Perfectus STUDENT 3/2026

Bi želeli objaviti strokovni članek?
Iščemo pisce in recenzente za naslednjo številko.
Pišite nam na: zalozba.perfectus@gmail.com
www.andrejraspor.com

Kontakti revije

Založnik revije Perfectus STUDENT

Perfectus
Svetovanje in izobraževanje, dr. Andrej Raspor, s. p.
Dolga Poljana 57
5271 Vipava, SI Slovenija
E-pošta: zalozba.perfectus@gmail.com

Glavni urednik

Andrej Raspor

Odgovorni uredniki

Anton Vorina
Bojan Macuh
Drago Papler

Uredniški odbor revije

Andrej Raspor
Bojan Macuh
Anton Vorina
Drago Papler

Jezikovni pregled

Za jezikovno ustreznost in vsebino prispevkov so odgovorni avtorice in avtorji, ki odgovarjajo tudi za morebitne kršitve avtorskih pravic.

Fotografije

Slika na naslovni strani: Image by storyset on Freepik
Slika na hrbtni strani: Image by storyset on Freepik

Arhiv revij Perfectus STUDENT

https://www.andrejraspor.com/index.php/sl/nase-storitve/zaloznistvo/perfectus-student

Mednarodna standardna serijska številka
(on line) **ISSN 3024-0018**

Perfectus STUDENT

Področje in opis revije

Revija Perfectus STUDENT je interdisciplinarna strokovna revija, ki objavlja prispevke študentov družboslovja. Vsebina ni omejena, ampak smo za vaše predloge odprti. Še posebej želimo objavljati prispevke, ki obravnavajo nove in aktualne teme in predstavljajo dosežke v razvoju ter njihovo uvajanje in uporabo v praksi. Vsled tega objavljamo tudi tematske številke.

Pogostost izhajanja

Revija Perfectus STUDENT izhaja dvakrat letno.

Politika za prosti dostop

Revija Perfectus STUDENT zagotavlja sprotni odprti dostop do vsebine v skladu z definicijo odprtega dostopa na podlagi načela, da prosti dostop do raziskav javnosti podpira večjo globalno izmenjavo znanja.

Navodila avtorjem

V reviji Perfectus STUDENT objavljamo strokovne članke, rezultate raziskovalnega dela študentov. Prispevki so napisani v slovenskem jeziku. Avtorji so odgovorni za ustrezen pravopis in vse morebitne kršitve avtorskih pravic. Prispevki niso honorirani.

Besedilo naj bo oblikovano po navodilih (interesenti nam pišite, da vam posredujemo predlogo z bolj podrobnimi navodili). Na začetku prispevka, takoj za naslovom naj bo povzetek dolžine 3–5 vrstic z do 4 ključnimi besedami. Obseg članka naj bo 4 – 12 strani. V primeru daljših tekstov si uredništvo pridržuje pravico, da besedilo krajša.

Predložite tudi kratek strokovni življenjepis vsakega od avtorjev (2–3 vrstice). Članki morajo biti pred objavo lektorirani. Ne uporabljajte opomb v besedilu. Eventualne opombe, ki naj bodo kratke, navedite na dnu besedila skupaj z literaturo. Seznam citirane literature oblikujte po APA-standardu.

Predložene prispevke pregledata in ocenita najmanj dva recenzenta. Na osnovi mnenj in predlogov recenzentov uredniški odbor ali urednik sprejmeta prispevek, zahtevata manjše ali večje popravke in dopolnitve ali ga zavrneta. Če urednik oziroma recenzenti predlagajo večje popravke, se dopolnjeni prispevek praviloma pošlje v ponovno recenzijo.

Iz tekoče vsebine stran

Pismo urednika

ANDREJ RASPOR

KO PRVA GENERACIJA EMUNI SOUSTVARJA DIGITALNO DRUŽBO

Spoštovane bralke, spoštovani bralci,

Pred vami je posebna številka revije *Perfectus STUDENT*, saj jo soustvarja **prva generacija študentov Univerze EMUNI**, ki so pri predmetu *Temelji digitalne družbe* (The foundations of the digital society) stopili v vlogo raziskovalcev, analitikov in avtorjev. Gre za izjemno mednarodno skupino – študenti prihajajo iz različnih kulturnih, jezikovnih in družbenih okolij, kar je v njihovo delo vneslo raznolikost perspektiv, bogastvo izkušenj in zrelost razmisleka, ki presega običajne študentske prispevke.

Digitalna družba je globalna, zato je prav, da jo prvi raziskujejo tisti, ki prihajajo iz globalnega okolja. Njihovi prispevki v tej številki odražajo to raznolikost: od vprašanj **dostopnosti in vključevanja**, do **umetne inteligence v računovodstvu**, **psiholoških učinkov digitalizacije**, **digitalnih pravic**, **e-uprave**, **algoritmov v politični komunikaciji**, **dinamičnega oblikovanja cen** in **preoblikovanja dela v dobi AI**. Vsak članek odpira pomembno razpravo o tem, kako tehnologija vpliva na človeka, družbo in institucije – in kako lahko kot družba odgovorno oblikujemo digitalno prihodnost.

Študenti EMUNI so pokazali, da razumejo bistvo sodobne digitalne družbe: da tehnologija ni le orodje, temveč prostor, kjer se prepletajo etika, politika, ekonomija, psihologija in človekove pravice. Njihovi prispevki dokazujejo, da mladi raziskovalci ne le sledijo trendom, temveč jih znajo kritično ovrednotiti in umestiti v širši družbeni kontekst.

Naj vas njihova razmišljanja spodbudijo k premisleku, dialogu in morda tudi k lastnemu raziskovanju. Digitalna družba je prostor, ki ga soustvarjamo vsi – in prav študenti prve generacije EMUNI so pokazali, kako dragocen je glas mladih v tem procesu.

Želim vam poglobljeno, navdihujoče in koristno branje.

dr. Andrej Raspor
glavni urednik

Uvodnik programske direktorice magistrskega študijskega programa

MAJA PUCELJ

KO ŠTUDENTI SKUPAJ USTVARJAJO ZNANJE

Z veseljem predstavljam izdajo revije Perfectus STUDENT za leto 2026. Ta številka obravnava, kako digitalne tehnologije spreminjajo človekove pravice, družbene strukture in strokovno prakso.

