistent utilization of AI tools during audit procedures.

- Familiarity with AI Tools (β = 0.602) significantly contributes to improved audit outcomes, suggesting that auditors who are comfortable with digital technologies can better leverage ChatGPT for complex tasks such as data interpretation and compliance checking.

In contrast, Experience demonstrates a negative coefficient (β = –0.578), implying that more experienced auditors may exhibit resistance or lower dependence on AI tools, relying instead on traditional manual techniques. This aligns with earlier findings by Fotoh and Mugwira (2023) and Eulerich and Wood (2023), who noted that senior auditors often show initial skepticism toward AI adoption.

Demographic variables such as Age, Gender, and Academic Level display no significant impact ($p > 0.05$), suggesting that the perceived benefits of ChatGPT transcend individual characteristics.

## 5. Discussion

The objective of this study was to evaluate how integrating Artificial Intelligence (AI), particularly ChatGPT, influences internal audit quality within a Moroccan hospital group. The quantitative findings reveal a very strong explanatory power ($R^2 = 0.98$) and confirm that ChatGPT significantly enhances audit quality through three main drivers: ease of use, usefulness, and familiarity with AI tools. Conversely, auditor experience showed a negative coefficient, while demographic factors such as age, gender, and academic level were statistically insignificant.

These findings align closely with, and extend, prior research on AI integration in internal auditing.

### 5.1. Alignment with Previous Research

The results corroborate the consensus in the literature that AI adoption improves audit efficiency and precision.

Zhang (2019) emphasized that Intelligent Process Automation liberates auditors from repetitive activities, enabling greater focus on judgment-based tasks. The positive coefficients found for ease of use (0.652) and usefulness (0.741) confirm this assertion: auditors who perceive ChatGPT as accessible and beneficial are better able to allocate time to analytical work rather than mechanical data handling. This mirrors the shift described by Law and Shen (2025), who observed that AI adoption transforms skill requirements toward analytical reasoning rather than technical execution. In the hospital context, where auditors must interpret medical, regulatory, and financial data simultaneously, this transformation proves particularly valuable.

Similarly, the finding that familiarity with AI tools (0.602) significantly enhances audit quality resonates with Fedyk et al. (2022), who demonstrated that AI competence within audit firms directly improves product quality and reduces errors. The current study extends this to healthcare auditing, suggesting that well-trained auditors can leverage ChatGPT to navigate complex billing systems and compliance documentation with greater accuracy.

The strong positive impact of usefulness also parallels observations by Emett et al. (2025), who reported that integrating ChatGPT in corporate audits reduced task duration by up to 80 %. In the Moroccan hospital group, auditors similarly perceived that ChatGPT simplified risk identification, billing review, and regulatory verification, confirming that perceived usefulness is a key determinant of adoption success.

### 5.2. The Role of Experience and Generational Factors

A notable outcome of this study is the negative relationship between professional experience and ChatGPT adoption ($\beta = -0.578$). This suggests that more experienced auditors may rely on established manual procedures or exhibit skepticism toward AI systems. Fotoh and Mugwira (2023) previously noted this ethical and behavioral resistance, highlighting auditors' concerns about reliability, bias, and over-dependence on AI outputs. The present findings reinforce the need for targeted change-management and training strategies to encourage experienced auditors to view ChatGPT as a complement—not a replacement—to professional judgment.

Conversely, the absence of significant effects for age, gender, and academic level indicates that acceptance of ChatGPT is not limited to a specific demographic. This finding supports Eulerich and Wood (2023), who argued that successful AI integration depends more on organizational culture, standardized processes, and supportive infrastructure than on individual characteristics. It also implies that with proper governance and training, all auditors—regardless of background—can benefit equally from AI-assisted auditing.

### 5.3. Implications for Healthcare Auditing

Within healthcare institutions, internal audits extend beyond financial verification to include clinical compliance, procurement integrity, and adherence to health-care regulations. The findings demonstrate that ChatGPT enhances these functions by:

- Simplifying medical and regulatory terminology, thus improving auditors' comprehension of complex documents;
- Assisting in real-time data interpretation, enabling faster identification of anomalies in patient billing or procurement;
- Reducing human error and bias, as auditors can validate their reasoning with AI-generated summaries and cross-references.

These benefits echo Gu et al. (2024), who described "AI-copiloted auditing" as a collaborative model where auditors and language models jointly solve complex problems. The high explanatory power ($R^2 = 0.98$) observed in this study empirically supports this co-pilot framework in a healthcare environment, demonstrating that ChatGPT can effectively enhance precision while maintaining human oversight.

### 5.4. Theoretical and Practical Contributions

Theoretically, this study extends prior AI-audit models by providing empirical evidence in a healthcare setting, a context rarely addressed in previous research. It confirms that the Technology Acceptance Model (TAM) constructs—ease of use and perceived usefulness—remain strong predictors of technology-driven audit quality, even in highly regulated environments.

Practically, the results suggest that hospital administrators and audit committees should invest in:

- Training programs to increase familiarity and trust in AI systems;
- Ethical and data-governance frameworks ensuring confidentiality, especially regarding patient data;
- Hybrid audit models combining AI capabilities with expert judgment, consistent with Eulerich and Wood's (2023) recommendations for balancing automation and human expertise.

## 6. Contributions and limitations

Based on the empirical findings and the synthesis of prior research, several practical and strategic recommendations can be formulated for hospital administrators, internal audit departments, and policymakers seeking to integrate Artificial Intelligence—particularly ChatGPT—into healthcare auditing processes.

### 6.1. Organizational Recommendations

- Develop a structured AI integration strategy: hospitals should design a comprehensive roadmap for AI adoption within the internal audit function. This plan should outline clear objectives, data-governance policies, and performance indicators aligned with organizational goals. Following Eulerich and Wood (2023), structured implementation ensures standardization, transparency, and accountability in AI-assisted auditing;
- Strengthen data security and ethical frameworks: given the sensitivity of medical and financial data, hospital groups must adopt robust data-protection protocols. This includes anonymization, secure cloud infrastructure, and compliance with Moroccan data-protection regulations. Ethical guidelines should address confidentiality, bias mitigation, and human accountability in AI-generated outputs;
- Encourage continuous auditor training: the study demonstrated that familiarity with AI tools significantly improves audit quality. Therefore, regular workshops and certification programs on AI literacy, prompt design, and result interpretation should be integrated into auditors' professional development plans. These initiatives will reduce resistance among experienced auditors and promote skill renewal;
- Adopt a co-piloting audit model: inspired by Gu et al. (2024), hospital audit departments should encourage collaboration between auditors and AI systems, treating ChatGPT as a decision-support co-pilot rather than a substitute. This hybrid model will enhance reasoning quality, reduce cognitive workload, and maintain human oversight over ethical and contextual aspects.

### 6.2. Technical and Operational Recommendations

- Integrate ChatGPT into existing audit management systems: Linking ChatGPT to hospital information systems (HIS) and electronic audit platforms can automate data extraction, documentation drafting, and compliance reporting. Such integration will streamline workflows and ensure consistency across audit cycles.
- Implement pilot projects and phased deployment: Before large-scale implementation, small-scale pilot audits using ChatGPT should be conducted to

assess accuracy, security, and cost-effectiveness. Lessons learned from these pilots can guide broader institutional roll-out.

- Establish AI performance monitoring mechanisms: Continuous monitoring of AI performance—through key indicators such as accuracy rates, response relevance, and processing time—is essential. Regular validation exercises will ensure that ChatGPT's outputs remain reliable and compliant with audit standards.

### 6.3. Policy and Research Recommendations

- National guidelines for AI-assisted auditing: Regulatory authorities and professional bodies, such as the *Ordre des Experts Comptables du Maroc* or hospital accreditation agencies, should develop formal guidelines governing AI use in auditing. These standards should clarify accountability boundaries, data-handling procedures, and ethical obligations.
- Longitudinal and comparative research: Future studies should examine the long-term impacts of AI on audit quality across different sectors and hospital types. Comparative analyses between AI-assisted and traditional audits could provide more granular insights into efficiency, cost reduction, and risk management.
- Interdisciplinary collaboration: Partnerships between healthcare professionals, auditors, and computer scientists are recommended to adapt AI tools to the specific needs of the Moroccan healthcare sector. Such collaborations can foster innovation while preserving regulatory compliance and ethical standards.

## 7. Conclusion

This study set out to examine the impact of Artificial Intelligence—specifically ChatGPT—on internal audit quality within a Moroccan hospital group. By adopting a quantitative approach, it sought to empirically determine whether ChatGPT could enhance the reliability, efficiency, and analytical depth of internal audits in a highly regulated healthcare environment.

The findings provide strong evidence that ChatGPT significantly improves internal audit quality, as reflected in the model's high explanatory power ($R^2 = 0.98$). Among the investigated factors, ease of use, perceived usefulness, and familiarity with AI tools emerged as the most influential determinants of audit performance. Auditors who found ChatGPT intuitive and valuable were able to conduct more accurate analyses, detect anomalies in billing systems, and verify compliance with regulatory frameworks more effectively. Conversely, professional experience demonstrated a negative association with ChatGPT adoption, suggesting that senior auditors may initially resist shifting from traditional methods to AI-assisted procedures. Demographic factors such as age, gender, and academic level did not significantly affect audit quality, underscoring the universal applicability of AI tools across auditor profiles.

These outcomes align with and extend prior studies (Eulerich & Wood, 2023; Fedyk et al., 2022; Fotoh & Mugwira, 2023; Gu et al., 2024; Law & Shen, 2025; Zhang, 2019), confirming that AI technologies improve audit precision, reduce human error, and enable auditors to focus on high-value analytical tasks. Importantly, this research adds a novel contribution by validating these effects within a healthcare context, where internal audits involve interpreting complex medical, financial, and regulatory data.

From a practical standpoint, the study emphasizes the importance of human–AI collaboration. ChatGPT should be perceived not as a replacement for professional expertise but as an analytical co-pilot that assists auditors in data interpretation, documentation, and compliance analysis. Hospital administrators are encouraged to invest in AI training, data governance, and phased implementation strategies to maximize the benefits while mitigating ethical and operational risks.

Despite its strong results, the study's limitations—notably the small sample size (n = 5) and focus on a single hospital group—restrict the generalizability of the conclusions. Future research should therefore expand the sample across multiple healthcare institutions and investigate longitudinal effects of AI adoption on audit quality, efficiency, and decision-making.

In conclusion, this research demonstrates that ChatGPT represents a transformative tool for internal auditing in healthcare. When integrated responsibly and supported by skilled human oversight, it can significantly enhance transparency, compliance, and organizational accountability, contributing to the ongoing digital transformation of hospital governance in Morocco and beyond.

## Bibliography

Bouziane, A. (2025). Assessing the Influence of technological tools on the quality of computer-assisted audits: The case of Moroccan audit firms. *EDPACS: The EDP Audit, Control and Security Newsletter, 70*(5), 45–59. https://doi.org/10.1080/07366981.2025.2454094

Bouziane, A., Tahri, W., Bouziane, K. (2023). Satisfaction and acceptance of ICT in learning activities during Covid19: The case of Moroccan students. In Ł,Tomczyk (Ed.), *New Media Pedagogy: Research Trends, Methodological Challenges and Successful Implementations.* (Vol 1916, pp.254-265). Springer. https://doi.org/10.1007/978-3-031-44581-1_18

DeAngelo, L. E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*(3), 183-199. https://doi.org/10.1016/0165-4101(81)90002-1

Emett, S., Eulerich, M., Lipinski, E., Prien, N., & Wood, D. A. (2025). Leveraging ChatGPT for enhancing the internal audit process—A real-world example from Uniper, a large multinational company. *Accounting Horizons, 39*(2), 125–135. https://doi.org/10.2308/HORIZONS-2023-111

Eulerich, M., & Wood, D. A. (2023, July 24). *A demonstration of how ChatGPT can be used in the internal auditing process* [Preprint]. *SSRN.* http://dx.doi.org/10.2139/ssrn.4519583

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x

Fotoh, L., & Mugwira, T. (2023, May 19). *Exploring large language models (ChatGPT) in external audits: Implications and ethical considerations* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4453835

Gu, H., Schreyer, M., Moffitt, K., & Vasarhelyi, M. (2024). Artificial intelligence co-piloted auditing. *International Journal of Accounting Information Systems, 54*, Article100698. https://doi.org/10.1016/j.accinf.2024.100698

IBM Corp. (n.d). *IBM SPSS Statistics (Version 26*) [Computer software]. IBM Corp. https://www.ibm.com/products/spss-statistics

Law, K. K. F., & Shen, M. (2025). How does artificial intelligence shape audit firms? *Management Science, 71*(5), 3601-3628. https://doi.org/10.1287/mnsc.2022.04040

Zhang, C. (2019). Intelligent process automation in audit. *Journal of Emerging Technologies in Accounting, 16*(2), 69–88. https://doi.org/10.2308/jeta-52653

# THE IMPACT OF AI-ASSISTED TRANSLATION ON INTERNAL AUDIT QUALITY IN AN INTERNATIONAL CORPORATION

**Abdelmounim Bouziane[1] & Karima Bouziane**

*[1]LIRO Lab, National School of Commerce and Management*
*Chouaib Doukkali University*

*[2]ALCS Lab, Faculty of Letters and Humanities*
*Chouaib Doukkali University*

**ABSTRACT**

This study examines the impact of AI-assisted translation on the quality of internal audits within a multinational corporation whose parent company is based in Morocco, with subsidiaries and participations in various countries. Given the multilingual nature of internal audit processes, AI-powered translation tools have become essential for ensuring clear communication, accurate documentation, and effective risk assessment. However, their impact on audit quality remains underexplored. Through a mixed-method approach combining qualitative and quantitative analysis, this research assesses how AI-assisted translation affects auditors' efficiency, comprehension of regulatory frameworks, and overall audit reliability. The study investigates whether these tools enhance or hinder the internal audit process, particularly in terms of accuracy, consistency, and compliance with international auditing standards. Findings will provide insights into the advantages and limitations of AI-driven translation in internal auditing and offer recommendations for optimizing its use in multinational corporate environments.

***Keywords**: AI, Translation, Audit, Audit Quality, ChatGPT.*

## 1. Introduction

Globalization has significantly transformed corporate governance and internal audit practices, compelling multinational corporations to operate within complex linguistic, cultural, and regulatory environments. Effective internal auditing depends heavily on clear communication, accurate interpretation of documents, and a precise understanding of local and international standards. However, in multilingual corporate structures, language barriers often hinder information flow, increase the risk of misinterpretation, and affect the consistency of audit conclusions. These challenges highlight the growing need for technological tools that can bridge linguistic gaps and enhance the reliability of audit processes.

The emergence of artificial intelligence (AI)-assisted translation tools, such as ChatGPT and other neural machine translation systems, offers promising opportunities to address these challenges. Unlike traditional translation software, AI-powered systems use deep learning and natural language processing to interpret context, detect nuances, and generate more accurate and coherent translations. Their application in internal auditing can facilitate the comprehension of complex regulatory frameworks, improve cross-border communication between audit teams, and ensure greater consistency in audit documentation.

Nevertheless, despite the growing use of AI-assisted translation in international business operations, its impact on internal audit quality remains largely underexplored. Questions persist regarding its reliability, the risk of semantic distortion, and its effect on auditors' professional judgment and independence.

This study seeks to assess the impact of AI-assisted translation on the quality of internal audits conducted within a multinational corporation headquartered in Morocco. By examining how these tools influence audit efficiency, accuracy, and compliance with international auditing standards, the research aims to determine whether AI-driven translation enhances or compromises the integrity of the audit process and to provide practical recommendations for its optimized use in multilingual audit environments.

## 2. Literature Review

Auditing firms play a central role in ensuring the credibility and reliability of financial information. Their primary responsibility is to control and certify financial statements, providing stakeholders with assurance that the data faithfully represents the entity's true financial position. However, as organizations become increasingly globalized and engage in thousands of transactions daily, auditors face growing complexity, particularly when coordinating across multilingual environments. In this context, the emergence of artificial intelligence (AI) has revolutionized the auditing profession, enabling automation, improved efficiency, and enhanced accuracy (Sahut et al., 2013; van den Broek & van Veenstra, 2018). For multinational corporations, AI-assisted translation tools now play a crucial role in bridging linguistic gaps within internal audit processes, supporting auditors in maintaining consistency and clarity across jurisdictions.

### 2.1. AI Integration and Transformation of Audit Firms

The integration of AI in audit firms has led to a transformation of their economic models and organizational structures. Law and Shen (2025) examined how AI adoption reshapes auditor roles, revealing that firms incorporating AI experience both an expansion of audit positions and a shift in skill requirements — emphasizing analytical and interpersonal competencies over technical coding skills. Importantly, these AI-enabled firms demonstrate fewer audit delays and restatements, reflecting improvements in audit quality. Such findings suggest that AI does not replace auditors but enhances their performance and adaptability within technologically augmented environments.

Fedyk et al. (2022) further provide empirical evidence that investment in AI talent significantly improves audit outcomes. Analyzing data from major U.S. audit firms, they found that AI-driven systems reduced material restatements and increased productivity. However, the study also cautions against uneven benefits distribution: while senior auditors profit from enhanced efficiency and reduced costs, junior auditors may face role redefinition. This underscores the need for strategic integration and training, particularly when AI tools, such as translation systems, are introduced to support multilingual audit engagements.

### 2.2. AI-Assisted Translation and Audit Quality

The increasing use of AI-based translation tools represents a crucial innovation in the audit profession, especially in multinational contexts where auditors must interpret and apply international standards in multiple languages. These tools facilitate communication and reduce misunderstandings that could affect audit reliability and compliance. Drawing on Intelligent Process Automation (IPA) frameworks (Zhang, 2019), AI translation can be understood as part of the automation continuum that enhances auditors' efficiency by managing repetitive linguistic and data verification tasks, allowing professionals to concentrate on judgment-based activities.

Fotoh and Mugwira (2023) explored the integration of ChatGPT in external audits, emphasizing its capacity to generate human-like responses and streamline repetitive audit tasks. Their analysis, based on dynamic capabilities and social presence theories, revealed that ChatGPT can enhance auditors' understanding, documentation, and collaboration—particularly beneficial when audits span multiple languages. However, concerns persist regarding accuracy, data reliability, and ethical implications, such as confidentiality and bias. These findings are essential for understanding how AI-assisted translation may improve comprehension and efficiency while posing risks to audit quality if outputs are not rigorously validated.

### 2.3. Practical Applications and Case Studies

Several practical studies demonstrate how ChatGPT and similar models can directly improve audit efficiency and cross-border communication. Emett et al. (2023) investigated the implementation of ChatGPT at Uniper, a multinational energy firm, and reported that AI integration reduced audit task time by 50–80%. The tool supported auditors in preparing reports, interpreting documents, and writing conclusions, illustrating its potential to streamline multilingual reporting and translation. Nonetheless, the study highlights ongoing challenges concerning data privacy, ethical compliance, and accuracy validation—issues that are equally relevant to AI-assisted translation in internal audits.

In parallel, Gu et al. (2023) introduced the concept of “AI co-piloted auditing,” where auditors collaborate with language models such as GPT-4 to perform reasoning-intensive tasks. Their chain-of-thought prompting method demonstrated improved precision in financial analysis and report interpretation. Adapting this approach to translation implies that AI and human auditors can work collaboratively to ensure contextually accurate, standardized translations of audit documentation and international standards, thereby enhancing reliability and compliance.

Finally, Eulerich and Wood (2023) offered practical frameworks for integrating ChatGPT into internal audit functions, emphasizing its role in task automation, risk assessment, and report generation. Their findings reveal that AI tools can strengthen internal audit efficiency while raising critical considerations regarding data governance, algorithm transparency, and cybersecurity. These insights directly inform the application of AI in translation tasks, suggesting that proper oversight and ethical frameworks are vital to safeguarding audit quality.

### 2.4. Synthesis and Research Gap

The reviewed literature consistently demonstrates that AI adoption—particularly through tools like ChatGPT—enhances efficiency, accuracy, and analytical depth within auditing processes. However, most existing research emphasizes automation, data analytics, and process optimization, while the specific impact of AI-assisted translation on audit quality remains underexplored. Given that internal audits in multinational corporations often operate across multiple languages, translation accuracy directly affects auditors’ comprehension of regulatory frameworks, risk assessment, and communication clarity.

Therefore, this study addresses a significant gap by investigating how AI-assisted translation tools influence audit reliability, consistency, and compliance in cross-border settings. Building on the foundational works of Emett et al. (2023), Eulerich and Wood (2023), Fedyk et al. (2022), Fotoh and Mugwira (2023), Gu et al. (2023), Zhang (2019), and it explores the extent to which AI-powered translation enhances or hinders internal audit quality. The findings will contribute to the ongoing discourse on balancing technological innovation with human professional judgment, offering actionable insights for auditors, regulators, and multinational corporations adopting AI in their internal control systems.

## 3. Methodology

### 3.1. Research Design

This study adopted a mixed-method approach combining qualitative and quantitative techniques to assess the impact of AI-assisted translation on internal audit quality within a multinational corporation headquartered in Morocco. The mixed design enabled both an in-depth understanding of auditors' perceptions and a statistical validation of the relationships among key variables derived from the Technology Acceptance Model (TAM) and the literature review.

### 3.2. Data Collection

Two complementary data-collection instruments were employed:

- Semi-structured interviews with four experienced auditors from the internal audit department. These interviews aimed to capture nuanced perceptions regarding the integration of AI-based translation tools—particularly ChatGPT—into audit tasks, focusing on perceived usefulness, ease of use, and perceived effects on audit quality.
- A structured questionnaire distributed to the corporation's 15-member audit team. The questionnaire included both closed-ended and Likert-scale questions assessing auditors' familiarity with AI technologies, their experience using ChatGPT for translation, perceived ease of use, perceived usefulness, and perceived impact on audit quality.

### 3.3. Variables and Measurement

Independent variables were conceptualized according to the Technology Acceptance Model (TAM) (Davis, 1989) and extended through factors identified in the literature:

- Perceived Ease of Use (PEOU): the extent to which auditors find AI translation tools easy to learn and operate.
- Perceived Usefulness (PU): the degree to which auditors believe that AI-assisted translation improves efficiency and accuracy.

- Experience with ChatGPT Translation (EXP): auditors' prior practical use of ChatGPT or similar AI tools for translation tasks.
- Familiarity with AI Tools (FAM): overall exposure to and comfort with artificial-intelligence applications.
- Control Variables: gender, academic level, and years of professional experience.

The dependent variable was Perceived Audit Quality (AQ), operationalized through items reflecting accuracy, consistency, and compliance of internal audit documentation following the adoption of AI translation tools.

All items were measured using a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

### 3.4. Data Analysis

Quantitative data were analyzed using multiple linear regression to identify which factors significantly influence perceived audit quality. Statistical tests included:

- Coefficient of Determination ($R^2$) to measure explanatory power;
- F-Statistic to evaluate overall model significance;
- Durbin–Watson (DW) test to verify independence of residuals.

Qualitative data from interviews were coded thematically to complement and contextualize quantitative findings, highlighting patterns of perception, acceptance, and resistance among auditors.

## 4. Results

### 4.1. Descriptive Statistics

Of the 15 questionnaires distributed, all were returned, yielding a 100 % response rate. Respondents comprised 60 % male and 40 % female auditors, with diverse academic backgrounds (mainly accounting, auditing, and finance) and professional experience ranging from 3 to 15 years. The majority reported having experimented with AI tools, though the level of familiarity varied.

### 4.2. Regression Analysis

The regression model produced a high explanatory power, with $R^2 = 0.897$, indicating that approximately 89.7 % of the variance in perceived audit quality is explained by the independent variables. The F-statistic was significant at a p-value $< 0.05$, confirming that the model is statistically meaningful. The Durbin–Watson coefficient (DW = 2.1) indicates no autocorrelation, supporting the model's robustness.

**Table 1**

*Summary of Regression Results*

| Variable | Standardized β | Significance (p) | Interpretation |
|---|---|---|---|
| **Ease of Use (PEOU)** | 0.750 | < 0.05 | Strong positive influence on audit quality |
| **Experience with ChatGPT (EXP)** | 0.768 | < 0.05 | Strong positive influence on audit quality |
| **Usefulness (PU)** | 0.432 | > 0.05 | Moderate, non-significant effect |
| **Familiarity with AI Tools (FAM)** | 0.318 | > 0.05 | Limited impact |
| **Academic Level** | 0.142 | > 0.05 | Minimal effect |
| **Gender** | –0.097 | > 0.05 | No significant effect |
| **Professional Experience** | 0.183 | > 0.05 | Weak positive effect |

### 4.3. Key Findings

The results demonstrate that Ease of Use and Experience with ChatGPT translation are the most influential factors affecting perceived audit quality. Auditors who find AI-assisted translation intuitive and have prior experience using ChatGPT report substantially higher levels of confidence in translation accuracy, consistency of documentation, and overall audit efficiency.

Other factors such as usefulness, familiarity with AI, and demographic variables (academic level, gender, and experience) showed limited or statistically insignificant effects, suggesting that the practical usability of AI tools outweighs general attitudes or background characteristics in determining audit quality.

### 4.4. Qualitative Insights

Interview analysis reinforced the quantitative results. Participants emphasized that AI-assisted translation notably reduced manual workload, enhanced cross-border communication, and minimized linguistic ambiguities in audit documentation. However, they also expressed caution regarding the need for human validation of AI-generated translations to ensure contextual and regulatory accuracy.

## 5. Discussion

The findings of this study confirm that integrating AI-assisted translation tools, particularly ChatGPT, can significantly enhance the perceived quality of internal audits in multinational contexts. The high explanatory power of the regression model ($R^2 = 0.897$) indicates that the selected variables—drawn from both the Technology Acceptance Model (TAM) and prior empirical studies—collectively provide a robust explanation of how auditors perceive and experience AI in their professional activities.

### 5.1. Influence of Ease of Use and Experience with ChatGPT

The results reveal that ease of use ($\beta = 0.750$) and experience with ChatGPT ($\beta = 0.768$) are the strongest predictors of audit quality. This outcome aligns closely with the TAM framework proposed by Davis (1989), which posits that perceived ease of use directly affects user acceptance and performance outcomes. In the context of auditing, when auditors find AI translation tools intuitive and user-friendly, they are more likely to trust and integrate them into their daily workflow, thereby improving efficiency and precision.

This finding is consistent with Zhang (2019), who demonstrated that intelligent automation can enhance auditors' efficiency by eliminating repetitive manual tasks and enabling greater focus on analytical reasoning. Similarly, Fotoh and Mugwira (2023) reported that ChatGPT's natural-language capabilities simplify complex or repetitive communication processes, allowing auditors to allocate more cognitive resources to judgment-intensive tasks.

Moreover, auditors who had prior experience using ChatGPT for translation tasks perceived significantly higher audit quality. This observation echoes Fedyk et al. (2022), who showed that experience and familiarity with AI systems positively correlate with audit quality improvements. The current study adds nuance by showing that not only AI familiarity in general, but specifically hands-on experience with ChatGPT translation, enhances perceived audit accuracy and consistency in multilingual audit environments.

### 5.2. Limited Role of Usefulness, Familiarity, and Demographic Variables

Contrary to expectations derived from TAM, perceived usefulness exhibited only a moderate and statistically non-significant influence on audit quality. One possible explanation is that auditors may recognize the usefulness of AI translation tools conceptually but remain cautious about their reliability, particularly regarding linguistic precision and contextual appropriateness—concerns that have been raised in earlier research. Fotoh and Mugwira (2023) and Eulerich and Wood (2023) both noted that ethical and accuracy-related issues, such as biased outputs or privacy risks, can limit users' perceived utility of ChatGPT in professional settings.

Similarly, familiarity with AI tools and demographic factors such as academic level, gender, and professional experience showed minimal influence. This outcome supports Law and Shen (2025), who found that while AI adoption changes the skill composition of audit teams, the performance benefits are driven more by task-specific competencies and training than by demographic or educational background. Thus, audit quality gains depend less on who the auditors are and more on how effectively they interact with AI systems.

### 5.3. Alignment with Prior Evidence on AI Integration

The positive association between ChatGPT use and audit quality reinforces previous evidence on AI's capacity to enhance both efficiency and accuracy within audit processes. For instance, Emett et al. (2023) demonstrated that ChatGPT reduced audit task time by 50–80 %, paralleling the present study's qualitative finding that AI-assisted translation reduces the time needed to interpret and prepare multilingual audit documentation. Likewise, Gu et al. (2023) conceptualized "AI co-piloted auditing," emphasizing how large language models can collaborate with human auditors to improve reasoning and standard interpretation. The current findings support this co-pilot framework by illustrating how ChatGPT's translation capabilities complement auditors' expertise, fostering clearer, standardized communication across subsidiaries operating in different linguistic contexts.

### 5.4. Implications for Audit Quality and Cross-Border Practice

From a practical standpoint, these results suggest that AI-assisted translation tools are not merely linguistic aids but strategic enablers of audit quality in multinational environments. By enhancing the clarity and consistency of audit documentation, they contribute directly to compliance with international auditing standards and minimize risks of misinterpretation arising from language barriers.

However, as Eulerich and Wood (2023) caution, the introduction of AI also entails risks related to data confidentiality, algorithmic transparency, and bias. In the Moroccan context—where audit firms often operate under dual linguistic frameworks (Arabic–French–English)—the challenge lies in establishing robust validation protocols to verify the accuracy of AI-generated translations and ensure alignment with regulatory terminology. Thus, while ease of use and experience promote adoption, safeguards for ethical and linguistic integrity remain essential to sustaining audit quality.

### 5.5. Theoretical and Managerial Contributions

Theoretically, this study extends the Technology Acceptance Model by integrating new constructs—experience with ChatGPT translation and AI familiarity—within an audit-quality framework. The high $R^2$ value demonstrates that extending TAM with task-specific variables strengthens its predictive validity in professional settings that combine linguistic and technical dimensions.

Managerially, the findings emphasize the importance of training and user experience in maximizing the benefits of AI tools. Firms should prioritize continuous learning programs that familiarize auditors with ChatGPT and similar platforms, encouraging experimentation and feedback loops to improve translation accuracy. Furthermore, developing AI governance policies covering confidentiality, data management, and ethical usage will be vital to ensuring sustainable and compliant adoption.

## 6. Recommendations

Based on the findings and discussion, several recommendations can be formulated for audit firms, regulators, and future researchers to optimize the integration of AI-assisted translation tools, particularly ChatGPT, in internal auditing processes.

### 6.1. Managerial Recommendations

- Promote Targeted Training and Capacity Building- Given that ease of use and experience with ChatGPT were the strongest predictors of audit quality, firms should invest in regular, hands-on training sessions for auditors. Workshops and simulation exercises can familiarize staff with AI translation interfaces, helping them develop confidence in verifying and editing AI-generated outputs.
- Establish AI Governance and Data Integrity Protocols-Audit firms should implement formal AI governance frameworks covering data confidentiality, privacy, and translation validation. Internal audit teams must ensure that AI-translated content complies with professional and ethical standards and that sensitive client information remains secure during processing.
- Integrate Human Oversight and Validation Mechanisms-Despite the efficiency gains from ChatGPT, human auditors must retain final responsibility for reviewing and approving translated reports. Introducing a “human-in-the-loop” validation model can prevent errors, mitigate biases, and ensure contextual accuracy across multilingual audit documentation.

- Develop Internal Databases and Terminology Libraries-Creating internal repositories of standardized audit terms in multiple languages will enhance translation consistency. Linking these glossaries with AI tools can reduce linguistic ambiguity and align translations with local regulatory and cultural nuances, particularly in cross-border engagements.
- Encourage a Balanced Human–AI Collaboration Model-AI tools should be positioned as co-pilots rather than substitutes. Audit leaders are encouraged to cultivate an environment where technology complements professional judgment. Such collaboration can reinforce auditors' analytical role while improving workflow efficiency.
- Evaluate and Monitor AI Performance Continuously-Firms should routinely assess the accuracy, reliability, and ethical compliance of AI-assisted translation systems. Performance indicators—such as translation error rates, turnaround time, and user satisfaction—should guide iterative improvements and inform future adoption strategies.

**6.2. Policy and Regulatory Recommendations**

- Incorporate AI Translation Standards in Professional Guidelines Regulatory bodies and professional associations should update auditing standards to include guidance on AI translation use. Clear policies regarding data handling, validation, and accountability will foster responsible adoption within the auditing profession.
- Promote Transparency and Accountability in AI Use-Audit firms should disclose the extent of AI involvement in audit reports, particularly where translations influence the interpretation of financial or compliance data. Transparency will strengthen stakeholder trust in AI-augmented auditing practices.
- Encourage Multilingual Compliance Training-Given the multilingual realities of multinational firms, regulators could support initiatives that promote the harmonization of audit terminology across languages, ensuring that AI-translated documents maintain consistency with international auditing standards (*e.g.,* ISA, IFRS).

### 6.3. Academic and Research Recommendations

- Expand Empirical Research on AI-Assisted Translation in Auditing-Future studies should replicate and extend this research across different industries, countries, and linguistic settings to validate the findings and generalize the relationship between AI translation and audit quality.
- Investigate Longitudinal Effects on Audit Efficiency and Ethics-While the current study focused on perceived quality, further research could examine the long-term implications of AI adoption on audit efficiency, ethical decision-making, and workforce transformation.
- Explore Hybrid Models of Translation Quality Assurance-Comparative studies assessing human-only, AI-only, and hybrid translation approaches would yield valuable insights into optimizing translation accuracy and reliability in audit contexts.
- Incorporate Additional Behavioral and Cultural Variables Future research could integrate cross-cultural communication factors, cognitive load, or organizational culture variables to better understand how auditors adapt to AI tools in multilingual settings.

## 7. Conclusion

This study set out to examine the impact of AI-assisted translation, particularly through ChatGPT, on the quality of internal audits within a multinational corporate context. Grounded in the Technology Acceptance Model (TAM) and supported by insights from prior literature, the research sought to determine how variables such as ease of use, perceived usefulness, familiarity with AI tools, and experience with ChatGPT translation influence perceived audit quality.

Through a mixed-method design combining interviews and quantitative analysis, the results revealed a strong and statistically significant relationship between ease of use and experience with ChatGPT translation and the perceived enhancement of audit quality ($R^2 = 0.897$). These findings indicate that when AI-assisted translation tools are intuitive and auditors gain practical experience with them, the accuracy, consistency, and efficiency of audit documentation improve substantially. Other variables—including perceived usefulness, familiarity, academic level, gender, and experience—had only marginal influence, suggesting that usability and hands-on engagement outweigh demographic factors in shaping audit quality outcomes.