Osem člankov v tej številki so napisali študenti EMUNI universe v okviru magistrskega študijskega programa Digitalizacija in človekove pravice, znotraj predmeta Človekove pravice v digitalni dobi. Posebej zanimiv se mi zdi razpon tem, ki so jih izbrali za svoje raziskovanje. Pisali so o različnih vprašanjih – od računalniških vmesnikov za osebe z invalidnostmi do vpliva algoritmov na zaupanje v politiko v državah, ki so doživele konflikte.

Ti študenti ne pišejo zgolj o tem, kako odlična je tehnologija. Ukvarjajo se tudi s izzivi, ki jih prinaša digitalizacija. Razmišljajo o stroških in kompromisih, ki jih sprejemamo ob uporabi tehnologij. Prav tako se sprašujejo, kdo nosi odgovornost, ko odločitve sprejemajo računalniški sistemi.

Skozi vse članke se pojavlja več osrednjih idej. Ena izmed njih je, da digitalni dostop in oblikovanje tehnologij niso enako pravični za vse ljudi. Lahko pomagajo posameznikom pri uresničevanju njihovih pravic ali pa jih pri tem omejujejo. To jasno pokažejo članki o digitalnem razkoraku in umetni inteligenci v javni upravi.

Druga pomembna ideja je vprašanje odgovornosti pri računalniško podprtih odločitvah. To vprašanje se pojavlja v vsakem članku. Študenti razmišljajo tudi o tem, kako tehnologija vpliva na naše mišljenje in medosebne odnose. Poudarjajo, da moramo o teh vprašanjih razmišljati skupaj s tehnološkimi, pravnimi in ekonomskimi vidiki.

Menim, da je izjemno pomembno, da si študenti že na začetku svoje akademske poti zastavljajo takšna vprašanja. Ne bojijo se kritičnega razmišljanja in ne iščejo preprostih odgovorov. Prav to potrebujemo od raziskovalcev v digitalni dobi.

Upam, da bo ta številka za vse bralce zanimiva ter spodbudna za nadaljnji razmislek.

doc. ddr. Maja Pucelj

Programska direktorica magistrskega študijskega programa Digitalizacija in človekove pravice na EMUNI Univerzi

DEVELOPING INTELLIGENT INTERFACES FOR PEOPLE WITH DISABILITY

AHMAD GUL AHMADZAI

Abstract

Purpose: *This paper reviewed contemporary intelligent human–computer interfaces designed to improve accessibility for people with disabilities, focusing on eye tracking, voice control, gesture recognition, adaptive interfaces, and brain–computer interaction. This paper examines the problems with current computer technology for people with disabilities and explores practical, intelligent solutions to improve digital accessibility.*
Methodology: *The study examines the problems with current computer technology for people with disabilities and explores practical, intelligent solutions to improve digital accessibility. To address this aim, a systematic literature review was conducted using IEEE Xplore, ACM Digital Library, PubMed, and Google Scholar. Peer-reviewed articles published between 2000 and 2025 were analyzed using thematic synthesis to identify technological capabilities, accessibility outcomes, and ethical–legal considerations.*
Results and Conclusions: *The findings of this systematic review highlight both the promise and the complexities of intelligent interfaces for accessibility. The analysis indicates that these technologies can significantly reduce interaction barriers for users with motor, visual, auditory, and cognitive impairments, with multimodal and adaptive systems demonstrating particularly high usability. However, persistent challenges in cost, standardization, data privacy, and long-term adoption were also identified. Consequently, while intelligent interfaces offer potent practical solutions to overcome digital barriers, their implementation must be pursued with deliberate adherence to ethical principles and established legal frameworks such as the ADA and the UN CRPD. Future work should therefore focus not only on technological innovation but also on developing sustainable, equitable, and policy-aligned models for deployment.*
Research limitations: *The research is based on literature review without original empirical investigation.*
Practical and/or Social implications: *The findings emphasize integrating accessibility into technology development from the initial design phase.*
Originality: *The paper synthesizes technical solutions with human rights perspectives on digital inclusion.*

Keywords: *accessibility, inclusivity, human-computer interaction, assistive technology, user experience*

Introduction

Technology is essential for communication, education, work, and social participation in the modern world. Yet, for people with disabilities, interaction with standard digital systems—designed for so-called “average users”—often creates prohibitive barriers, leading to digital exclusion (Shneiderman et al., 2022).

This research is grounded in the social model of disability, which posits that disability arises from the interaction between individuals and environmental barriers, rather than from impairments themselves (Burgstahler, 2020). From this perspective, inaccessible technology is a socially constructed barrier. Consequently, accessibility must be a core component of usability and user experience (Dix et al., 2003), not addressed through separate, often-limiting assistive tools.

Intelligent interfaces—such as eye tracking, voice control, and adaptive systems—represent a paradigm shift toward interaction models that adapt to the user. They align with Universal Design principles and offer a path to dismantle digital barriers. This review therefore addresses the following question: How do intelligent interfaces contribute to accessibility and inclusion for people with disabilities, and what ethical, legal, and practical imperatives govern their implementation?

Supporting Research Questions:

1. What are the main limitations of current computer technologies in meeting the needs of people with disabilities?
2. How can intelligent interfaces (eye tracking, voice control, gesture recognition, adaptive systems, brain-computer interfaces) improve accessibility and inclusivity?
3. To what extent is accessible design a moral duty and a legal requirement in digital society?
4. What are the real-world implications of implementing intelligent interfaces for people with disabilities in education, employment, and social participation?
5. How can multimodal, adaptive systems be developed to provide context-aware, user-centered interaction for diverse disabilities?

Literature Review

Human-Computer Interaction (HCI) has long studied how people and computers communicate, with emphasis on usability and efficiency. Traditional interaction methods such as keyboards and mice have proven inaccessible for many disabled users, creating barriers to

participation in the digital society. Prior research has explored assistive devices such as screen readers, specialized keyboards, and switch systems, but these often suffered from high cost, discomfort, and limited applicability.

Recent studies have advanced intelligent interfaces significantly. For example, Alvarez-Marquez et al. (2022) demonstrated AI-driven adaptive interfaces that personalize interaction in real-time. Majaranta and Bulling (2021) provided a comprehensive review of eye-tracking applications for people with motor impairments, showing significant improvements in communication and control. Similarly, Kumar et al. (2023) presented a low-cost gesture recognition system using depth sensors, making gesture-based control more accessible. There are already some methods and studies that align with this topic, as mentioned in my reference section. For example, research has been conducted on Smart Assistive Devices for People with Disabilities, the development of intelligent wheelchairs for the handicapped, and various smart home appliances designed to support disabled users. However, my research differentiates itself by focusing on technology as an essential need of modern life for people with disabilities, whether partial or complete. Unlike prior work that targets specific impairments, the aim here is to develop inclusive frameworks and solutions that address a wide spectrum of disabilities.