The results are consistent with previous studies (Emett et al., 2023; Eulerich & Wood, 2023; Fedyk et al., 2022; Fotoh & Mugwira, 2023; Law & Shen, 2025; Zhang, 2019), which collectively affirm that AI enhances audit processes when supported by effective user interaction and governance mechanisms. The study also highlights the crucial role of human oversight and ethical frameworks in ensuring that AI-generated translations remain reliable, contextually accurate, and compliant with international auditing standards.

From a practical standpoint, the findings underscore that AI-assisted translation tools are not a replacement for human expertise, but a complementary instrument that augments auditors' analytical capacity and facilitates multilingual communication. In multinational corporations—such as those operating in Morocco with subsidiaries abroad—these tools can significantly improve cross-border audit coordination and understanding of regulatory frameworks, provided that proper governance, validation, and training systems are in place.

Finally, the study contributes to both academic and professional discourse by extending the TAM model to the auditing domain and introducing AI translation experience as a new determinant of audit quality. Future research should continue to explore this evolving field by assessing longitudinal impacts, cross-cultural variables, and hybrid human–AI collaboration models in different auditing contexts.

In conclusion, the integration of ChatGPT-powered translation tools represents a promising advancement for internal auditing, offering new pathways for improving efficiency, communication, and reliability—while reminding practitioners that technological progress must remain anchored in ethical responsibility and professional judgment.

## References

Emett, S. A., Eulerich, M., Lipinski, E., Prien, N., & Wood, D. A. (2023, July 18). *Leveraging ChatGPT for enhancing the internal audit process: A real-world example from a large multinational company.* [Preprint]. SSRN. http://dx.doi.org/10.2139/ssrn.4514238

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Gu, H., Schreyer, M., Moffitt, K., & Vasarhelyi, M. (2024). Artificial intelligence co-piloted auditing. *International Journal of Accounting Information Systems, 54*, Article100698. https://doi.org/10.1016/j.accinf.2024.100698

Law, K. K. F., & Shen, M. (2025). How does artificial intelligence shape audit firms? *Management Science, 71*(5), 3601-3628. https://doi.org/10.1287/mnsc.2022.04040

Sahut, J. M., Hikkerova, L., & Khalfallah, M. (2013). Business model and performance of firms. *International Business Research, 6*(2), 64–76.https://doi.org/10.5539/ibr.v6n2p64

van den Broek, T., & van Veenstra, A. F. (2018). Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. *Technological Forecasting and Social Change, 129*, 42–52. https://doi.org/10.1016/j.techfore.2017.09.040

Zhang, C. (2019). Intelligent process automation in audit. *Journal of Emerging Technologies in Accounting, 16*(2), 69–88. https://doi.org/10.2308/jeta-52653

Fotoh, L., & Mugwira, T. (2023, May 19). *Exploring large language models (ChatGPT) in external audits: Implications and ethical considerations* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4453835

Eulerich, M., & Wood, D. A. (2023, July 24). *A demonstration of how ChatGPT can be used in the internal auditing process* [Preprint]. *SSRN*. http://dx.doi.org/10.2139/ssrn.4519583

# TOWARDS CONTINUOUS AUDITING THROUGH AI: OPPORTUNITIES, CHALLENGES, AND FUTURE PERSPECTIVE

**Sarah OUBERNI, Doha HARRAG & Issam EL FILALI**

*Chouaib Doukkali University, LIRO, ENCG,*
*El Jadida, Morocco*

**ABSTRACT**

In the era of digital transformation, Artificial Intelligence (AI) is reshaping traditional auditing practices by enabling the transition from periodic to continuous auditing. This paper explores the potential of AI in enhancing audit processes through real-time data analysis, anomaly detection, and predictive insights. While the integration offers substantial opportunities such as increased audit efficiency, improved risk management, and enhanced fraud detection it also presents significant challenges.

This research aims to explore how AI contributes to the emergence and development of continuous auditing, with a focus on identifying the main opportunities, challenges, and strategic perspectives associated with this shift. The objective is to provide a comprehensive understanding of the current state of research on this topic and to map the intellectual structure of the field. It is of particular interest as it addresses a critical evolution in the audit profession, responding to the growing demands for real-time assurance and greater transparency. The shift towards continuous auditing through AI is not merely a technological advancement but a shift in how audits are conceptualized and executed.

A bibliometric analysis will be conducted to map the academic landscape related to continuous auditing and AI. This method will provide a comprehensive overview of the evolution, trends, and knowledge gaps in the field, laying the groundwork for future empirical research.

The findings will contribute to a deeper understanding of how AI is reshaping the audit landscape and will offer valuable insights for academics, professionals, and regulators seeking to adapt audit practices to the digital age. Additionally, the study will highlight underexplored areas, suggesting avenues for future empirical research.

***Keywords:*** *Continuous Auditing; Artificial Intelligence; Audit Automation; Real-time Assurance; Machine Learning in Auditing.*

## 1. Introduction

The accelerating digital transformation of financial and corporate environments is reshaping how assurance and audit functions are conceptualized and executed. In an era where organizations operate with vast, real-time data flows, the traditional model of periodic auditing characterized by retrospective assessments and static verification appears increasingly inadequate. Stakeholders now demand continuous assurance, transparency, and agility in financial reporting, urging auditors to move beyond episodic evaluations toward a more dynamic, technology-enabled paradigm. Artificial Intelligence (AI), through its capabilities in automation, anomaly detection, and predictive analytics, has emerged as a transformative enabler of this paradigm shift.

The integration of AI into auditing represents more than a technological innovation; it signifies a conceptual reorientation of the audit process. By leveraging machine learning and data analytics, AI systems can monitor transactions continuously, detect irregularities in real time, and generate predictive insights that enhance decision-making. Such tools promise increased efficiency, enhanced fraud detection, and improved risk management (Leocádio et al., 2024; Noordin et al., 2022). However, this transformation also introduces profound challenges. The growing reliance on algorithmic systems raises critical concerns regarding data governance, algorithmic transparency, and ethical accountability in audit practices. This perspective is consistent with the literature emphasizing that intelligent technologies are intended to support rather than replace auditors' professional judgment (Omoteso, 2012).

Despite its promising potential, the path toward continuous auditing through AI remains underdeveloped in both theory and practice. Current audit frameworks often struggle to align technological innovation with established ethical and regulatory standards. AI applications are frequently implemented without clear guidelines for data protection, explainability, and audit independence. This gap between innovation and governance creates uncertainty about how to design AI-driven audit systems that are secure, reliable, and transparent.

Consequently, this study aims to explore how AI contributes to the emergence and development of continuous auditing, focusing on identifying its principal opportunities, challenges, and future perspectives. Specifically, the research addresses two guiding questions inspired by the problem statement and central inquiry formulated in the project:

1. How can AI be used to implement secure, reliable, and transparent continuous auditing models?

2. What are the main opportunities, challenges, and strategic implications associated with the integration of AI in continuous auditing?

By addressing these questions, the article seeks to provide a comprehensive understanding of the transition from static to dynamic audit models. Through a bibliometric analysis of academic literature, combined with a review of technological trends and regulatory developments, this research aims to map the intellectual landscape of AI-driven auditing. The objective is to clarify how this emerging field responds to the growing demand for real-time assurance and trustworthiness in the digital economy, while highlighting the conditions under which AI can enhance audit quality, independence, and reliability.

This study makes three main contributions to the existing literature. First, it provides a structured bibliometric mapping of research on AI in continuous auditing, offering a comprehensive overview of the field's intellectual evolution. Second, it identifies and organizes the main thematic clusters shaping current research, including technological innovation, ethical challenges, and professional transformation. Third, it highlights critical research gaps and proposes avenues for future investigation, particularly regarding regulatory frameworks, algorithmic transparency, and the integration of human judgment in AI-driven audit systems.

## 2. Research Design and Analytical Approach

The exploration of how intelligent technologies are reshaping auditing practices requires a methodological framework that reflects both the evolution of ideas and the dynamics of technological adoption. In this study, the analytical design combines bibliometric mapping of academic research with an interpretative analysis of current professional trends, offering a dual lens that captures the theoretical and applied dimensions of continuous auditing's emergence.

This framework was developed in line with the increasing scholarly interest in understanding how automation and intelligent data processing redefine the foundations of financial assurance. The academic analysis focused on identifying the main intellectual contributions that have structured the debate around continuous auditing, ethics in algorithmic decision-making, and the professional adaptation of auditors. Publications and sources from the Scopus and Web of Science databases were examined to trace how the field has evolved conceptually, revealing the progressive transition from traditional, retrospective assurance models toward dynamic systems of ongoing verification.

The bibliometric perspective allowed the mapping of influential authors, recurrent concepts, and emerging thematic clusters. Using visual analysis tools such as VOSviewer, the study examined how key notions such as transparency, automation, governance, and ethical accountability interconnect within the scientific discourse. This mapping offered a panoramic understanding of how the academic community conceptualizes the relationship between intelligent technologies and assurance processes. The results highlight a convergence of research around three thematic poles: technological innovation in auditing, ethical and regulatory considerations, and professional competencies for a data-driven environment (van Eck & Waltman, 2010).

Beyond academic literature, technological and contextual analysis was inspired by recent professional developments in audit and financial governance. Industry reports and international standards increasingly emphasize the adoption of real-time audit mechanisms supported by algorithmic systems, blockchain integration, and cryptographic protocols for data validation. By situating these developments within the academic debate, the analysis underscores how technological infrastructures are progressively shaping new expectations for audit transparency, reliability, and speed. This interdisciplinary reading aligns with the growing recognition that the transformation of auditing is not purely technical but deeply epistemological and ethical.

Through this combined analytical approach, the study establishes a foundation for understanding how the convergence of automation, data intelligence, and ethical governance is redefining the assurance function. The bibliometric component elucidates the intellectual evolution of the field, while the contextual analysis translates these insights into the realities of practice and regulation. This integrative methodology thus provides a coherent basis for discussing the opportunities, challenges, and strategic implications of the ongoing transition toward continuous auditing, setting the stage for the interpretation of results that follows.

Bibliometric analysis has been widely recognized as a robust method for mapping scientific knowledge and identifying research trends (Donthu et al., 2021).

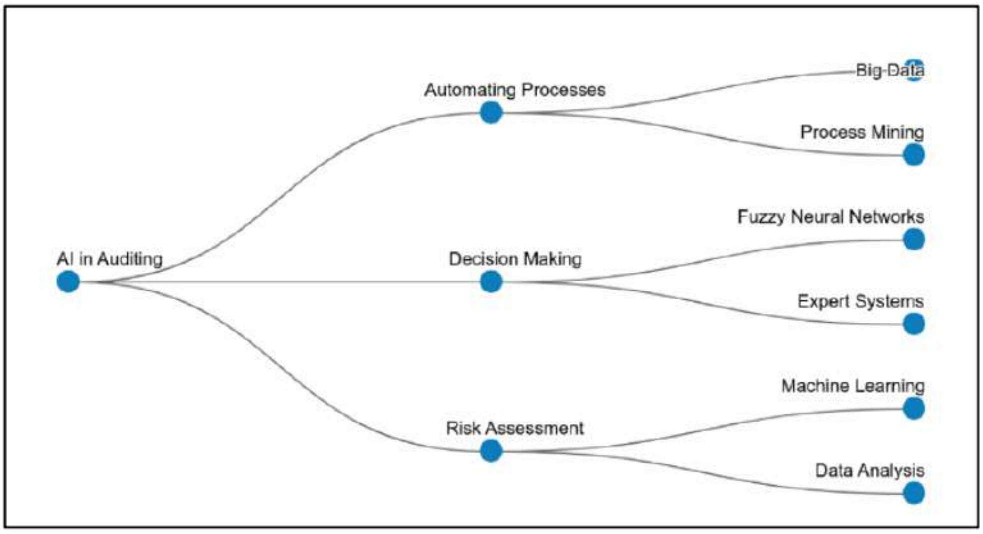
### 2.1. The Evolution of Continuous Auditing in the Intelligent Era: Opportunities and Transformations

The rise of intelligent systems within auditing marks a profound shift in the epistemology of assurance. Beyond the technological enthusiasm that often surrounds innovation, this evolution reflects a redefinition of auditing's purpose and practice. The bibliometric mapping and interpretative analysis undertaken in this study reveal that the transformation toward continuous auditing unfolds along two main trajectories: the emergence of strategic opportunities and the persistence of ethical and organizational challenges.

The first conceptual map (see Figure 1: Concept Map of AI Opportunities in Auditing) visually illustrates the thematic clusters identified in the bibliometric analysis. These clusters converge around three pivotal dimensions: efficiency and automation, risk prediction and fraud detection, and strategic value creation within organizations.

**Figure 1**

*Concept Map of AI Opportunities in Auditing*




**Note**. Source : Scopus AI https://www.scopus.com/search/form.uri?display=basic#scopus-ai (13.08.2024)

Figure 1 presents a conceptual map derived from keyword co-occurrence analysis. The clusters were generated using VOSviewer based on the frequency and co-occurrence of terms within the dataset.

Three main clusters emerge:

- Cluster 1 (Technological Innovation): Includes terms such as Big Data, process mining, machine learning, and fuzzy neural networks, highlighting the technological foundations of continuous auditing.
- Cluster 2 (Risk and Fraud Detection): Focuses on anomaly detection, predictive analytics, and risk management, emphasizing the proactive role of AI in auditing.
- Cluster 3 (Strategic and Organizational Impact): Encompasses audit quality, decision-making, and value creation, reflecting the broader implications of AI integration.

These clusters illustrate how technological tools, analytical capabilities, and strategic outcomes are interconnected in the evolution of continuous auditing. They confirm that intelligent systems enable auditors to move from retrospective examination to continuous, data-driven observation. Real-time monitoring capabilities transform assurance into a living process that evolves with organizational data flows. This capacity not only improves transparency but also enhances the timeliness of corrective action and managerial oversight.

Another major opportunity lies in predictive analytics and fraud anticipation. Intelligent tools can detect irregularities that human observation may overlook, enabling a proactive stance toward risk management and anomaly detection. Leocádio et al. (2024) emphasize that such systems reorient auditing from compliance verification toward strategic foresight. As a result, auditors are increasingly positioned as interpreters of data patterns and advisors on systemic vulnerabilities rather than mere verifiers of transactions.

Equally significant is the transformation of professional identity. As automation reduces repetitive tasks, auditors gain greater capacity to exercise critical reasoning and ethical judgment. The profession thus experiences an evolution toward roles that combine technical understanding with analytical discernment and human empathy (Raschke et al., 2018).

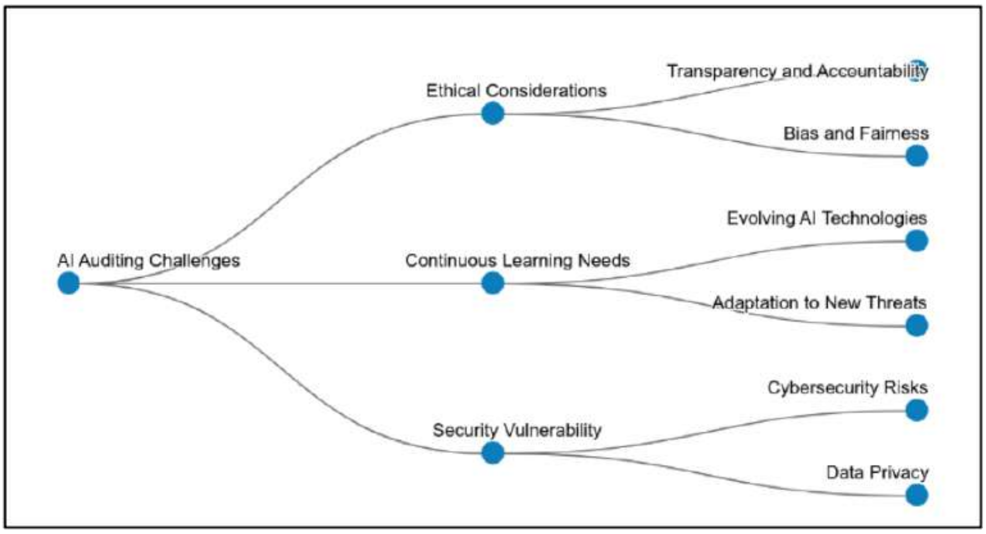
Collectively, these findings show that the movement toward continuous auditing opens a new strategic horizon in which assurance becomes both preventive and participatory embedding auditors more deeply in the governance structures of organizations.

### 2.2. Ethical, Regulatory, and Human Challenges of Intelligent Assurance

The second conceptual map (see Figure 2: Concept Map of AI Challenges in Auditing) reveals the other side of this transformation: the structural constraints that accompany innovation. Three key areas dominate the challenge landscape: data governance and transparency, regulatory adaptation, and professional competence and trust.

**Figure 2**

*Concept Map of AI Challenges in Auditing*




***Note**. Source : Scopus AI https://www.scopus.com/search/form.uri?display=basic#scopus-ai (13.08.2024)*

Figure 2 illustrates the main challenges associated with AI-driven auditing. The clusters highlight three interconnected dimensions:

- Data governance and transparency, including issues of algorithmic opacity and explainability;
- Regulatory challenges, reflecting the absence of standardized frameworks;
- Human and professional adaptation, emphasizing skills transformation and trust.

The clustering confirms that ethical, regulatory, and human factors are deeply interdependent and must be addressed simultaneously.

Intelligent systems often operate through processes that are not easily interpretable, creating uncertainty regarding accountability and decision traceability. Laine et al. 2024, caution that when auditors rely on complex models without fully understanding their mechanisms, professional skepticism may be weakened. Ensuring algorithmic transparency thus becomes an ethical imperative and a regulatory necessity.

The second challenge pertains to the absence of mature regulatory frameworks capable of governing the integration of intelligent technologies into assurance. Current standards remain largely descriptive rather than prescriptive, leaving gaps in responsibility allocation when automated systems err or produce biased outputs (Laine et al., 2024). This regulatory delay risks undermining stakeholder trust, particularly in contexts where financial data integrity and independence of judgment are paramount.

Finally, the human dimension remains central to the debate. The transformation of auditing requires a new blend of analytical, ethical, and technological literacy. Continuous learning and re-skilling are no longer optional but essential for maintaining relevance in a digital assurance environment. As Davenport (2018) aptly observed, “technology will not replace auditors, but auditors who embrace intelligent technologies will replace those who do not.” This statement encapsulates the cultural and professional metamorphosis currently reshaping the field.

The conceptual visualization in Figure 2 illustrates how these ethical and organizational clusters are interlinked. Issues of transparency are inseparable from regulation, and both are deeply intertwined with human adaptation. Together, they delineate the boundaries within which innovation can enhance rather than compromise the credibility of the audit function.

### 2.3. Integrative Interpretation and Strategic Implications for the Transformation of Auditing

The convergence of insights from the two conceptual maps reveals the multidimensional character of the ongoing transformation of auditing. The movement toward continuous and intelligent assurance cannot be understood merely as a technological evolution; it represents a paradigm shift in how organizations construct, validate, and communicate trust. The bibliometric and contextual analyses jointly suggest that the integration of intelligent technologies redefines the audit function across three intersecting levels: the operational, the epistemological, and the ethical.

At the operational level, the findings highlight the restructuring of audit processes. Intelligent systems enable permanent monitoring of financial flows and transactions, turning auditing into a living process embedded within organizational information architectures. This reconfiguration drastically reduces the temporal gap between the occurrence of events and their verification, thereby improving risk detection and responsiveness. Yet, this acceleration of assurance cycles also generates new forms of dependency on data infrastructures, algorithms, and automated decision models. The profession must therefore reconcile technological speed with methodological rigor, ensuring that efficiency does not come at the expense of depth and reflection.

At the epistemological level, the results point to a shift in the way knowledge and judgment are produced within the audit process. Traditionally, auditors relied on sampling, experience, and professional skepticism as foundations of reliability. Continuous auditing, however, introduces a model in which knowledge is increasingly algorithmic, probabilistic, and predictive. This new mode of reasoning demands from auditors not only technical literacy but also a renewed critical capacity to interpret machine-generated insights and to question automated reasoning when necessary. In this sense, the auditor's role evolves from verifier of evidence to interpreter of computational intelligence, a transformation that fundamentally alters professional identity and practice.

At the ethical and societal level, the expansion of intelligent auditing raises important questions about trust, transparency, and accountability. The risk of over-reliance on opaque systems whose decision paths remain invisible could undermine the very legitimacy of the audit profession. Ensuring transparency and explainability becomes a prerequisite for maintaining the social contract between auditors, institutions, and the public. Moreover, as intelligent auditing technologies become more autonomous, regulatory frameworks must evolve to define clear boundaries of responsibility and to protect the independence of assurance activities.

The synthesis of these dimensions suggests that the transition to continuous auditing is not a linear progression but rather a complex negotiation between innovation and ethics, between automation and human discernment. The strategic implications are considerable. For professional bodies, there is an urgent need to integrate digital ethics and data analytics into education and certification systems. For regulators, the challenge is to establish governance standards that preserve both innovation and accountability. For researchers, the field remains open for empirical inquiry into how intelligent systems reshape the cognitive, organizational, and ethical dimensions of auditing practice.

Recent studies highlight the growing importance of AI in enhancing audit judgment, professional skepticism, and risk assessment capabilities (Puthukulam et al., 2021; Nogueira et al., 2024; Mishra et al., 2024).

Ultimately, the future of auditing depends on achieving a balanced coexistence between technological intelligence and human judgment. Intelligent systems can enhance transparency and efficiency, but only human reflection can confer meaning, responsibility, and legitimacy upon their outcomes. The transformation of auditing, therefore, must be guided not only by what technology can do but by what society expects it to safeguard: fairness, reliability, and public trust.

## 3. Limitations of the Study

Despite its contributions, this study presents several limitations. First, the bibliometric analysis is limited to Scopus and Web of Science databases, which may exclude relevant publications from other sources. Second, the selection of keywords may have influenced the scope of the dataset. Third, bibliometric methods focus on quantitative patterns and may not fully capture the qualitative depth of individual studies.

Finally, the rapid evolution of AI technologies means that some findings may quickly become outdated, highlighting the need for continuous research in this area.

## 4. Conclusion

The evolution of auditing toward intelligent and continuous assurance represents more than a technical modernization; it signifies a structural redefinition of how trust and accountability are established in the digital economy. This article has highlighted the central role of intelligent systems in transforming audit processes, revealing both the potential and the vulnerabilities that accompany this transition.

Through the combined use of bibliometric mapping and interpretative analysis, the research has shown that continuous auditing embodies a paradigm in which automation, ethical reasoning, and strategic decision-making converge. The findings confirm that intelligent technologies can substantially improve the quality, timeliness, and analytical depth of assurance activities. Yet, the same innovations introduce new layers of complexity related to transparency, governance, and professional adaptation.

What emerges from this analysis is that the future of auditing will depend not on technological sophistication alone, but on the capacity of institutions and professionals to redefine the epistemic and ethical boundaries of the field. The integration of intelligent systems requires new modes of collaboration between human judgment and computational inference, ensuring that audit conclusions remain interpretable, accountable, and socially meaningful.

This study demonstrates that the integration of AI into auditing is not merely a technological enhancement but a structural transformation of the assurance function. By combining bibliometric analysis with conceptual interpretation, it provides a comprehensive mapping of the field and identifies key thematic areas shaping its evolution.

The findings highlight both the opportunities and the critical challenges associated with AI-driven continuous auditing, emphasizing the need for methodological rigor, ethical governance, and professional adaptation. Future research should focus on empirical validation, regulatory development, and the interaction between human judgment and algorithmic decision-making.

## Bibliography


Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017) Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory,* 36(4), 1-27. https://doi.org/10.2308/ajpt-51684

Davenport, T. H. (2018*). The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. MIT Press. https://doi.org/10.7551/mitpress/11781.001.0001

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Kokina, J., & Davenport, T. H. (2017) The Emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting, 14*(1), 115-122. https://doi.org/10.2308/jeta-51730

Laine, J., Minkkinen, M. & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information & Management, 61*(5), 103969. https://doi.org/10.1016/j.im.2024.103969

Leocádio, D., Malheiro, L.& Reis, J. (2024). Artificial intelligence in auditing: A conceptual framework for auditing practices. *Administrative Sciences, 14*(10), 238. 10.3390/admsci14100238

Mishra, A. K., Anand, S., Debnath, N. C., Pokhariyal, P. & Patel, A. (Eds.). (2024). *Artificial intelligence for risk mitigation in the financial industry*. Scrivener Publishing LLC. https://doi.org/10.1002/9781394175574

Nogueira, J., Ribeiro, D., & Marques, R. P. (2024). Factors influencing statutory auditors' perception of the role of artificial intelligence in auditing. In Á. Rocha, H. Adeli, G. Dzemyda, F.Moreira, & A. Poniszewska-Marańda (Eds.), Good practices and new perspectives in information systems and technologies. WorldsCist 2024 (*Lecture Notes in Networks and Systems,Vol. 990*, pp.306–316) Springer. https://doi.org/10.1007/978-3-031-60328-0_31

Noordin, N. A., Hussainey, K. & Hayek, A. F. (2022). The use of artificial intelligence and audit quality: An analysis from the perspectives of external auditors in the UAE.

*Journal of Risk and Financial Management, 15*(8), 339. https://doi.org/10.3390/jrfm15080339

Omoteso, K. (2012). The application of artificial intelligence in auditing: Looking back to the future. *Expert Systems with Applications, 39*(9), 8490–8495. https://doi.org/10.1016/j.eswa.2012.01.098

Puthukulam, G., Ravikumar, A., Sharma, R. V. K.& Meesaala, K. M. (2021). Auditors' perception on the impact of artificial intelligence on professional skepticism and judgment in Oman. *Universal Journal of Accounting and Finance, 9*(5), 1184–1190. https://doi.org/10.13189/ujaf.2021.090527

Raschke, R., L. Saiewitz, A., Kachroo, P. & Lennard, J. B. (2018). AI-enhanced audit inquiry: Research Note. *Journal of Emerging Technologies in Accountings, 15* (2), 111–116. https://doi.org/10.2308/jeta-52310

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016) Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), 1-20. https://doi.org/10.2308/jeta-10511

Van Eck, N. J. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84,* 523–538. https://doi.org/10.1007/s11192-009-0146-3

# UNDERSTANDING THE CONTINUANCE USE OF GENERATIVE AI IN HIGHER EDUCATION: A CROSS-NATIONAL STUDY OF PORTUGAL AND INDIA

**Agostinho Sousa Pinto[1], António Abreu[1], Manuel Pérez Cota[2], Sindhu Singh[3] & Jerónimo Paiva[1]**

*[1]Instituto Superior de Contabilidade e Administração do Porto (ISCAP)*
*CEOS.PP, Portugal*
*Porto, Portugal*

*[2]Universidade de Vigo,*
*Vigo, Spain*

*[3]K.J. Somaiya Institute of Management Studies & Research,*
*India*

**ABSTRACT**

Generative Artificial Intelligence (GenAI) tools such as ChatGPT and Gemini have transformative and disruptive impacts across various sectors, and education is not an exception. GenAI tools are becoming increasingly popular among university students to provide a more personalized study experience. In this paper, through the lens of IS continuance theory, the authors propose a conceptual model consisting of perceived anthropomorphism (ANT), perceived intelligence (INT), perceived technological novelty (TN), perceived usefulness (PU), perceived ease of use (PEOU), confirmation (C) and satisfaction (SA) to explore continuous intention (CI) and actual use (U) to use GenAI in higher education. Data were collected via an anonymous online questionnaire (5-point Likert scale), pilot-tested and translated into English, Portuguese, and Spanish, with 611 valid responses from students in India (n=306) and Portugal (n=305). Using a hybrid SEM-PLS and Artificial Neural Network (ANN) approach, results show that Technological Novelty (TN) is the strongest predictor of continuance intention—twice as influential in India (75%) as in Portugal (37%). These findings have implications for the application of AI in the context of higher education.

***Keywords:*** *Generative AI (GenAI), Artificial Intelligence (AI), ChatGPT, Cross-Country Comparison, Higher Education, IS Continuance, Technology Acceptance, Artificial Neural Network (ANN), SEM-PLS*

## 1. Introduction

In the 2020s, the widespread adoption of Artificial Intelligence (AI) and Machine Learning (ML) technologies is disrupting a field that is naturally slow to embrace innovation: higher education. Indeed, the COVID-19 pandemic has significantly accelerated the digital transition of higher education institutions (HEIs), forcing them to carry out an unprecedented experiment in education (Paiva, 2022; Paiva et al., 2021; Pinto et al., 2022). Generative AI (GenAI), as exemplified by ChatGPT and Copilot, is a subset of deep learning that is capable of generating new content in the form of text, images, audio, and even video (Miquel-Vergés, 2022; Rocha, 2023).

Although AI chatbots have existed since the release of ELIZA, the release of a new generation of AI chatbots like OpenAI´s ChatGPT, Google´s Gemini, and X´s Grok have opened new opportunities in education (Lemos, 2018; Xavier et al., 2024). Nonetheless, several academics and educators have raised concerns about issues such as plagiarism, ethical concerns, bias, reduced cognitive ability, and a lack of educational guidelines (Galdames, 2025; Saúde et al., 2024).

Based on the IS continuance model as well as external variables that reflect the characteristics of AI, this study explores the continuance intention of university students to use GenAI. To explore the underlying mechanisms that lead students to use GenAI in their studies, an online questionnaire was conducted, and a conceptual model was proposed and tested. The research questions are presented below.

- RQ1: Are students willing to continue using GenAI in higher education?
- RQ2: What factors influence the continuance intention and actual use of GenAI by university students?

The research paper seeks to explore the continuance intention to use GenAI tools in higher education by university students from two contrasting countries: India and Portugal. To this end, an online questionnaire was conducted in two different higher education institutions: Somaiya Vidyavihar University and the Polytechnic Institute of Porto. By comparing responses from these two distinct institutions, the research intends to uncover regional similarities and differences in the adoption and usage of GenAI in higher education.

## 2. Literature Review

The rapid uptake of GenAI technologies like ChatGPT by universities fuelled broad examinations of adoption drivers and barriers. To aggregate these results, in this research, we conducted a series of Meta-Analytic Structural Equation Model (MASEM) studies on 61 papers (overall N = 21,499 participants across 30 nations). The results showed that Perceived Usefulness (PU) and Ease of Use (PEOU) remained the strongest predictors of adoption based on Technology Acceptance Model (TAM) (Davis, 1989; Davis, 2005; Tam & Oliveira, 2019). However, cross-cultural nuances emerged: Technological Novelty (TN) and anthropomorphism (ANT) showed higher salience in collectivist societies (*e.g.,* India, Pakistan) versus utilitarian expectations in Western contexts (Paiva et al., 2021; Singh & Paiva, 2025; Venkatesh et al., 2003; 2012).

Notably, 10 studies (16%) originated from the Indian subcontinent (India: 5; Pakistan: 2; Nepal: 2; Sri Lanka: 1), highlighting regional interest in GenAI's educational potential. These studies emphasized sociocultural factors, such as students' openness to human-like AI interactions (Mishra et al., 2023), whereas European and North American research prioritized ethical concerns (*e.g.,* plagiarism, bias). MASEM Confirmed Satisfaction (SA) as a universal mediator between confirmation and continuance intention, yet its effect size was 20% stronger in low-resource settings, suggesting GenAI's role in bridging educational gaps (Bhattacherjee, 2001).

Building on these results, this study proposes a model that builds on prior models by introducing: (1) intelligence and novelty as precedents to ease of use, (2) the mediating variable of satisfaction, and (3) continuance-to-usage relationships - without loss of theoretical parsimony, while addressing cultural differences (Table 1).

**Table 1**

*Key constructs and hypothesis*

| Constructs | Hypothesis |
|---|---|
| Perceived Intelligence (INT) reflects the degree to which users believe that a GenAI tool (*e.g.,* ChatGPT) understands their needs and provides intelligent answers.<br><br>Perceived Technological Novelty (TN) refers to the perception that GenAI tool is technologically innovative.<br><br>Perceived Anthropomorphism (ANT) relates to the extent to which users perceive that a GenAI tool has human-like traits sensibility.<br><br>Perceived Usefulness (PU) refers to the belief that GenAI tools will increase productivity and performance.<br><br>Perceived Ease of Use (PEOU) refers to the extent to which GenAI tools are easy to use and interact with.<br><br>Confirmation (C) is the degree to which users' expectations of the GenAI tools are met. Satisfaction (SA) is the level of contentment that users feel whilst experimenting with GenAI tools.<br><br>Continuance intention refers to a user's decision to continue using a technology, system, or service after its initial adoption.<br><br>Use refers to the actual behavior of utilizing a technology or system. | H1: Perceived Intelligence will have a positive effect on Perceived Ease of Use<br><br>H2: Perceived Technological Novelty will have a positive effect on Perceived Ease of Use.<br><br>H3: Perceived Anthropomorphism will have a positive effect on Perceived Ease of Use.<br><br>H4a: Perceived Ease of Use will have a positive impact on Perceived Usefulness.<br><br>H4b: Perceived Ease of Use will have a positive influence on Satisfaction.<br><br>H4c: Perceived Ease of Use will have a positive impact on Continuance Intention.<br>H5a: Perceived Usefulness will have a positive influence on Satisfaction.<br><br>H5b: Perceived Usefulness will have a positive impact on Continuance Intention.<br><br>H6a: Confirmation will have a positive impact on Perceived Usefulness.<br><br>H6b: Confirmation will have a positive impact on Satisfaction.<br><br>H7: Satisfaction will have a positive impact on Continuance Intention.<br><br>H8: Continuance Intention will have a positive impact on Use. |

## 3. Material and Methods

### 3.1. Hybrid PLS-SEM-ANN

This study employed a two-stage hybrid approach combining Partial Least Squares Structural Equation Modeling (PLS-SEM) and Artificial Neural Network (ANN) analysis (Ashaari et al., 2021; Liébana-Cabanillas et al., 2017). Partial Least Squares (PLS) regression, a type of Structural Equation Modelling (SEM) in Smart PLS, was conducted to confirm the reliability and validity of the constructs and model (Hair & Sarstedt, 2019). Artificial Neural Network (ANN) captured non-linear patterns among the independent variables (C, SA, TN, INT, PEOU, ANT, PU) and the dependent variable (CI) (Taud & Mas, 2018). ANN provides greater prediction precision, with its ability to handle non-normality and small samples, ranked the relative importance of predictors.