The central aim of this research is to design and evaluate inclusive technological solutions that empower individuals across diverse disability categories. This holistic approach recognizes that technology is not merely a convenience but a basic right for disabled people in today's interconnected world.

Methodology

This paper employs a systematic literature review methodology to examine intelligent interfaces for people with disabilities. The search was conducted across academic databases including IEEE Xplore, ACM Digital Library, PubMed, and Google Scholar, using keywords such as "assistive technology," "eye tracking," "voice control," "gesture recognition," and "brain-computer interface" combined with "disability" and "accessibility." Articles published between 2000 and 2025 were considered, with emphasis on recent studies (2020–2025). Inclusion criteria focused on empirical studies, review articles, and case studies demonstrating practical implementations. The analysis followed a thematic approach to identify key technologies, challenges, and ethical-legal frameworks.

Our approach aligns with accessibility-by-design principles as outlined in international standards (e.g., EN 301 549).

Search Strategy Details: Boolean operators were used to combine search terms: ("intelligent interface" OR "adaptive interface") AND ("disability" OR "accessibility") AND ("eye tracking" OR "voice control" OR "gesture recognition"). This approach ensured comprehensive coverage while maintaining relevance to the research questions.

Data Extraction: From initially identified articles, those meeting inclusion criteria were analyzed using a structured protocol that captured: study objectives, technological specifications, outcomes, and limitations. Thematic analysis identified recurring patterns across the literature.
Limitations of Methodology: This review-based approach relies on published literature and does not include original empirical data collection with disabled users. Future research should address this through participatory design approaches.

Results

Current Technology Limitations

Mainstream digital technology often assumes an “average” user with full visual, motor, auditory, and cognitive abilities, thereby excluding individuals with disabilities. Conventional assistive devices, while available, frequently suffer from high cost, limited adaptability, and usability constraints—highlighting a clear need for more inclusive, intelligent solutions.

Table 1: Global Disability Prevalence Estimates (WHO, 2011)

Category	Population (in Millions)	% of World Population
Total World Population	8,000	100%
Disabled Population (all types)	1,300	16.25%
Blindness / Visual Impairment	253	3.16%
Deafness / Hearing Loss	466	5.82%
Intellectual Disability	200	2.50%
Wheelchair Users	75	0.94%

Note: The 2011 WHO report is used as it provides comprehensive data across different disability categories.

Intelligent Interface Modalities:

Eye Tracking: This system uses a camera to follow where a user is looking on the screen. Software then moves the computer cursor to that spot. The user can "click" by looking at a button for a few seconds or by blinking. This is very helpful for people who cannot move their hands but can control their eye movements, such as those with ALS or severe spinal injuries.

Voice Control and Speech Recognition: With this method, users control their computer by speaking. Using technology like Google Assistant, Apple Siri, or Windows Voice Recognition, they can say commands like "open document," "scroll down," or "delete file." This helps people with motor impairments or visual disabilities. Future systems could also understand the user's tone of voice to detect if they are confused or need help, making the interaction more natural.

Gesture Recognition: Cameras or sensors can detect a user's body or hand movements and turn them into commands. For example, a wave of the hand could act like a mouse click, or a nod could mean "yes." This is already used in some smart TVs and gaming systems. For computer users, it offers a way to interact without touching any device.

Other Advanced Methods:

- Adaptive User Interfaces: Software that changes automatically based on the user
- Brain-Computer Interfaces (BCI): Systems that read brain signals to control computers
- Haptic Feedback: Devices that give touch feedback through vibration
- Multimodal Systems: Approaches that combine several methods

Table 2: Comparison of Intelligent Interface Modalities

Modality	*Best For*	*Key Advantages*	*Primary Challenges*
Eye-Gaze Tracking	Users with severe motor impairments	Hands-free, fast pointing, uses preserved ability	Midas Touch, calibration, cost of high-end systems
Gesture Recognition	Users with motor control but unable to grasp	Intuitive, direct manipulation, no wearables	Ambient light sensitivity, limited vocabulary, user fatigue
Speech Recognition	Users with motor/visual impairments	Hands-free, natural for commands, widely available	Background noise, privacy concerns, poor for spatial tasks
Brain-Computer Interfaces	Users with complete motor paralysis	Direct neural control, high potential	Invasive (some types), low bandwidth, high cost, training intensive
Haptic Feedback	Visually impaired users	Conveys spatial/tactile information	Limited resolution, can be bulky, not standardized

Implementation Examples

Tobii Eye Tracker: A commercial product that lets people with disabilities communicate, learn, and control smart home devices using only their eyes (Tobii, 2023).

Microsoft Accessibility Suite: Tools like the Adaptive Controller for Xbox (with large, customizable buttons) and Windows Eye Control show how big companies are building accessibility into mainstream products (Microsoft, 2022).

Google Voice Access: A free Android app that allows complete hands-free control of a phone or tablet.

Assistive Robotic Arms: Research projects have successfully used eye-tracking to let people with paralysis control a robotic arm to eat or drink independently (Wolpaw & Wolpaw, 2022).

Low-Cost Eye Tracking: Academic researchers are developing software that works with ordinary webcams, making eye-controlled computing cheaper and more available (Majaranta & Bulling, 2021).

Ethical and Legal Frameworks

Analysis reveals that accessibility represents both ethical imperative and legal requirement. The Americans with Disabilities Act (1990) establishes general accessibility mandates in the United States. The EU Accessibility Act (2019) specifies technical requirements for digital products and services in Europe. The UN Convention on the Rights of Persons with Disabilities (2006) frames accessibility as a fundamental human right, influencing national legislation in 182 ratifying countries.
Ethical considerations emphasize autonomy, privacy, and fairness—particularly important for technologies like BCIs that involve neural data.