### 3.2. Data Collection Procedure

Data was collected via an online questionnaire using a 5-point Likert scale (Abreu et al., 2015). After pretesting and translation (English/Portuguese/Spanish), questionnaires were distributed to university students in India from the Somaiya Vidyavihar University (n=306 valid responses) and Portugal (n=305) from the Polytechnic Institute of Porto from December 2024- to February 2025. Both samples are sufficiently large to meet the required minimum sample size of 200 (Boomsma & Hoogland, 2001). In the Appendix 1, it is possible to find a Table that presents the individual outer loading (λ), the Mean (M), and Standard Deviation (SD) for all the questionnaire items (Hussain et al., 2018; Nunnally, 1978).

Table 2 presents the demographic characteristics of the questionnaire respondents for both samples:

- The first sample primarily consists of young adults, with 85% of respondents under 25. None of the Indian respondents is over 35 years old. There is a slight gender imbalance toward male respondents (56.21% male vs. 42.81% female). Additionally, most Indian respondents have higher education, either at the bachelor's or master's level.
- None of the Portuguese respondents are under 18 years old, and over two-thirds are between 18 and 24 years old. The second sample has a slight gender imbalance toward female respondents (43.61% male vs. 56.07% female). More than 40% of Portuguese respondents have a high school education, and one-third of them hold a BSc degree.

**Table 2**

*Demographic Characteristics of the Questionnaire Respondents*

| | | India | | Portugal | |
|---|---|---|---|---|---|
| | | **Number** | **Percentage** | **Number** | **Percentage** |
| **Gender** | **Male** | 172 | 56.21% | 133 | 43.61% |
| | **Female** | 131 | 42.81% | 171 | 56.07% |
| | **Prefer Not to Answer** | 3 | 0.98% | 1 | 0.33% |
| **Age** | **Under 18** | 22 | 7.19% | 0 | 0% |
| | **18-24** | 238 | 77.78% | 213 | 69.84% |
| | **25-34** | 46 | 15.03% | 44 | 14.43% |
| | **Over 35** | 0 | 0% | 48 | 15.74% |
| **Education** | **High School Diploma** | 25 | 8.07% | 142 | 46.56% |
| | **Bachelor's degree** | 191 | 61.61% | 103 | 33.77% |
| | **Master's degree** | 81 | 26.13% | 36 | 11.80% |
| | **Any Other** | 13 | 4.19% | 24 | 7.87% |
| | **Total** | **306** | **100%** | **305** | **100%** |

### 3.3. Validity and Reliability Analysis

Firstly, to determine whether to use parametric or non-parametric methods, it is important to establish whether all variables follow a normal distribution. To this end, the Kolmogorov-Smirnov (K-S) was used to test the assumption of normality for all variables, and they significantly deviate from a normal distribution ($p < 0.05$) (Oliveira et al., 2011). Common Method Bias (CMB) was assessed using a Common Latent Factor (CLF), which showed no signs of bias (Podsakoff et al., 2003). The standardized regression weights differed by less than 0.200 between models with and without the CLF.

Secondly, items that exhibited an outer loading below the 0.7 threshold were removed. U3, U6, and U7 were eliminated from the Indian sample, while ANT4, ANT5, PEOU3, and U3-U7 were removed from the Portuguese sample.

Thirdly, the reliability and validity of the survey were assessed using Cronbach's alpha($\rho\_T$) and the Composite Reliability (CR) using exploratory factor analysis (Netemeyer et al., 2003). In Tables 3 and 4, all the key constructs have a Cronbach's Alpha and CR above 0.7, which proves the construct reliability. It also shows that the AVE values of all constructs are larger than the recommended threshold of 0.5 (Fornell & Larcker, 1981). The discriminant validity of constructs is assessed with the Fornell-Larcker criterion, which postulates that the positive square root of the AVE for all factors is higher than the highest correlation with any other factor (Fornell & Larcker, 1981).

**Table 3**

*Validity and Reliability Analysis - India*

| | $\rho_T$ | **R** | **VE** | **NT** | | **I** | **NT** | **EOU** | **U** | **A** | **N** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NT** | .84 | .85 | .62 | .78 | | | | | | | | |
| | .84 | .84 | .68 | .59 | .82 | | | | | | | |
| **I** | .86 | .86 | .7 | .53 | .72 | .84 | | | | | | |
| **NT** | .87 | .87 | .65 | .72 | .61 | .52 | .81 | | | | | |
| **EOU** | .82 | .82 | .64 | .48 | .59 | .55 | .57 | .8 | | | | |
| **U** | .84 | .84 | .68 | .45 | .53 | .54 | .47 | .67 | .82 | | | |
| **A** | .87 | .87 | .73 | .6 | .77 | .7 | .66 | .57 | .54 | .85 | | |
| **N** | .8 | .82 | .62 | .61 | .66 | .71 | .62 | .53 | .52 | .7 | .79 | |
| | .76 | .76 | .58 | .37 | .54 | .5 | .39 | .57 | .62 | .47 | .44 | .76 |

**Table 4**

*Validity and Reliability Analysis - Portugal*

| | $\rho_T$ | R | VE | NT | | I | NT | EOU | U | A | N | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NT** | .66 | .81 | .59 | .77 | | | | | | | | |
| | .83 | .9 | .75 | .43 | .87 | | | | | | | |
| **I** | .89 | .92 | .75 | .38 | .44 | .86 | | | | | | |
| **NT** | .82 | .88 | .65 | .41 | .55 | .45 | .81 | | | | | |
| **EOU** | .79 | .88 | .7 | .3 | .46 | .24 | .52 | .84 | | | | |
| **U** | .86 | .9 | .7 | .4 | .54 | .67 | .5 | .36 | .84 | | | |
| **A** | .89 | .92 | .75 | .44 | .64 | .63 | .66 | .49 | .63 | .87 | | |
| **N** | .86 | .91 | .71 | .44 | .43 | .66 | .49 | .31 | .65 | .61 | .84 | |
| | .74 | .89 | .79 | .34 | .37 | .7 | .4 | .32 | .57 | .5 | .49 | .89 |

### 3.4. Validity and Reliability Analysis

Multicollinearity can distort the regression model, preventing us from distinguishing between the individual effects of independent variables on dependent variables because two or more independent variables are highly correlated. The Variance Inflation Factor (VIF) measures the extent to which the standard error increases due to collinearity. An Average Full Collinearity VIF (AFVIF) of less than 3.3 is ideal and indicates that there are no signs of multicollinearity (Kock, 2010; Shin, 2003). VIF values range from 1 to 2.336, suggesting that there is no multicollinearity among the predictor variables.

Table 5 presents the model fit indicators for both the saturated and estimated models to evaluate the model's fitness to the observed data: SRMR, d_ULS, d_G, Chi-square, and NFI (Hu & Bentler, 1998; Lohmöller, 1989). All the indicator values fall within the acceptable range.

**Table 5**

*Model Fit Indicators*

| Indicators | India | | Portugal | |
|---|---|---|---|---|
| | Saturated Model | Estimated Model | Saturated Model | Estimated Model |
| **SRMR** | 0.057 | 0.099 | 0.06 | 0.129 |
| **d_ULS** | 2.401 | 7.267 | 1.78 | 8.273 |
| **d_G** | 0.945 | 1.127 | 0.717 | 0.9 |
| **Chi-square** | 1660.398 | 1814.224 | 1343.701 | 1509.132 |
| **NFI** | 0.777 | 0.756 | 0.777 | 0.75 |

### 3.5. Analysis of Hypothesis Tests

Using the consistent PLS-SEM algorithm technique, the Path Coefficients (β) and Cohen's F Statistic ($F^2$) were obtained for each hypothesis, as shown in Table 6 and 7 (Esposito Vinzi et al., 2010). For the Indian sample, the results indicate that 11 out of 13 hypotheses cannot be rejected. H3: ANT → PEOU and H5b: PU → CI were the only hypotheses that were rejected.

**Table 6**

*Results of the SEM-PLS model - India*

| Hypothesis | Coefficient | F Statistic | *p* |
|---|---|---|---|
| **H1: TN -> PEOU** | 0.261 | 0.061 | 0.000*** |
| **H2: INT -> PEOU** | 0.377 | 0.098 | 0.001** |
| **H3: ANT -> PEOU** | 0.049 | 0.002 | 0.541 |
| **H4a: PEOU -> PU** | 0.553 | 0.379 | 0.000*** |
| **H4b: PEOU -> SA** | 0.107 | 0.014 | 0.002** |
| **H4c: PEOU -> CI** | 0.147 | 0.023 | 0.001** |
| **H5a: PU -> SA** | 0.134 | 0.025 | 0.001** |
| **H5b: PU -> CI** | 0.15 | 0.025 | 0.185 |
| **H6a: C -> PU** | 0.199 | 0.049 | 0.004** |
| **H6b: C -> SA** | 0.635 | 0.659 | 0.000*** |
| **H7: SA -> CI** | 0.535 | 0.389 | 0.000*** |
| **H8: CI -> U** | 0.5 | 0.334 | 0.000*** |

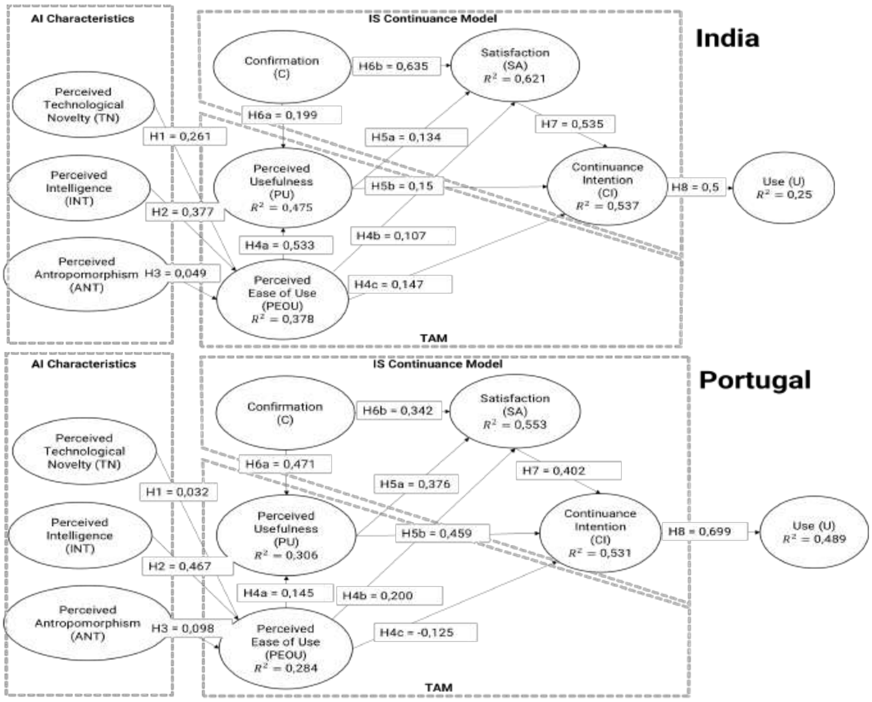
**Table 7**

*Results of the SEM-PLS model – Portugal*

| Hypothesis | Coefficient | F Statistic | P |
|---|---|---|---|
| **H1: TN -> PEOU** | 0.032 | 0.061 | 0.000*** |
| **H2: INT -> PEOU** | 0.467 | 0.098 | 0.001** |
| **H3: ANT -> PEOU** | 0.098 | 0.002 | 0.541 |
| **H4a: PEOU -> PU** | 0.145 | 0.379 | 0.000*** |
| **H4b: PEOU -> SA** | 0.2 | 0.014 | 0.002** |
| **H4c: PEOU -> CI** | -0.125 | 0.023 | 0.001** |
| **H5a: PU -> SA** | 0.376 | 0.025 | 0.001** |
| **H5b: PU -> CI** | 0.459 | 0.025 | 0.185 |
| **H6a: C -> PU** | 0.471 | 0.049 | 0.004** |
| **H6b: C -> SA** | 0.342 | 0.659 | 0.000*** |
| **H7: SA -> CI** | 0.402 | 0.389 | 0.000*** |
| **H8: CI -> U** | 0.699 | 0.334 | 0.000*** |

Figure 1 presents the structural model generated in SmartPLS for both samples.

**Figure 1**
*Model estimation results*




### 3.6. Artificial Neural Network (ANN)

Table 8 presents the sample data divided into training and testing sets, with 90% allocated for training and 10% for testing.

**Table 8**

*Training and testing data for ANN*

| Network | India | | | Portugal | | |
|---|---|---|---|---|---|---|
| | Sample Size (Training) | Sample Size (Testing) | Total Sample Size | Sample Size (Training) | Sample Size (Testing) | Total Sample Size |
| **1** | 271 | 35 | 306 | 277 | 28 | 305 |
| **2** | 278 | 28 | 306 | 271 | 34 | 305 |
| **3** | 279 | 27 | 306 | 265 | 40 | 305 |
| **4** | 272 | 34 | 306 | 275 | 30 | 305 |
| **5** | 266 | 40 | 306 | 266 | 39 | 305 |
| **6** | 276 | 30 | 306 | 274 | 31 | 305 |
| **7** | 267 | 39 | 306 | 271 | 34 | 305 |
| **8** | 275 | 31 | 306 | 277 | 28 | 305 |
| **9** | 272 | 34 | 306 | 274 | 31 | 305 |
| **10** | 278 | 28 | 306 | 269 | 36 | 305 |

To assess the possibility of overfitting, we carried out a ten-fold cross-validation procedure and obtained the Root Mean Square Errors (RMSE). It is important to note that the dependent variable (continuance intention) was standardized to z-scores (mean=0, SD=1) prior to analysis. Consequently, all RMSE values are reported on this standardized scale. Table 9 shows the average Sum of Square Errors (SSE) and RMSE for the training and testing procedures of the Indian sample. The testing errors demonstrate greater variability, with RMSE ranging between 0.298 and 0.488. In contrast, the training errors are consistent across the networks, with RMSE ranging between 2.090 and 2.463. The RMSE difference (Training - Testing) indicates the level of generalization capability. The low value (1.891) suggests that the model may generalize well. The RMSE difference of 1.611 also suggests it may generalize well.

**Table 9**

*Model fitness for RMSE – India*

| Network | Sum of Square Error | | RMSE | | |
|---|---|---|---|---|---|
| | Training | Testing | Training | Testing | Training-Testing |
| **1** | 51.146 | 6.389 | 2.302 | 0.427 | 1.875 |
| **2** | 53.324 | 2.486 | 2.283 | 0.298 | 1.985 |
| **3** | 55.549 | 4.62 | 2.241 | 0.414 | 1.827 |
| **4** | 56.309 | 6.241 | 2.198 | 0.428 | 1.769 |
| **5** | 60.87 | 9.177 | 2.09 | 0.479 | 1.611 |
| **6** | 49.235 | 5.863 | 2.368 | 0.442 | 1.926 |
| **7** | 47.193 | 5.662 | 2.379 | 0.381 | 1.998 |
| **8** | 45.344 | 7.376 | 2.463 | 0.488 | 1.975 |
| **9** | 46.555 | 5.821 | 2.417 | 0.414 | 2.003 |
| **10** | 49.669 | 4.961 | 2.366 | 0.421 | 1.945 |
| **Means** | 51.519 | 5.86 | 2.311 | 0.419 | 1.891 |
| **S.D.** | 4.692 | 1.658 | 0.106 | 0.05 | |

Table 10 illustrates the average Sum of Square Errors (SSE) and Root Mean Square Errors (RMSE) for the training and testing procedures of the Portuguese sample. Additionally, the RMSE values for both training and testing are relatively low, indicating a high level of generalization capability. As a baseline for comparison, a null model (predicting the mean of CI for all cases) would yield an RMSE equivalent to the standard deviation of CI. Since CI was standardized to a standard deviation of 1.0, our testing RMSE values of 0.419 (India) and 0.344 (Portugal) represent substantial improvements over this baseline—reducing prediction error by approximately 58% and 66%, respectively.

**Table 10**

*Model fitness for RMSE – Portugal*

| | Sum of Square Error | | RMSE | | |
|---|---|---|---|---|---|
| | **Training** | **Testing** | **Training** | **Testing** | **Training-Testing** |
| **1** | 42.517 | 3.535 | 2.552 | 0.355 | 2.197 |
| **2** | 40.161 | 7.196 | 2.598 | 0.46 | 2.138 |
| **3** | 40.197 | 4.752 | 2.568 | 0.345 | 2.223 |
| **4** | 46.511 | 4.21 | 2.432 | 0.375 | 2.057 |
| **5** | 43.985 | 3.677 | 2.459 | 0.307 | 2.152 |
| **6** | 46.61 | 3.149 | 2.425 | 0.319 | 2.106 |
| **7** | 43.864 | 2.22 | 2.486 | 0.256 | 2.23 |
| **8** | 55.075 | 1.261 | 2.243 | 0.212 | 2.03 |
| **9** | 47.865 | 5.988 | 2.393 | 0.44 | 1.953 |
| **10** | 42.377 | 4.996 | 2.519 | 0.373 | 2.147 |
| **Means** | 44.916 | 4.098 | 2.467 | 0.344 | 2.123 |
| **S.D.** | 4.2 | 1.655 | 0.098 | 0.072 | |

To identify the relative contribution of each predictor to continuance intention, we performed a sensitivity analysis using the connection weight approach. This method computes variable importance by summing the products of input-hidden and hidden-output connection weights for each input variable. Regarding the Indian sample, Table 11 indicates that Technical Novelty (75%) is by far the most significant predictor of Continuance Intention (CI), followed by Perceived Usefulness (PU) (32%), and Confirmation (18%).