Discussion

Addressing Research Questions:

The findings provide comprehensive responses to the research questions:

- Research Question 1: The study confirmed that most mainstream technologies are designed for "average" users, leaving out those with motor, visual, auditory, or cognitive impairments. Existing assistive devices often suffer from high cost, discomfort, and limited applicability, reinforcing the need for more inclusive solutions.
- Research Question 2: Evidence from commercial products and research prototypes shows that intelligent interface modalities provide practical, hands-free, and adaptive ways for disabled users to interact with computers. Eye tracking enables precise control, voice recognition supports natural commands, and gesture systems allow intuitive interaction. Together, they demonstrate significant improvements in accessibility.
- Research Question 3: The analysis highlights that accessibility is both an ethical obligation and a legal mandate. Frameworks such as the ADA (1990), EU Accessibility Act (2019), and UN CRPD (2006) establish enforceable standards, while the moral dimension emphasizes fairness, autonomy, and inclusivity. Accessibility is therefore not optional but fundamental.
- Research Question 4: Case studies such as Microsoft's Adaptive Controller, Tobii Eye Tracker, and Google Voice Access demonstrate tangible impacts: enabling communication, supporting independent living, and opening opportunities in education and work. These examples show that accessible design directly enhances participation in society.
- Research Question 5: The findings suggest that the future of HCI lies in adaptive, multimodal systems that can switch between gaze, voice, and gesture depending on user context. Such systems would reduce fatigue, personalize interaction, and ensure flexibility. Developing these requires participatory design with disabled users as co-creators, alongside advances in sensors and algorithms.

Recent updates to accessibility guidelines, such as WCAG 2.2, emphasize embedding inclusive design principles throughout the software lifecycle.

The findings suggest that the future of human–computer interaction lies in adaptive, multimodal systems that can switch between gaze, voice, gesture, and other modalities depending on user context. Such systems reduce fatigue, personalize interaction, and ensure flexibility. Developing these requires participatory design with disabled users as co-creators, alongside advances in sensors and algorithms. *To illustrate this integration, Figure 1* presents a conceptual framework showing how diverse intelligent interfaces can be combined into a unified adaptive system for accessibility.

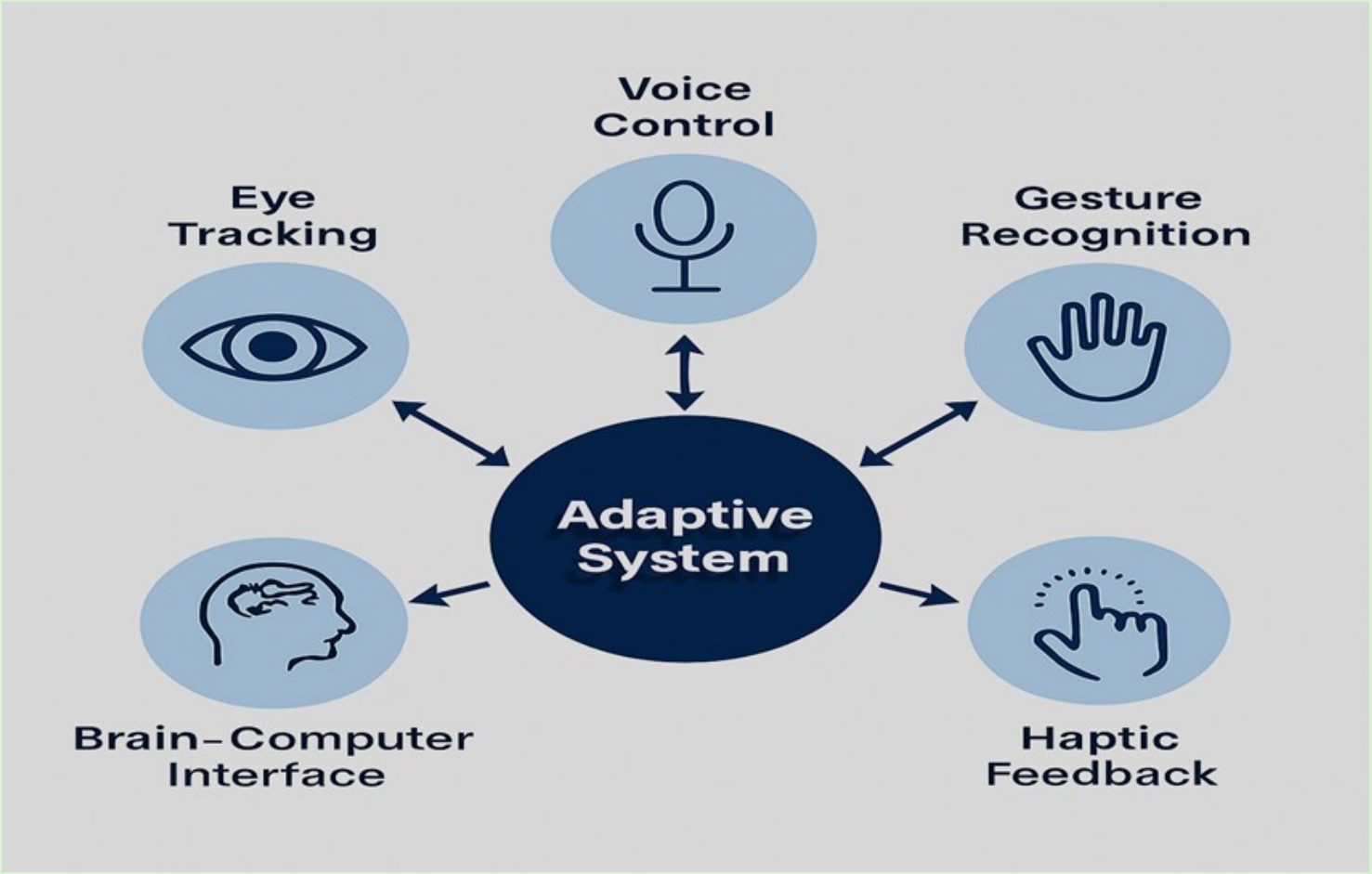


Figure 1: Integration of Multimodal Interfaces for Accessibility

Figure 1: Conceptual framework illustrating the integration of multimodal intelligent interfaces (eye tracking, voice control, gesture recognition, brain–computer interface, and haptic feedback) into an adaptive system for accessibility.

As shown in Figure 1, multimodal systems integrate diverse interaction modalities into a single adaptive hub. This model demonstrates how eye tracking, voice control, gesture recognition, brain–computer interfaces, and haptic feedback can work together to provide personalized, context-aware accessibility for users across diverse disability categories. Such integration aligns with Universal Design principles and emerging accessibility-by-design standards (WCAG 2.2, ISO/IEC TR 29138), reinforcing the need to embed accessibility into mainstream digital innovation.

Theoretical and Practical Contributions

Theoretical Contributions

First, this review extends human–computer interaction theory by synthesizing intelligent interface technologies as a coherent accessibility paradigm rather than isolated interaction techniques. Second, it contributes to digital society theory by framing accessibility as a structural condition for participation, positioning intelligent interfaces as mediators between technological innovation and social inclusion.