**Table 11**

*Sensitivity Analysis – Indian Sample*

| | TN | INT | ANT | C | SA | PU | PEOU |
|---|---|---|---|---|---|---|---|
| **1** | 71% | 36% | 7% | 40% | 56% | 8% | 14% |
| **2** | 70% | 7% | 11% | 89% | 100% | 70% | 18% |
| **3** | 66% | 14% | 13% | 66% | 79% | 31% | 15% |
| **4** | 93% | 37% | 22% | 100% | 39% | 34% | 10% |
| **5** | 100% | 30% | 68% | 89% | 49% | 29% | 19% |
| **6** | 100% | 22% | 14% | 67% | 72% | 31% | 10% |
| **7** | 48% | 4% | 5% | 41% | 54% | 15% | 22% |
| **8** | 76% | 21% | 22% | 69% | 82% | 30% | 18% |
| **9** | 52% | 39% | 22% | 56% | 50% | 11% | 24% |
| **10** | 72% | 27% | 5% | 77% | 52% | 5% | 24% |
| **Average Normalized Importance** | **75%** | **24%** | **19%** | **69%** | **63%** | **26%** | **17%** |

Regarding the Portuguese sample, Table 12 indicates that Technical Novelty (37%) is the most important predictor of Continuance Intention (CI), followed by Perceived Usefulness (32%) and Confirmation (18%). Consequently, ranking the most critical factors remains consistent across both samples.

**Table 12**

*Sensitivity Analysis – Portuguese Sample*

| | TN | INT | ANT | C | SA | PU | PEOU |
|---|---|---|---|---|---|---|---|
| **1** | 30% | 5% | 9% | 15% | 15% | 28% | 9% |
| **2** | 39% | 10% | 3% | 18% | 12% | 34% | 5% |
| **3** | 37% | 19% | 18% | 10% | 5% | 25% | 4% |
| **4** | 32% | 9% | 4% | 11% | 6% | 29% | 13% |
| **5** | 19% | 14% | 6% | 11% | 4% | 20% | 4% |
| **6** | 64% | 8% | 4% | 17% | 31% | 29% | 10% |
| **7** | 35% | 5% | 12% | 12% | 3% | 33% | 1% |
| **8** | 50% | 10% | 15% | 55% | 25% | 54% | 42% |
| **9** | 40% | 4% | 4% | 21% | 35% | 46% | 10% |
| **10** | 21% | 21% | 6% | 11% | 6% | 20% | 4% |
| **Average Normalized Importance** | **37%** | **11%** | **8%** | **18%** | **14%** | **32%** | **10%** |

## 4. Discussion

This study reveals meaningful cross-cultural patterns in how Indian and Portuguese university students perceive GenAI in academic contexts. While both populations recognize the learning potential of GenAI, notable differences emerge in their orientations toward anthropomorphism, technological novelty, and satisfaction.

Indian students expressed greater willingness to ascribe human-like qualities to GenAI, with 60% to 70% agreeing that it is lifelike, compared to 20% to 30% of Portuguese students, who were more skeptical. This pattern suggests that cultural factors may be associated with the degree to which individuals anthropomorphize AI. Indian respondents' more evenly distributed, mesokurtic response patterns point toward cautious but balanced judgment, while Portuguese students exhibited a leptokurtic, positively skewed pattern—indicating stronger enthusiasm for GenAI's technical novelty than for its human-like qualities. This finding aligns with Portugal's growing emphasis on digital literacy and technological innovation, which may foster greater appreciation of AI for its functional benefits rather than perceived anthropomorphic traits.

Despite these differences, both groups reported high satisfaction with GenAI, with Portuguese students slightly more satisfied (93%) than their Indian counterparts (91%).

Portuguese students were also more likely to endorse GenAI's utility in education (76.4% agreement vs. 59.4% in India), suggesting that satisfaction may be more closely associated with perceived functionality than with experiential factors in the Portuguese context. This interpretation is reinforced by the finding that Perceived Usefulness (PU) showed a weaker relationship with continuance intention in the Indian sample, indicating that non-utilitarian factors—such as novelty or interest—may be stronger correlates of attitudes toward GenAI among Indian students.

The application of Expectation Confirmation Theory (ECT) and the Technology Acceptance Model (TAM) provided additional interpretive lenses. Perceived intelligence and technological novelty both showed positive associations with Perceived Ease of Use (PEOU) for both groups, consistent with H1 and H2. Perceived anthropomorphism was not significantly associated with PEOU (H3 not supported), a finding that may reflect the uncanny valley effect, wherein excessive human-likeness can detract from user comfort. Notably, while PEOU was positively associated with Perceived Usefulness (PU) in both groups (H4a, H4b supported), PU showed no significant relationship with continuance intention (H5b not supported). This suggests that sustained adoption may be more closely tied to experiential considerations than to perceived utility alone.

Confirmation was positively associated with satisfaction (H6b supported) but not with PU (H6a not supported), indicating that satisfaction may stem from fulfilled expectations independent of usefulness perceptions. Satisfaction, in turn, was positively associated with continuance intention (H7 supported). Taken together, the findings suggest that while Indian and Portuguese students differ in how they relate to GenAI, both groups appear more likely to continue using it when they hold positive attitudes—regardless of their satisfaction levels.

These findings offer practical considerations for integrating AI in education. Developers and educators may benefit from accounting for cultural differences when designing and introducing GenAI tools. For instance, in contexts where anthropomorphism resonates more strongly—such as India—designing for relatable engagement may enhance appeal. In technology-oriented settings like Portugal, emphasizing capability and innovation may prove more effective. Future research could examine additional cultural and contextual factors associated with AI perceptions, as well as track how these orientations evolve with sustained use over time.

## 5. Conclusions

This study provides critical insights into GenAI adoption and impact in higher education across Indian and Portuguese student populations. The research confirms that GenAI holds transformative potential, enhancing tailored learning, faculty effectiveness, and digital competence through applying sound psychometric validation, hybrid SEM-ANN analysis, and descriptive statistical analysis. However, its successful deployment requires adaptive policy frameworks that balance innovation with ethical safeguards.

## 6. Contributions and Limitations

The current study has certain limitations related to both samples. The questionnaire was only distributed in one university in India (Somaiya Vidyavihar University) and in Portugal (Polytechnic Institute of Porto). Naturally, both samples can contain biases. Future research should gather data from more universities and other countries to enhance the generalizability of the findings. Moreover, to examine the factors explaining the continuance intention of GenAI tools, future studies should investigate other comparable models to investigate additional factors to better understand the acceptance of AI technologies. It would also benefit from incorporating classroom-based observations or ethnographic methods to capture the real-time dynamics of GenAI interaction among students and educators.

# References


Abreu, A., Rocha, A., Pérez Cota, M., & Carvalho, J. V. (2015). Caderneta eletrónica no processo ensino-aprendizagem: Visão de professores e pais de alunos do ensino básico e secundário. *Revista Ibérica de Sistemas e Tecnologias de Informação, (16),* 108–128. https://doi.org/10.17013/risti.16.108-128

Ashaari, M. A., Singh, K. S. D., Abbasi, G. A., Amran, A., & Liebana-Cabanillas, F. J. (2021). Big data analytics capability for improved performance of higher education institutions in the industry 4.0 era: A multi-analytical SEM & ANN perspective. *Technological Forecasting and Social Change, 173*, 121119. https://doi.org/10.1016/j.techfore.2021.121119

Bhattacherjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, *25(3)*, 351–370. https://doi.org/10.2307/3250921

Boomsma, A., & Hoogland, J. J. (2001). *The robustness of LISREL modeling revisted.* https://api.semanticscholar.org/CorpusID:17339260

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Davis, T.R.V. (2005). Integrating Shared Services with the Strategy and Operations of MNEs. *Journal of General Management*, *31(2)*, 1–17. https://doi.org/10.1177/030630700503100201

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50. https://doi.org/10.2307/3151312

Galdames, I. S. (2025). Impact of artificial intelligence on higher education: A literature review. In A. Abreu, J. V. Carvalho, A. Mesquita, A. Sousa Pinto, M. Mendonça Teixeira (Eds.), *Perspectives and Trends in Education and Technology (*ICITED 2024, Lecture Notes in Networks and Systems, Vol. 859, pp. 373–392). Springer. https://doi.org/10.1007/978-3-031-78155-1_35

Hair, J. F., & Sarstedt, M. (2019). Factors versus Composites: Guidelines for Choosing the Right Structural Equation Modeling Method. *Project Management Journal*, *50*(6), 619–624. https://doi.org/10.1177/8756972819882132

Hu, L. -t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hussain, S., Fangwei, Z., Siddiqi, A. F., Ali, Z., & Shabbir, M. S. (2018). Structural Equation Model for Evaluating Factors Affecting Quality of Social Infrastructure Projects. *Sustainability*, *10*(5). https://doi.org/10.3390/su10051415

Kock, N. (2010). Using WarpPLS in E-collaboration Studies: An Overview of Five Main Analysis Steps. *International Journal of e-Collaboration*, *6*(4), 1–11. https://doi.org/10.4018/jec.2010100101

Liébana-Cabanillas, F., Marinković, V., & Kalinić, Z. (2017). A SEM-neural network approach for predicting antecedents of m-commerce acceptance. *International Journal of Information Management*, *37*(2), 14-24. https://doi.org/10.1016/j.ijinfomgt.2016.10.008

Lemos, E. das C. (2018). Technologies in the Classroom: Reflections on Teaching, Research, and Outreach in Public Schools of Rio Grande do Norte, Brazil. *International Journal of Research in Engineering and Technology*. https://api.semanticscholar.org/CorpusID:240401027

Lohmöller, J. B. (1989). Basic principles of model building. In J.-B. Lohmöller (Ed.), *Latent variable path modeling with partial least squares* (pp. 13–26). Physica-Verlag. https://doi.org/10.1007/978-3-642-52512-4_1

Miquel-Vergés, J. (2022). New technologies in virtual higher education: The deepfake and the face-to-face translation technologies. *HUMAN REVIEW. International Humanities Review / Revista Internacional de Humanidades, 12*(4), 1–20. https://doi.org/10.37819/revhuman.v12i4.1060

Mishra, A., Shukla, A., Rana, N. P., Currie, W. L., & Dwivedi, Y. K. (2023). Re-examining post-acceptance model of information systems continuance: A revised theoretical model using MASEM approach. *International Journal of Information Management*, *68*, 102571. https://doi.org/10.1016/j.ijinfomgt.2022.102571

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling Procedures: Issues and Applications.* SAGE Publications. https://books.google.pt/books?id=woiECgAAQBAJ

Nunnally, J. C. (1978). *Psychometric Theory*. (2nd ed.) McGraw-Hill. https://books.google.pt/books?id=WE59AAAAMAAJ

Oliveira, M. M. de, Santos, L. D., & Fortuna, N. (2011). Econometria. *Escolar Editora*.

Paiva, J. (2022). *Análise do ensino a distância no ISCAP-IPP em tempos de pandemia da COVID-19: Uma nova realidade educativa* [master's thesis, Instituto Superior de Contabilidade e Administração do Porto]. Repositório Científico do Instituto Politécnico do Porto.

Paiva, J., Abreu, A., & Costa, E. (2021). Distance learning in higher education during the COVID-19 pandemic: A systematic literature review. *Cadernos de Investigação do Mestrado em Negócio Eletrónico, 1*(1). https://doi.org/10.56002/ceos.0024_cimne_1_1

Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2022). Factors influencing the success of Moodle among students at a Portuguese business school. In J. V. d. Carvalho, P.

Liberato, & A. Peña (Eds.), *Advances in tourism, technology and systems (Smart Innovation, Systems and Technologies, Vol. 284,* pp. 479–489). Springer. https://doi.org/10.1007/978-981-16-9701-2_39

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Rocha, N. (2023). Oportunidades e Riscos do ChatGPT-um aliado ou inimigo? *The Trends Hub, 3*(1). https://doi.org/10.34630/tth.vi3.5039

Saúde, S., Barros, J. P., & Almeida, I. (2024). Impacts of generative artificial intelligence in higher education: research trends and students' perceptions. *Social Sciences*, *13*(8), 8. https://doi.org/10.3390/socsci13080410

Shin, B. (2003). An exploratory investigation of system success factors in data warehousing. *Journal of the Association for Information Sciences, 4*(1),*141-170.* https://doi.org/10.17705/1jais.00033

Singh, S., & Paiva, J. (2025). The role of AI characteristics and their influence on higher education students' continuance intention to use GenAI tools. *Information Discovery and Delivery*. Advance online publication. https://doi.org/10.1108/IDD-03-2025-0060

Tam, C., & Oliveira, T. (2019). Does culture influence m-banking use and individual performance? *Information & Management, 56*(3), 356–363. https://doi.org/10.1016/j.im.2018.07.009

Taud, H., & Mas, J. F. (2018). Multilayer perceptron (MLP). In M. Camacho Olmedo, J. M., Paegelow, J. F, & F. Escobar (Eds.), *Geomatic Approaches for Modeling Land Change Scenarios* (Lecture Notes in Geoinformation and Cartography, pp. 451–455). Springer. https://doi.org/10.1007/978-3-319-60801-3_27

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.). (2010). *Handbook of partial least squares: Concepts, methods and applications*. Springer. https://doi.org/10.1007/978-3-540-32827-8https://books.google.co.in/books?id=PPUbvBUvmWoC&printsec=frontcover&source=gbs_atb#v=onepage&q&f=false

Xavier, C. C. P., Gonçalves, V. M., Filho, A. S. de A., Lemos, E. das C., & Maia, L. C. G. (2024). Inteligência artificial aplicada à educação: Contribuição nas avaliações formativas de aprendizagem. *Revista Processando o Saber, 16*(01), 44–56. https://doi.org/10.5281/zenodo.14225907

**Appendix 1**

| Question | India | | | Portugal | | |
|---|---|---|---|---|---|---|
| | | | D | | | D |
| INT1 - GenAI tools are competent in understanding my educational needs. | .76 | .24 | .91 | .78 | .99 | .87 |
| INT2 - GenAI tools are knowledgeable. | .84 | .40 | .91 | .84 | .66 | .01 |
| INT3 - GenAI tools exhibit responsibility during interactions. | .82 | .38 | .97 | .64 | .74 | .89 |
| INT4 - Generative AI tools have intelligent functions to address complex educational tasks. | .78 | .32 | .95 | .74 | .24 | .91 |
| INT5 - GenAI tools give correct answers to my questions. | .82 | .50 | .92 | .83 | .40 | .91 |
| ANT1 – GenAI tools are sensible during replies. | .80 | .07 | .07 | .80 | .50 | .92 |
| ANT2 – Using GenAI feels natural; I do not perceive it as fake. | .81 | .91 | .06 | .74 | .07 | .07 |
| ANT3 – Using GenAI appears more human-like. | .79 | .03 | .99 | .79 | .91 | .06 |
| ANT4 – GenAI is conscious of its actions. | .79 | .13 | .01 | .65 | .03 | .99 |
| ANT5 – GenAI feels lifelike and not artificial. | .74 | .43 | .84 | .66 | .13 | .01 |
| TN1 GenAI tools are an innovative technology in higher education. | .84 | .69 | .89 | .84 | .75 | .96 |
| TN2 GenAI offers unique features not found in other educational tools. | .79 | .59 | .89 | .79 | .66 | .94 |
| TN3 The technology behind GenAI in education is novel to me. | .72 | .41 | .90 | .88 | .90 | .90 |
| TN4 GenAI is at the forefront of educational technology. | .79 | .49 | .93 | .86 | .75 | .93 |
| PU1 - GenAI improves my academic performance. | .80 | .46 | .00 | .83 | .41 | .06 |

| Question | India | | | Portugal | | |
|---|---|---|---|---|---|---|
| | | | D | | | D |
| PU2 - GenAI enhances my productivity in coursework. | .86 | .75 | .96 | .87 | .60 | .00 |
| PU3 - GenAI is useful for my educational activities. | .83 | .66 | .94 | .84 | .42 | .14 |
| PU4 - Using GenAI increases my efficiency in completing academic tasks. | .80 | .90 | .90 | .80 | .46 | .00 |
| PEOU1 - I find it easy to get GenAI tools to do what I want it to do. | .82 | .75 | .93 | .85 | .75 | .96 |
| PEOU2 - I find the GenAI tools to be easy to use. | .83 | .99 | .87 | .82 | .23 | .06 |
| PEOU3 - Interaction with the GenAI tools does not require a lot of my mental effort. | .75 | .66 | .01 | .61 | .57 | .06 |
| PEOU4 - My interaction with the GenAI tools is clear and understandable. | .81 | .74 | .89 | .82 | .72 | .95 |
| C1 - My experience with GenAI in education was better than what I expected. | .84 | .45 | .92 | .81 | .43 | .84 |
| C2 - The benefits I receive from using GenAI align with my expectations. | .82 | .61 | .93 | .68 | .69 | .89 |
| C3 - Using GenAI has confirmed my initial expectations about its utility. | .86 | .61 | .88 | .88 | .59 | .89 |
| C4 - The performance of GenAI in my studies matches my initial expectations. | .83 | .53 | .82 | .83 | .41 | .90 |
| SA1 - I am satisfied with the performance of GenAI in my education. | .86 | .59 | .88 | .88 | .49 | .93 |
| SA2 - The GenAI meets my academic needs. | .86 | .56 | .86 | .85 | .64 | .85 |
| SA3 - I am pleased with my experience using GenAI in my studies. | .85 | .64 | .79 | .86 | .60 | .86 |
| SA4 - Overall, I am happy with the effectiveness of GenAI in education. | .84 | .63 | .88 | .88 | .63 | .84 |

| Question | India | | | Portugal | | |
|---|---|---|---|---|---|---|
| | | | D | | | D |
| CI1-I intend to continue using GenAI tools in my future educational activities. | .84 | .75 | .91 | .88 | .61 | .88 |
| CI2- I plan to increase my use of GenAI tools over time in my coursework. | .82 | .51 | .93 | .83 | .53 | .82 |
| CI3 -I will use GenAI tools regularly in my studies. | .86 | .48 | .98 | .87 | .59 | .88 |
| CI4 - I am likely to recommend the use of GenAI tools to others in higher education. | .83 | .76 | .97 | .89 | .56 | .86 |
| U1 - I use GenAI frequently | .77 | .75 | .96 | .76 | .64 | .79 |
| U2 - I utilize GenAI for completing my assignments. | .78 | .23 | .06 | .77 | .63 | .88 |
| U3 - I rely on GenAI for studying and learning new concepts. | .20 | .57 | .06 | .66 | .75 | .91 |
| U4 - I use GenAI to assist me in research projects. | .76 | .72 | .95 | .63 | .51 | .93 |
| U5 - I use GenAI tools to review and revise my coursework. | .70 | .41 | .06 | .66 | .48 | .98 |
| U6 - I interact with GenAI for exam preparation and practice questions. | .20 | .60 | .00 | .58 | .76 | .97 |
| U7 - I use GenAI to enhance my presentations and class projects. | .16 | .42 | .14 | .67 | .62 | .79 |