Practical Contributions

From a practical perspective, the findings provide guidance for designers and developers by demonstrating that accessibility can be embedded within mainstream systems through adaptive and multimodal interfaces. Additionally, the review offers policymakers and organizations a conceptual basis for aligning technological innovation with ethical responsibility and legal accessibility requirements, reinforcing accessibility-by-design as a standard practice in digital innovation.

Limitations and Future Research

This study has several limitations. First, the review primarily focuses on recent technological solutions, potentially overlooking longer-term socio-technical challenges in adoption and sustainability. Second, while multiple intelligent interfaces are discussed, the comparative analysis is based on existing literature rather than original empirical testing. Third, the legal and ethical discussion, though informed by key frameworks like the ADA and UN CRPD, does not address regional variations in policy implementation. Fourth, the perspectives of people with disabilities, while acknowledged as crucial, are not directly incorporated through primary data collection.

Future research should address these limitations through:

1. Empirical studies evaluating intelligent interface effectiveness across disability categories
2. Participatory design research involving disabled users as co-investigators
3. Longitudinal studies examining long-term impacts on independence and quality of life
4. Cross-cultural comparisons of accessibility implementation
5. Technical development of low-cost, open-source solutions

Future Directions and Emerging Standards

Looking ahead, the evolution of intelligent interfaces will be shaped by advances in explainable AI (XAI), wearable sensor networks, and ambient intelligence. Emerging standards such as the W3C's WCAG 3.0 (Silver) and ISO/IEC 23026-4 on accessibility-by-design provide frameworks for embedding inclusivity from the earliest stages of development. Additionally, the rise of federated learning offers a privacy-preserving approach to personalizing adaptive interfaces without centralizing sensitive user data. Future systems are likely to integrate real-time context awareness—using environmental and physiological data to adjust interaction modes dynamically. These developments should be guided by participatory design methodologies that position people with disabilities not as passive users, but as co-creators of the technologies intended to serve them.

Conclusion

This review has systematically examined how intelligent interfaces—including eye tracking, voice control, gesture recognition, and adaptive systems—contribute to digital accessibility and inclusion for people with disabilities. Grounded in human–computer interaction theory and principles of digital inclusion, the findings demonstrate that these adaptive, multimodal technologies effectively address the limitations of conventional "one-size-fits-all" interaction models.

The analysis confirms that intelligent interfaces not only reduce technical barriers but also uphold core ethical values such as autonomy, dignity, and fairness, while providing pathways for compliance with legal frameworks like the ADA, the EU Accessibility Act, and the UN CRPD. This work aligns with global commitments such as the United Nations Sustainable Development Goals (SDGs), particularly Goal 4 (Quality Education) and Goal 10 (Reduced Inequalities), reinforcing the role of accessible technology in building equitable societies.

Although persistent challenges in affordability, privacy, standardization, and equitable access must be addressed, these issues do not negate the transformative potential of intelligent interfaces in inclusive technology design. In conclusion, intelligent interfaces should be regarded as foundational components of contemporary digital systems, not as optional enhancements. By integrating accessibility at the theoretical, technical, and societal levels, they offer a robust framework for building a more inclusive digital society. Future progress depends on sustained collaboration between researchers, developers, policymakers, and disability communities to translate this potential into widespread, practical reality. Future work could benefit from closer alignment with emerging accessibility standards, including ISO/IEC 40500.

References

1. Alvarez-Marquez, J. Z., López, G., & Guerrero, L. A. (2022). AI-driven adaptive user interfaces for people with motor impairments. Universal Access in the Information Society, 21(3), 589–605. https://doi.org/10.1007/s10209-021-00841-z
2. Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 et seq.
3. Bennett, C. L., Brady, E., & Branham, S. M. (2018). Interdependence as a frame for assistive technology research and design. Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, 89–100. https://doi.org/10.1145/3234695.3236348
4. Burgstahler, S. (2020). Universal design in higher education: From principles to practice (2nd ed.). Harvard Education Press.
5. Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2003). Human-computer interaction (3rd ed.). Pearson.
6. Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services*, 2019 O.J. (L 151) 70. https://eur-lex.europa.eu/eli/dir/2019/882/oj
7. Floridi, L. (2019). The ethics of information. Oxford University Press.
8. Gajos, K. Z., Hurst, A., & Findlater, L. (2020). Personalized dynamic accessibility. ACM Transactions on Accessible Computing, 13(1), 1–32. https://doi.org/10.1145/3368424
9. Goggin, G. (2021). Disability and the media. Routledge.

10. Google. (2023). Voice Access: Hands-free control of your Android device. Retrieved January 26, 2026, from https://support.google.com/accessibility/android/answer/6151848
11. Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. Medical Reference Services Quarterly, 37(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391
12. Kumar, P., Verma, A., & Singh, S. (2023). Low-cost gesture recognition system for assistive technology using depth sensors. Journal of Ambient Intelligence and Humanized Computing, 14(5), 6123–6135. https://doi.org/10.1007/s12652-021-03546-6
13. Majaranta, P., & Bulling, A. (2021). Eye tracking and eye-based human–computer interaction. In S. H. Fairclough & K. Gilleade (Eds.), Advances in physiological computing (pp. 39–65). Springer. https://doi.org/10.1007/978-1-4471-6392-3_3
14. Microsoft. (2022). Microsoft accessibility: Technology for everyone. Retrieved January 26, 2026, from https://www.microsoft.com/en-us/accessibility
15. Oviatt, S., & Cohen, P. (2022). The paradigm shift to multimodal interaction. MIT Press.
16. Pittarello, F. (2021). Haptic feedback for blind users: State of the art and future directions. Universal Access in the Information Society, 20(2), 305–319. https://doi.org/10.1007/s10209-020-00718-7
17. Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., & Elmqvist, N. (2022). Designing the user interface: Strategies for effective human-computer interaction (7th ed.). Pearson.
18. Tobii. (2023). Eye tracking for communication and control. Retrieved January 26, 2026, from https://www.tobii.com
19. United Nations. (2006). Convention on the Rights of Persons with Disabilities. Retrieved January 26, 2026, from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html
20. World Health Organization. (2011). World report on disability. WHO Press.
21. Wolpaw, J. R., & Wolpaw, E. W. (Eds.). (2022). Brain-computer interfaces: Principles and practice. Oxford University Press.
22. Chen, J., Li, Y., & Wang, H. (2024). Adaptive Multimodal Interaction for Users with Motor Impairments: A Deep Learning Approach. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3613904.3642507
23. World Health Organization. (2023). Global Report on Assistive Technology. WHO Press. https://www.who.int/publications/i/item/9789240049451
24. W3C Web Accessibility Initiative. (2023). WCAG 3.0 Working Draft. Retrieved from https://www.w3.org/TR/wcag-3.0/

AI IN ACCOUNTING: SMARTER REPORTING, CSR, SECURITY, AND ETHICS

MAROUAN BEN TORKIA

Abstract

Purpose: *The article examines how artificial intelligence transforms accounting by improving reporting accuracy, strengthening CSR and governance practices, enhancing fraud detection and security, and reshaping the role of accountants. It aims to identify both opportunities and ethical, professional, and organizational challenges arising from AI adoption. As zapisano v dokumentu, "AI… may perform repetitive and tedious tasks, decrease humanity mistakes… but can also show challenges."*
Methodology: *The study is based on a qualitative literature review of recent academic and professional sources. It synthesizes findings on AI applications in accounting, including machine learning, deep learning, expert systems, robotics, and CSR-related technologies.*
Results and Conclusions: *AI significantly improves accounting efficiency, accuracy, and decision-making by automating repetitive tasks, detecting fraud, predicting financial trends, and supporting strategic analysis. It also enhances CSR reporting, environmental monitoring, and governance transparency. However, AI introduces challenges such as high implementation costs, ethical risks, data privacy concerns, accountability issues, and the need for continuous skill development. The profession is shifting from manual execution to analytical, managerial, and advisory roles.*
Research limitations: *The study relies exclusively on secondary literature and does not include empirical field research. The rapid evolution of AI technologies means that some findings may quickly become outdated. The document itself notes that "there seems to be insufficient data to support companies' readiness to adopt AI solutions."*
Practical and/or Social implications: *Organizations must invest in training, ethical governance, and secure AI infrastructures to fully benefit from AI-driven accounting. AI can strengthen CSR performance, reduce fraud, and improve sustainability reporting, but only when combined with human oversight. Accountants must develop hybrid competencies in finance, analytics, and technology to remain relevant.*
Originality: *The article integrates AI applications across accounting, CSR, security, and ethics into a unified conceptual framework. It highlights the interconnected nature of technological, organizational, and ethical dimensions and emphasizes the emergence of new professional roles shaped by AI-driven accounting systems.*

Keywords: *Artificial Intelligence; Accounting Automation; Corporate Social Responsibility; Fraud Detection; Ethical Governance*

Introduction

Artificial Intelligence is changing the accounting profession At an exceptional speed. What was once a field overcoming by manual data entry, routine reconciliations, and standardized reporting is now Developing into a strategic, technology-driven discipline. AI helps accountants to do jobs faster, and leads to better decisions, also highlight on higher-value tasks like analysis, advisory services, and ethical judgment.

On other hand, the implementation of AI into accounting is not without challenges. despite AI improves financial reporting accuracy, increase Corporate Social Responsibility (CSR) practices, and reiforce security systems, it indeed give intention to serious ethical concerns in cohesion to data privacy, transparency, bias, and professional responsibility. This article aims to show how AI is reshaping accounting through smarter reporting, CSR, security, and ethics, while focusing the importance of responsible and human-centered adoption.

This study seeks to answer the following research question: *How does artificial intelligence transform accounting practices while raising ethical, governance, and professional challenges?*

Theoretical Foundations of AI

Concept and Definition

Artificial intelligence is one of The key and fascinating subjects in computer science and technology. In general, AI related to systems enable to performe tasks that typically demand human intelligence and cognition, like learning, reflexion, decision-making, pattern recognition, language translation, and more. AI may be broadly divided into two groups: Strong AI and Weak AI Ablameyko & Ablameyko (2021).

Powrful AI, or Artificial General Intelligence, defined to systems that leads performing all smart activities to the same degree or better than humans. This kind of artificial intelligence has not yet get the reality and is mainly focusing in science-fiction stories and films. Giving example, robots that can talk with humans, knowing emotions, exhibit creativity, and release self-awareness Alghafiqi & Munajat (2022). Weak AI, or Artificial Specialized Intelligence, defined as a systems that producing intelligent tasks only within a specific domain or task. This kind of AI currently be present and is employ in different technological applications and services. Also, facial recognition systems, machine translation, computer games, self-driving cars, and so on. AI may indeed be categorized refer to the methods and techniques used. Some of these types contain:

- Transitional AI: This type of AI function related to a defined and limited rules and does not nreeds large amounts of data for learning. It is utilize in some control systems and computer games, like chess.
- Evolutionary AI: This category of AI operates refer to evolutionary algorithms and can adjust to the environment and elevate over time. It is operated in optimization problems and simulations.
- Neural Network AI: This kind of AI uses guided by artificial neural networks and can learn from large datasets, discover patterns and relationships within them. It is operates in pattern recognition, natural language processing, and image analysis Alghafiqi & Munajat. (2022). According to historical financial data. Employing machine learning algorithms and neural networks, historical patterns can be learned to predict growth, profitability, and future financial conditions.

Generic AI Applications

AI in Data Analysis

AI plays a critical role in data analysis. By using AI models and algorithms, large and complex datasets can be analyzed and converted into actionable knowledge and information. Below are some applications of AI in data analysis in accounting Goel et al. (2023):

Pattern Recognition

AI can identify hidden and discernible patterns in financial data. These patterns may include trends, correlations, sudden changes, and hidden relationships between data. Recognizing these patterns helps identify financial issues, risks, and opportunities.

Financial Prediction and Analysis

AI can be used to predict and analyze financial performance based on historical financial data. Using machine learning algorithms and neural networks, historical patterns can be studied to forecast growth, profitability, and future financial conditions.

Risk Analysis

AI become like an assistant in analyzing risks correlated to accounting, like credit risk, market risk, and tax risk. Indeed, by using AI algorithms, more complex models can be created for risk analysis, helping for making better decision in financial risk management.

Fraud Detection

AI become more useful in finding fraudulent patterns in financial and accounting data. By emplying machine learning algorithms and neural networks, fraudulent patterns become easily detected, and suspicious activities can be flagged.

Improved Decision-Making

AI leads to increase decision-making in financial and accounting processes. For this utilizing AI algorithms, forecast models and decision-making systems may be created that making guidance focusing on data and financial conditions, enhance decision-making processes Goel et al. (2023).