# USING AI TO MAKE SMARTER PRICING DECISIONS IN THE HOTEL INDUSTRY

**Majlinda Godolja[1] & Arjola Sinani[2]**

*[1]Faculty of Economy,*
*University of Tirana, Tirana, Albania*

*[2]Universiteti i Vlorës Ismail Qemali*
*Albania*

**ABSTRACT**

This study examines how artificial intelligence (AI) can support revenue management in data-constrained hotel environments by combining demand forecasting with constrained price optimization. Using reservation data from a 40-room independent hotel in Vlora, Albania, the study develops a hybrid analytical framework that integrates Prophet and XGBoost forecasting models with a mixed-integer linear programming (MILP) optimization module. Forecast accuracy was evaluated through rolling 30-day prediction windows across six room categories. The results indicate that forecasting performance varied by category, with XGBoost outperforming Prophet in most cases and achieving statistically significant improvements for Double with Sea View, Quadruple, and Triple rooms, while Prophet remained a useful interpretable benchmark. These forecasts were then used to generate daily pricing recommendations under capacity and pricing constraints. Supported by a prototype dashboard, the framework demonstrates how a transparent AI-assisted decision-support approach can provide practical pricing guidance for small hotels in data-scarce settings.

***Keywords***: *artificial intelligence; revenue management; Prophet; XGBoost; forecasting; optimization; hospitality*

## 1. Introduction

Artificial Intelligence (AI) is increasingly shaping hospitality management by supporting demand forecasting, price adjustment, and operational decision-making. Large hotel groups already use data-rich revenue management systems to analyse booking patterns and support dynamic pricing, but comparable adoption remains more difficult for small and medium-sized hotels, especially in emerging markets. In such settings, limited historical data, strong seasonality, and weaker analytical infrastructure constrain the use of advanced pricing models (Alrawadieh et al., 2021; Buhalis & Leung, 2018; Henriques & Nobre Pereira, 2024).

These challenges are particularly relevant in small-hotel contexts, where pricing decisions are often still based on managerial intuition or static rate calendars. Although research on AI-enabled forecasting and pricing has expanded, much of the literature remains focused on larger firms or richer data environments. Smaller independent hotels are comparatively underrepresented, despite their importance in many tourism economies (Henriques & Nobre Pereira, 2024; Pereira & Cerqueira, 2022). This gap is both empirical and practical: methods that perform well in high-data settings are not always suitable for properties that require transparent, computationally manageable, and managerially usable tools. In such contexts, interpretability matters because adoption depends not only on predictive accuracy, but also on whether model outputs can be understood and trusted by non-technical users (Nowak & Pawłowska-Nowak, 2024).

In response to this gap, the present study develops a hybrid forecasting-optimization framework for hotel pricing that combines Prophet and Extreme Gradient Boosting (XGBoost) with a mixed-integer linear programming (MILP) model. Prophet is included as an interpretable benchmark for modelling recurring temporal structure, whereas XGBoost is used to capture nonlinear and short-term demand variation (Chen & Guestrin, 2016; Taylor & Letham, 2018). Rather than assuming that one forecasting method is universally superior, the study evaluates how alternative forecasting structures perform across room categories and how these differences affect downstream pricing recommendations.

## 2. Literature Review

Revenue management in hospitality has evolved from rule-based, descriptive approaches to more data-driven systems that support demand forecasting, pricing adjustment, and inventory control. Within this transition, artificial intelligence has expanded the role of revenue management from retrospective description to predictive and decision-support applications (Buhalis & Leung, 2018; Ivanov & Webster, 2024). However, adoption remains uneven across the sector. Large hotel groups are more likely to benefit from integrated analytics infrastructure, whereas small and medium-sized hotels often face constraints related to data availability, implementation cost, technical expertise, and organisational readiness (Alrawadieh et al., 2021; Pereira & Cerqueira, 2022). In such settings, the practical value of AI depends not only on predictive accuracy but also on interpretability and ease of managerial use (Nowak & Pawłowska-Nowak, 2024).

Demand forecasting is central to hotel revenue management because pricing and inventory decisions depend on expectations about future bookings and room-night demand. Classical models such as ARIMA, exponential smoothing, and regression remain useful because they are transparent, but their performance may weaken when demand is shaped by irregular seasonality, abrupt shocks, or multiple interacting drivers (Henriques & Nobre Pereira, 2024; Huang & Zheng, 2023). Machine-learning approaches offer greater flexibility in such contexts. XGBoost has attracted attention because it can model nonlinear relationships and accommodate a broad set of predictors, while Prophet provides an interpretable framework for modelling trend and seasonal structure with accessible parameterisation (Chen & Guestrin, 2016; Taylor & Letham, 2018). Rather than functioning only as substitutes, these approaches may offer complementary strengths, which has encouraged comparative and hybrid forecasting designs in hospitality analytics (Ampountolas & Legg, 2025; Henriques & Nobre Pereira, 2024).

Forecasting alone does not improve revenue unless predictions are translated into pricing decisions. For this reason, the link between forecasting and optimization is central to revenue management. In hotel settings, optimization models are used to convert expected demand into room-rate recommendations under capacity, pricing, and operational constraints, with mixed-integer linear programming offering a structured and transparent way to formalise such decisions across room categories (Mariani & Baggio, 2022; Talluri & Van Ryzin, 2005). Yet, much of this literature assumes richer data and stronger analytical capacity than are typically available to smaller properties. This creates a gap between the sophistication of many proposed models and the operational realities of independent hotels. A second gap is that forecasting accuracy is often evaluated in isolation, without examining how differences in predictive performance affect downstream pricing recommendations. The present study is positioned within these gaps. It examines a small independent hotel and develops an integrated framework linking demand forecasting, constrained price optimization, and dashboard-based decision support, with an emphasis on operationally feasible and interpretable pricing support in a data-constrained environment.

## 3. Material And Methods

### 3.1. Case Context

The empirical analysis was conducted on a 40-room independent hotel in Vlora, Albania, a coastal destination characterized by strong seasonality and increasing tourism activity. The case is representative of a small-hotel setting in which pricing decisions must be made under limited analytical capacity, a relatively short historical record, and substantial reliance on digital distribution channels.

The analysis used reservation-level data exported from the hotel's booking records. After the data preparation, the final analytical sample contained 2,255 reservations covering the period from 1 April 2024 to 14 February 2025. The study focused on six room categories: Double, Double with Sea View, Deluxe Double, Twin, Triple, and Quadruple. The raw dataset included room category, unit identifier, check-in and check-out dates, reservation duration, booking source, and reservation price. Records with blank room labels, observations outside the six focal categories, and rows with missing booking-price information were excluded.

For forecasting and pricing analysis, the reservation data were transformed into a daily analytical structure. Reservation-level prices were converted into implied nightly rates by dividing the total booking value by the reservation duration, and each reservation was expanded into a stay-date panel with each occupied night contributing a room-night observation. The cleaned panel contained 4,341 room nights and was aggregated by stay date and room category. Calendar-based variables, including weekends and locally relevant events, were incorporated to improve sensitivity to short-term demand variation.

A single-case design was considered appropriate because the aim of the study is to evaluate an applied decision-support framework under realistic small-hotel constraints rather than to claim broad statistical generalization.

### 3.2. Forecasting Module

The forecasting stage compared two complementary methods, Prophet and Extreme Gradient Boosting (XGBoost), across the six room categories. Prophet was included as an interpretable time-series model for capturing trend and seasonal structure (Taylor & Letham, 2018). In the final specification, Prophet was estimated using additive seasonality, a changepoint prior scale of 0.01, and a seasonality prior scale of 1, with weekend and local-event indicators included as external regressors.

XGBoost was used as a more flexible forecasting model capable of capturing nonlinear relationships and sharper short-term variation (Chen & Guestrin, 2016). The final XGBoost specification used a learning rate of 0.03, a maximum tree depth of 2, a minimum child weight of 1, a subsample ratio of 0.8, a column sampling ratio of 0.8, and 100 boosting rounds. Predictor construction was based on the category-level daily series and incorporated recent demand history and calendar-related effects.

Forecast validation followed a rolling-origin design with repeated 30-day prediction windows across sequential folds, approximating the operational rhythm of hotel revenue management. Forecast accuracy was assessed using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). To complement these descriptive measures, paired t-tests were conducted on daily absolute forecast errors.

The results showed that forecasting performance was category-specific. XGBoost achieved lower MAE and RMSE in Deluxe Double, Double with Sea View, Quadruple, Triple, and Twin rooms, while Prophet achieved a slightly lower MAE in the Double category, and XGBoost achieved a lower RMSE in that same category. Statistically significant differences in favor of XGBoost were observed for Double with Sea View, Quadruple, and Triple rooms. Prophet was therefore retained as an interpretable benchmark, while XGBoost provided a stronger predictive basis for several room categories.

### 3.3. Optimization Module

The second stage of the framework translated forecasted demand into pricing recommendations through a Mixed-Integer Linear Programming (MILP) model. This approach was selected because it provides a transparent and operationally tractable way to support pricing decisions under explicit business constraints.

The optimization objective was defined as the maximization of projected revenue across the six room categories:

$$\text{Maximize } R = \sum_{i=1}^{n} p_i q_i$$

where $p_i$ denotes the daily price assigned to room category $i$ and $q_i$ denotes the forecasted demand for that category.

In the baseline specification, forecasted demand was treated as an exogenous input. Accordingly, the optimization stage did not estimate the full behavioral response of demand to price but generated feasible pricing recommendations conditional on the forecasts produced by Prophet or XGBoost. This simplifying assumption was deliberately adopted to preserve interpretability and practical usability in a data-constrained setting.

The principal decision variables were the daily room rates assigned to the six focal categories. These were constrained by four conditions: total allocated demand could not exceed the hotel's capacity of 40 rooms; category-level prices were restricted to managerially acceptable intervals anchored to observed nightly-rate patterns; prices had to remain above minimum operating thresholds; and non-negativity conditions were imposed on prices and forecast quantities. In the analytical sample, the average implied nightly rates used as empirical anchors were approximately €68.41 for Double rooms, €82.23 for Double with Sea View, €84.76 for Deluxe Double, €90.08 for Quadruple, €78.88 for Triple, and €70.59 for Twin rooms.

The optimization model was implemented in R using the lpSolve and ompr packages. Forecast inputs from Prophet and XGBoost were alternately supplied to compare how different forecasting structures affected downstream pricing recommendations and projected revenue scenarios. Because demand was treated as fixed conditional on the forecast, the optimization outputs should be interpreted as scenario-based decision-support results rather than as structural estimates from a fully endogenous price-demand system. The resulting rate recommendations were designed for subsequent visualization in a managerial dashboard.

## 4. Results And Analysis

### 4.1. Revenue and Room-Night Distribution

**Table 1**

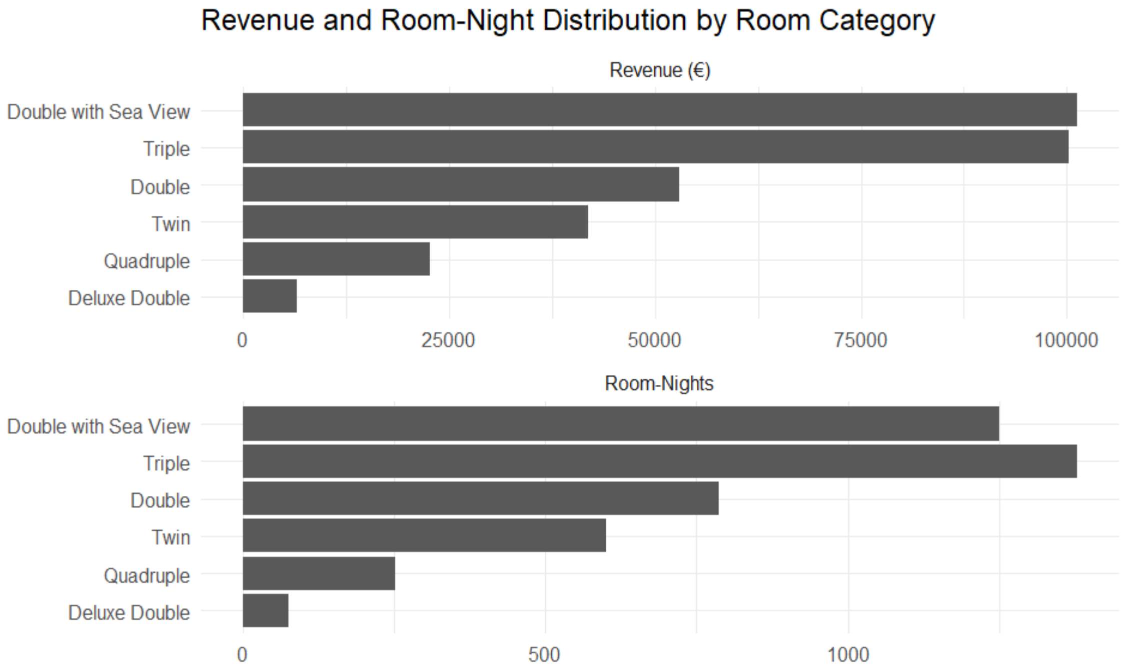
*Revenue and Room-Night Distribution by Room Category (1 April 2024–14 February 2025).*

| Room Category | Avg. Nightly Rate (€) | Revenue (€) | Reservations | Room-Nights |
|---|---|---|---|---|
| **Deluxe Double** | 84,76 | 6,556 | 38 | 74 |
| **Double** | 68,41 | 52,964 | 436 | 787 |
| **Double with Sea View** | 82,23 | 101,234 | 657 | 1,250 |
| **Quadruple** | 90,08 | 22,634 | 111 | 251 |
| **Triple** | 78,88 | 100,261 | 664 | 1,379 |
| **Twin** | 70,59 | 41,887 | 349 | 600 |
| **Total** | — | 325,536 | 2,255 | 4,341 |

The descriptive statistics show substantial heterogeneity across room categories. Double with Sea View generated the highest total revenue (€101,234), closely followed by Triple rooms (€100,261), while Triple rooms also accounted for the highest number of room-nights (1,379). Quadruple rooms recorded the highest average nightly rate (€90.08), whereas Deluxe Double accounted for only a small share of total revenue because of its limited volume. Figure 1 confirms that room categories differ not only in price level but also in sales volume, supporting the use of category-specific forecasting and pricing decisions rather than a single aggregated rule.

**Figure 1**

*Revenue and Room-Night Distribution by Room Category.*



### 4.2. Forecasting Model Performance

**Table 2**

*Forecast Accuracy: XGBoost vs. Prophet*

| Room Category | MAE (XGBoost) | MAE (Prophet) | RMSE (XGBoost) | RMSE (Prophet) |
|---|---|---|---|---|
| **Deluxe Double** | 0.313 | 0.332 | 0.382 | 0.410 |
| **Double** | 1.557 | 1.510 | 1.905 | 2.270 |
| **Double with Sea View** | 2.279 | 2.957 | 2.743 | 3.739 |
| **Quadruple** | 0.548 | 0.761 | 0.670 | 0.966 |
| **Triple** | 2.337 | 3.872 | 2.917 | 4.854 |
| **Twin** | 1.244 | 1.332 | 1.642 | 1.749 |

Forecasting performance varied by room category. XGBoost achieved lower MAE and RMSE for Deluxe Double, Double with Sea View, Quadruple, Triple, and Twin room types. The largest improvements were observed in Triple rooms, where MAE decreased from 3.872 to 2.337 and RMSE from 4.854 to 2.917, and in Double with Sea View rooms, where MAE declined from 2.957 to 2.279 and RMSE from 3.739 to 2.743. In the Double category, the comparison was mixed: Prophet achieved a slightly lower MAE, whereas XGBoost achieved a lower RMSE. Figure 2 visually supports the same pattern, indicating that XGBoost generally performed better in several of the hotel's more variable categories, while Prophet remained a useful interpretable benchmark.