For That, AI, by using various algorithms and models, can aid help in data analysis in accounting and finance, aid become actionable insights through various financial domains. But, domain-specific expertise and specialized knowledge are important to fully leverage AI and release optimal outcomes.

AI in Fraud and Risk Detection

AI plays an important contibution for detection fraud and risk. By adoption AI algorithms and models, suspicious patterns and cases can be detected, assisting to mitigate fraud and associated risks. In this case some applications of AI in fraud and risk detection are :

Financial Fraud Detection

AI can defiy suspicious patterns and signals in financial transactions. Employing machine learning and neural network algorithms, suspecious behaviors in financial transactions can be identifie, aiding decrease financial fraud. This application is widely employed in banking, insurance, online payments, and other financial sectors Harayama et al. (2021).

Cybersecurity Fraud Detection:
AI help in finding cyber threats and security fraud. By using AI algorithms and models, network traffic can be controled, and patterns of cyberattacks can be detected. This application is essential for saving information systems, not allowing breaches, and deceasing cybersecurity risks.

Fraud Detection in Loyalty Programs
AI may detect frauds and abuses in loyalty programs by investigate patterns and user behavior. Applying AI models, unreal patterns in loyalty program emplying can be identified and related frauds can be reduced. This application is employed in industries such as retail, hospitality, service providers, and other loyalty program-based sectors.

Risk Analysis and Hazard Prediction
For That AI provide identification and prediction risks and hazards associated with multiple activities accross data analysis and prediction algorithms. In case of using AI algorithms, patterns in data Can be observed, helping dor detection risks and hazards. This application is important in industries like insurance, finance, e commerce, and other related fields Harayama et al. (2021).

It is essential to say that AI is only a tool, and for providing accurate and confident results, quality data and a cohesion of suitable algorithms and models are essentiel. Also, final decision-making note the responsibility of humans, with AI acting as a supporting tool Harayama et al. (2021).

AI in Financial Robotics

A crucial role of AI finding in financial robotics. Financial robotics is the employing of robots and automated systems in financial tasks .in this case are some applications of AI in financial robotics Jrad (2023).

Investment Management: AI become effective in investment decision-making processes. For this uses of machine learning algorithms and data analysis, financial robots permitted to investigate market patterns and improve investment decisions refering on available information and market conditions.

Automated Trading: AI in financial robotics leads to perform trading transactions more automate. Or applying AI algorithms and neural networks, financial robots can analyse market patterns and automatically become buy and sell decisions. So, this application provide an increase in the efficiency and quick of transactions, minimaze errors, and enhance financial outcomes.

Market Prediction: AI for prediction market trends and financial prices. By combining and Examining market data and Enforcing predictive algorithms, financial robots Can manage predictive analyses and Produce investment decisions Derived from these forecasts.

Risk Management: Artificial intelligence in financial robotics give assistance in risk management. By examing risk-related data and applying AI algorithms, financial robots can analyse investment risks and generate decisions that decrease financial risks.

Customer Service: AI has a positive impact on customer services in finance. Financial robots provide an automate respond to customer inquiries and requests applying AI algorithms and natural language processing, investigate financial data, and offer support to customers. This application can elevate the speed and quality of customer service, decrease costs, and enhance customer satisfaction Hisan & Amri. (2022).

AI in Banking and Taxation

AI plays a significant role in banking and taxation. Below are some applications of AI in these fields: Kureljusic & Karger. (2023)

Data Analysis and Prediction: The importance of AI appears in analyzing financial and banking data. By employing machine learning algorithms and data analysis techniques, many patterns in financial and banking data can be given, producing more accurate predictions in regard of market behavior, risks, and other financial and banking factors.

Intelligent Decision-Making: AI become essential in decision-making processes in banking and taxation.

Data Analysis and Prediction: Applying AI algorithms, decisions in case of loan granting, capital management, investment policies, and other financial and banking decisions may be increased.

Intelligent Decision-Making: For that artificial intelligence leads to smarter decisions thanks to analyzing data and considering various factors.

Fraud and Scam Detection: AI give assistance in case of detection of fraudulent tasks and scams in the banking and taxation sectors. So applying machine learning algorithms and neural networks, common patterns in Dubious behavior can be detected, allowing fraudulent activities to be identify. This application produce more security and minimaze financial risks related with fraud in banking and taxation.

Accounting Theory in the Age of AI

To start with the implementation of AI and accounting is not restricted to the application of financial robots however also reflects the advancement of accounting theory over times. Actually, theories in the accounting field are composed into three categories: new management accounting theory, value creation theory, and intelligent management mechanism theory.

New Management Accounting Theory

The aims of Management accounting is to adapt internal users with the managerial information needs for proper business decision-making and improvement of business management Sath (1978). Indeed, The new management accounting theory said that in the context of AI, financial personnel have ample space for survival and development. Also, Financial robots as a result of human inventions, and the inside working mechanism is pre-defined by humans. Another thing, technology changes programmed and mechanical work, and the human accounting mind become irreplaceable. An investigation done in 2021 prided that even though the emergence of robo advisors, 68% of clients have the preference to utilize both human and robo advisors. In the era of AI, the importance of accountants has shifted from traditional financial accounting to using financial data to give informational support for business units. For that it is crucial to change in accounting personnel from "executors" to "managers," using financial robots like essential means to guide work, effectively combining business and financial information, and providing informational consulting services for business sectors. In other way, accountants are needed to be more professional and have comprehensive analytical skills. For that future accountants have to be skilled in both finance and management Harayama et al (2021).

Value Creation Theory

The theory of organizational value creation derives from the theory of modern commodity circulation. Indeed, Marx thought that a valuable modern resource circulation includes money initial processing into goods, and then goods transforming into money.
The aims is to increase value creation by integrating various value activities. In regard of the perspective of investors' rights and interests, the objective of accounting in business operations is to understand valuation and enhance company value. Accounting have to provide support for investors, saving the added value of investor capital, and safeguard value creation and risk control aspects Peng et al. (2023).

Artificial Intelligence in Accounting: Applications and Impact

Overview

Artificial intelligence has becomed like a transformative needing in accounting profession, reshaping processes, increasing efficiency, and finding valuable insights. AI's enable to automate tasks, ameliorate accuracy, and offer data-driven insights modify the accounting landscape, allowing accountants to concentrate on more strategic and value-added work. AI's capability for automating repetitive processes and time-consuming tasks has become a feature of its usinf in accounting. AI algorithms can deal with efficiently data entry, reconciliation, and invoice processing, permmitting accountants to keep attention on more complex and strategic tasks Ahmad et al. (2021). This change empowers accountants to participate in activities that needs higher-level decision-making and expertise, like financial analysis, tax planning, and business advisory services.