**Figure 2**

*Forecast Accuracy Comparison by Room Category.*

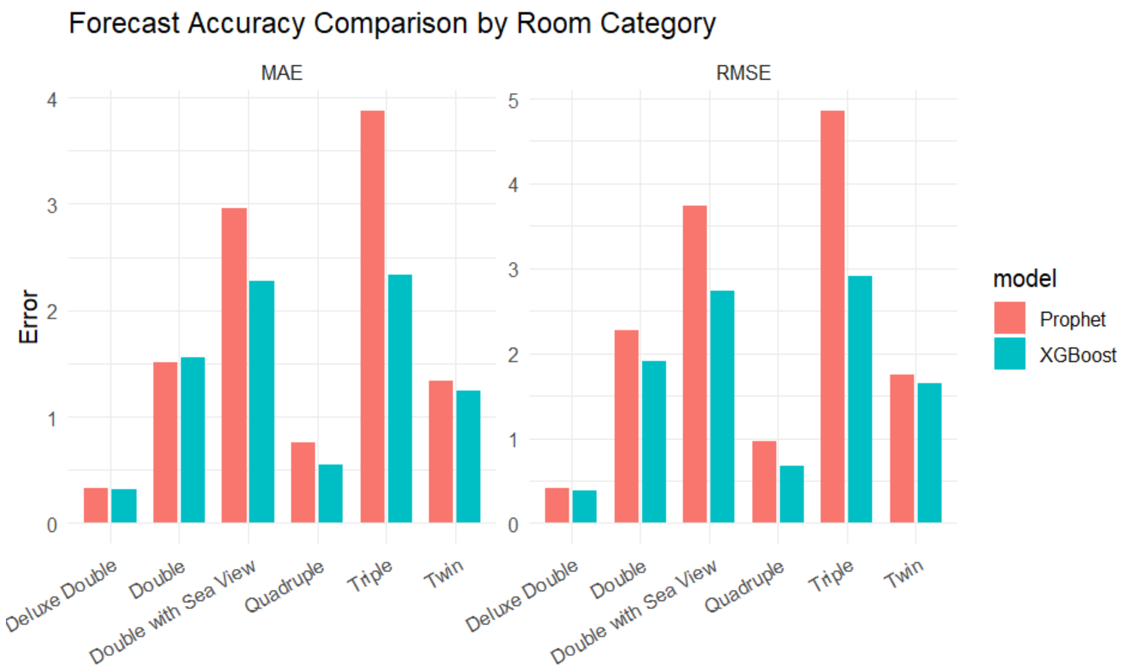


**Table 3**

*Mean Forecast Errors and Standard Deviations Across Validation Folds.*

| Room Category | MAE (XGBoost) Mean | MAE (XGBoost) SD | MAE (Prophet) Mean | MAE (Prophet) SD | RMSE (XGBoost) Mean | RMSE (XGBoost) SD | RMSE (Prophet) Mean | RMSE (Prophet) SD | Folds |
|---|---|---|---|---|---|---|---|---|---|
| **Deluxe Double** | 0.313 | 0.021 | 0.332 | 0.106 | 0.381 | 0.016 | 0.407 | 0.067 | 3 |
| **Double** | 1.557 | 0.760 | 1.510 | 1.272 | 1.774 | 0.802 | 1.964 | 1.314 | 4 |
| **Double with Sea View** | 2.279 | 0.867 | 2.957 | 1.798 | 2.663 | 0.804 | 3.395 | 1.917 | 3 |
| **Quadruple** | 0.548 | 0.252 | 0.761 | 0.476 | 0.614 | 0.310 | 0.861 | 0.507 | 4 |
| **Triple** | 2.337 | 0.854 | 3.872 | 2.486 | 2.863 | 0.645 | 4.433 | 2.283 | 4 |
| **Twin** | 1.244 | 0.314 | 1.332 | 0.370 | 1.619 | 0.333 | 1.703 | 0.489 | 3 |

Table 3 complements the point estimates in Table 2 by reporting mean errors and standard deviations across validation folds. In several categories, especially Double with Sea View, Quadruple, and Triple, XGBoost showed both lower mean errors and lower fold-level variability than Prophet, indicating a more stable predictive profile across repeated prediction windows.

### 4.3. Statistical Comparison of Forecast Errors

**Table 4**

*Paired t-Tests on Daily Absolute Forecast Errors (XGBoost − Prophet).*

| Room Category | n pairs | Mean Difference | t | df | p | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|---|---|
| **Deluxe Double** | 90 | -0.019 | -0.783 | 89 | 0.436 | -0.068 | 0.029 |
| **Double** | 120 | 0.047 | 0.340 | 119 | 0.734 | -0.226 | 0.319 |
| **Double with Sea View** | 90 | -0.679 | -3.458 | 89 | 0.0008 | -1.069 | -0.289 |
| **Quadruple** | 120 | -0.213 | -3.889 | 119 | 0.0002 | -0.322 | -0.105 |
| **Triple** | 120 | -1.535 | -6.411 | 119 | <0.001 | -2.009 | -1.061 |
| **Twin** | 90 | -0.088 | -0.913 | 89 | 0.364 | -0.279 | 0.103 |

The paired t-tests confirm that XGBoost's comparative advantage was concentrated in selected room categories. Negative mean differences indicate lower daily absolute forecast errors for XGBoost relative to Prophet. Statistically significant differences in favor of XGBoost were found for Double with Sea View, Quadruple, and Triple rooms, whereas differences for Deluxe Double, Double, and Twin rooms were not statistically significant.

### 4.4. Temporal Patterns in Predicted and Observed Demand

**Figure 3**

*Actual vs. Predicted Daily Demand for Selected Room Categories.*

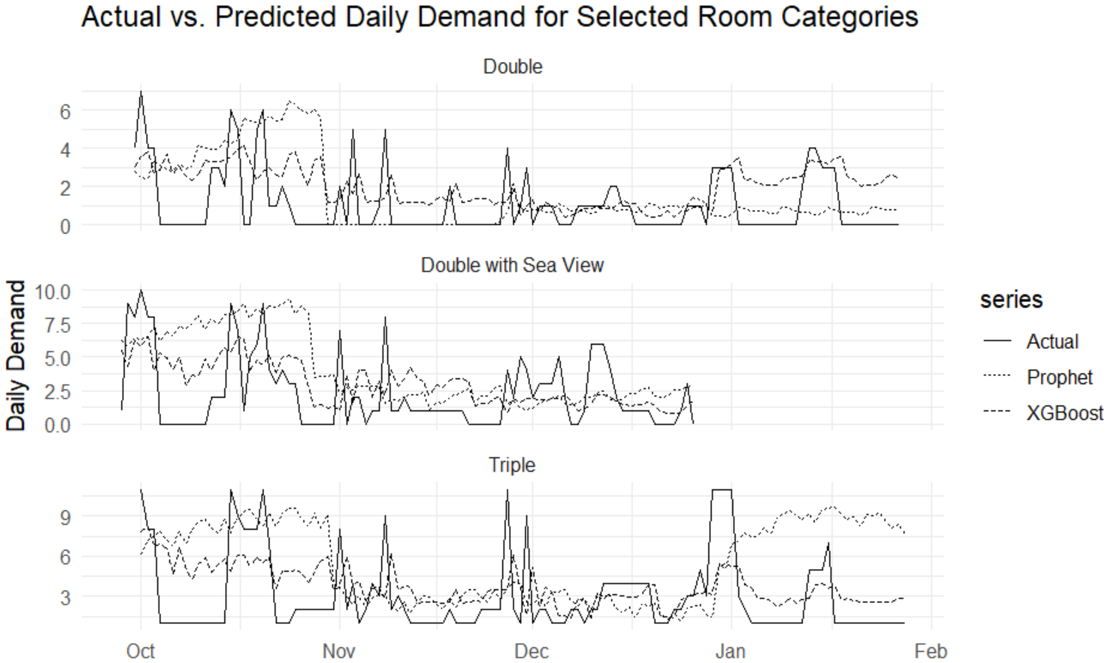



Figure 3 illustrates daily demand trajectories for Double, Double with Sea View, and Triple rooms. Both models broadly follow changes in demand over time, but Prophet produces smoother trajectories, whereas XGBoost reacts more closely to short-term fluctuations. This difference is most visible in Double with Sea View and Triple rooms, where XGBoost tracks declines and rebounds more sharply.

**Figure 4**

*Observed Daily Average Nightly Rate by Room Category.*

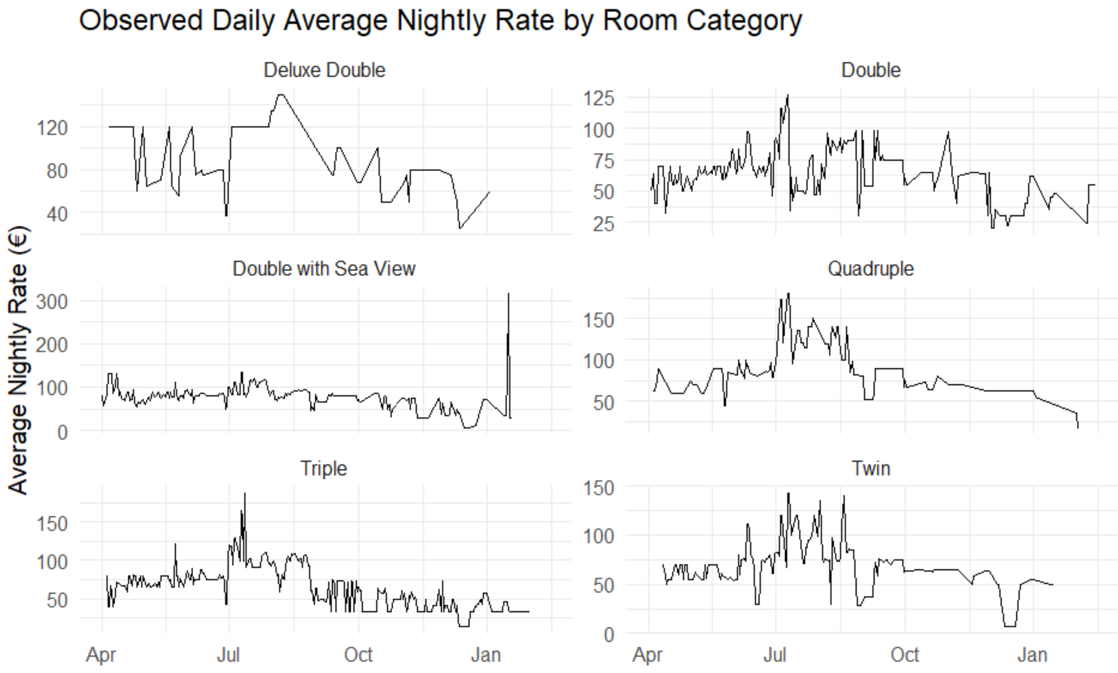



Figure 4 presents observed daily average nightly rates by room category. The figure shows substantial variation both between categories and over time, with Quadruple and Deluxe Double rooms generally maintaining higher rate levels than Double and Twin rooms. Several short-term spikes are also visible, especially in Double with Sea View, indicating that rate dynamics were neither fully stable nor uniformly volatile across the hotel's inventory.

### 4.5. Optimization Outcomes

Forecast outputs were used as inputs to the mixed-integer linear programming model in order to generate daily pricing recommendations under room-capacity and pricing constraints. Because the optimization stage treated forecasted demand as exogenous, the resulting outputs should be interpreted as scenario-based pricing recommendations rather than as estimates from a fully endogenous price-demand system. The main contribution of this stage is therefore practical: it shows that forecast-informed pricing scenarios can be generated transparently, under explicit constraints, and in a form suitable for managerial review.

Overall, the results indicate that AI-assisted pricing support is feasible in a small-hotel setting with limited data. The descriptive analysis confirms substantial heterogeneity across room categories, while the forecasting comparison shows that model performance is category-specific rather than universal. In the present case, XGBoost generally achieved lower prediction errors, particularly in Double with Sea View, Quadruple, and Triple rooms, whereas Prophet remained a useful interpretable benchmark. Taken together, the findings support a context-sensitive decision-support approach that balances predictive performance, interpretability, and operational feasibility.

## 5. Discussion, Implications, Limitations and Conclusions

The study examined whether an AI-assisted pricing framework could support revenue management in a small-hotel setting characterized by limited data and operational constraints. The results show that forecasting performance was category-specific rather than uniform across the hotel's inventory. XGBoost generally achieved lower prediction errors, with statistically significant advantages in Double with Sea View, Quadruple, and Triple rooms, whereas Prophet remained useful as an interpretable benchmark. This suggests that, in data-constrained hospitality settings, model choice should depend on the behavior of individual room categories rather than on the assumption that one forecasting approach will dominate universally.

The results also clarify the practical value of combining forecasting with constrained optimization. In the present framework, forecast outputs were translated into daily pricing recommendations through a mixed-integer linear programming model subject to room-capacity and pricing constraints. Although the optimization stage did not estimate a fully endogenous demand-price relationship, it demonstrated that forecast-informed pricing scenarios can be generated transparently and, in a way, suitable for managerial use. The contribution of the framework, therefore, lies less in claiming a universally optimal pricing rule and more in showing how predictive outputs can be converted into operationally feasible decision support.

From a managerial perspective, the findings suggest that small hotels do not necessarily require highly complex systems to benefit from AI-assisted pricing. More important is the use of methods that balance predictive performance with interpretability and usability. In this respect, the dashboard component remains relevant as a decision-support interface because it allows category-level demand signals and pricing scenarios to be reviewed in a practical and accessible form. More broadly, the study suggests that relatively lightweight analytical tools can support more systematic pricing decisions in small hospitality businesses and may help lower barriers to digital adoption in emerging tourism markets.

Several limitations should be acknowledged. The analysis is based on a single hotel, which limits external generalizability, and the observation window may not capture longer-term structural changes in demand. In addition, the optimization model treats forecasted demand as exogenous and should therefore be interpreted as a scenario-based decision-support tool rather than as a structural model of market-clearing prices. The dashboard concept was also not evaluated through user testing.

Future research could apply the framework to multiple hotels and destinations, incorporate additional predictors such as booking-channel composition, weather, events, or online review signals, and extend the optimization stage through explicit price-response assumptions or richer sensitivity analysis. User-oriented evaluation of dashboard design would also strengthen understanding of how forecasting and pricing models are interpreted and adopted in practice.

Overall, the study shows that AI-assisted pricing support can be implemented in a small-hotel context using a transparent and operationally feasible framework. Its main contribution lies in demonstrating how category-level forecasting, constrained optimization, and managerial interpretation can be linked within a practical decision-support pipeline.

## References

Alrawadieh, Z., Alrawadieh, Z., & Cetin, G. (2021). Digital transformation and revenue management: Evidence from the hotel industry. *Tourism Economics*, *27*(2), 328–345. https://doi.org/10.1177/1354816620901928

Ampountolas, A., & Legg, M. (2025). A comprehensive approach to enhancing short-term hotel cancellation forecasts through dynamic machine learning models. *Tourism Economics*, *32*(1), 321-341. https://doi.org/10.1177/13548166251318768

Buhalis, D., & Leung, R. (2018). Smart hospitality—Interconnectivity and interoperability towards an ecosystem. *International Journal of Hospitality Management*, *71*, 41–50. https://doi.org/10.1016/j.ijhm.2017.11.011

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Henriques, H., & Nobre Pereira, L. (2024). Hotel demand forecasting models and methods using artificial intelligence: A systematic literature review. *Tourism & Management Studies, 20*(3), 39–51. https://doi.org/10.18089/tms.20240304

Huang, L., & Zheng, W. (2023). Hotel demand forecasting: A comprehensive literature review. *Tourism Review, 78*(1), 218–244. https://doi.org/10.1108/TR-07-2022-0367

Ivanov, S., & Webster, C. (2024). Automated decision-making: Hoteliers' perceptions. *Technology in Society*, *76*, 102430. https://doi.org/10.1016/j.techsoc.2023.102430

Mariani, M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231–278. https://doi.org/10.1108/IJCHM-03-2021-0301

Nowak, M., & Pawłowska-Nowak, M. (2024). Dynamic pricing method in the e-commerce industry using machine learning. *Applied Sciences, 14*(24), 11668. https://doi.org/10.3390/app142411668

Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, *25*(17), 2733–2750. https://doi.org/10.1080/13683500.2021.1999397

Talluri, K. T., & Van Ryzin, G. (2005). *The theory and practice of revenue management* (pbk). Springer.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Rua Jaime Lopes Amorim
4465-004 S. Mamede de Infesta
Porto, Portugal

3rd
INTERNATIONAL
CONFERENCE ON
HUMANITIES AND
SOCIAL SCIENCES