AI has an important role in advancing the accuracy of accounting work, reducing errors, and guaranteeing compliance with financial regulations. Also, AI algorithms can check and correct errors in financial statements, tax returns, and other accounting documents, enhancing the overall quality and reliability of financial reporting Bui et al. (2020). This improved accuracy, minimaze the risk of financial misstatements and empower regulatory compliance, assuring adherence to accounting standards and legal requirements.

AI has the capability to examine extensive amounts of data to give accountants with valuable insights that were previously strict or impossible to get manually. AI algorithms may check trends, patterns, also anomalies in financial data, catching profond relation and

connections that traditional methods may have overlooked Goh & Seah. (2022). These insights encourage accountants to generate informed decisions, defiy potential risks and opportunities, and assure strategic path for businesses.
AI has become an indispensable tool in the accounting profession, automating tasks, improving accuracy, and providing data-driven insights. By automating repetitive tasks, AI frees up accountants' time, allowing them to focus on more strategic and value-added work.
AI enable to find and correct errors increases the accuracy of accounting work and assure compliance with financial regulations. Also, AI's ability to examine large datasets offers accountants with valuable insights, helping them to generate informed decisions, check potential risks and opportunities, and delivers strategic guidance to businesses. While AI continues to progress, its role in accounting is become expanded further, guiding innovation and transformation across the industry.

Machine Learning in Accounting:

Machine learning increases accounting processes by enhance efficiency and decision-making. Indeed , it automates transaction classification, checking fraudulent activities based on analysing historical data, and assists predictive analysis for forecasting financial trends and risks.

Also, it helps in expense management, reduces costs by categorizing expenditures, and increses data processing for accurate reporting. Machine learning systems completely adapt to new data, providing accounting processes stay current and effective **Li et al. (2023, pp. 1, 4, 6).**

Machine learning is an important means in finding tax evasion and providing compliance by checking anomalies and potential tax non-compliance. It uses pattern recognition to detect fraudulent transactions by analyz ing large financial datasets.

Deep Learning in Accounting

Financial Forecasting with LSTM Networks.
Long Short-Term Memory (LSTM) networks are important for financial prediction, as they effectively analyze complex time-series data and predict future trends Zaheer et al. (2023, p. 1)
LSTMs escape traditional methods by highlighting long-term dependencies and nonlinear links in historical data. Also, Their implimentation with Convolutional Neural Networks (CNNs) increases forecasting accuracy Zaheer et al. (2023, p. 4)

Anomaly Detection for Fraud and Accounting Information Quality
Deep learning has an essential role for detecting anomalies in financial data. Schreyer et al. (2019) demonstrated that adversarial auto-encoding neural networks can find irregularities in financial data, making them highly valuable for fraud detection.
Another thing, multilayer neural networks is an important tool in evaluating the quality of accounting in formation Wang (2023, p. 609).
Deep learning techniques are also utilized in forecast ing financial crises. Chen and Zhang proposed that artificial neural network models, based on accounting data, provide effective early warnings for financial crises (Chen & Zhang, 2022, p. 1).

Expert Systems:
Expert systems help managers in identifying weaknesses in internal control and facilitate training processes. This is crucial for organizations, as it leads to more reliable financial reporting and improved internal control practices (Changchit & Holsapple, 2004, p. 103).

Benefits, Disadvantages, and Future Outlook of AI in Accounting

Benefits:

- Enhanced accuracy and speed in performing accounting tasks, like data entry, reconciliation, recording, and reporting.
- Cost decrease and human resource optimization through changing repetitive and low-value tasks with automation.
- •Increased security and transparency of financial data by finding fraud and operational errors.
- Creation of innovative and creative opportunities for accountants across the analysis of large data sets and prediction of financial trends (Harayama et al., 2021).

Concrete Applications

AI in accounting provide assistance with tasks that are typically done manually and time-consuming. Some of these tasks contain:
- Data entry and Harmonisation, Invoice matching, Invoice creation and sending, Expense reporting, Price change tracking, Account reconciliation, Transaction sorting , Data recording and reporting Tandiono (2023) By using algorithms that are free of human error and enable of doing tasks around the clock, artificial intelligence may perform these works with high speed and accuracy. This protect time,

costs, and resources for accountants, capable them to Concentrate on more valuable and creative tasks. Furtehermor, artificial intelligence may indeed help to assist in examining and forecasting financial and accounting data. Employing artificial neural networks, AI may uncover patterns and relationships in large, complex datasets and make insights that are essential for decision-making by managers and accountants. Some of these insights include:

- Stock price estimation and trends
- Auditing and examining processes
- •Valuation of securities and other assets
- Credit rating predictions
- Credit agreement
- Cost prediction
- Future prediction and estimation
- Bankruptcy evaluation and loan risk assesment for banks and other financial institutions By giving these insights, AI assist to improve the quality and efficiency of accounting and minimaze risks and problems.

AI in accounting complete to evolve and promote, providing new applications for the field. A best example and case was, Alibaba's facial recognition technology in payment and financial services, and intelligent robot services for customers in the shopping platform, are examples of how AI ischanging and increasing business operations and services for Alibaba.

Disadvantages of AI

AI for accounting can however have some inconvinient. Some of these showed in The significant cost of acquiring and installing AI-based systems.The demand for training managers and accountants to adapt with these systems and update their knowledge. Obstacle in determining responsibility and accountability in situations involving errors or fraud in AI systems. Potential devaluation and decreased status of human accountants in the job market. The potential loss of human connection and confidence between accountants and clients.

Future and Outlook and Challenges:

AI is a technology leads to accounting processes improvement, minimaze errors, and enhance efficiency. AI may analyze financial data, making audits, issue invoices, report expenses, and recommend financial decisions. AI permitted for human accountants to keep intention more on strategic, creative, and value-added activities. But, AI also has its own challenges. Some of the challenges of AI for accounting appear in:

The highly costs of AI-based systems
The important needing for training and promoting the skills of accountants and financial managers to adapt with these systems
Difficulty in providing responsibility and ethics for decisions made by AI

The possibility for errors, flaws, frauds, or security threats in AI systems
The necessity to comply with accounting regulations and protect privacy These challenges need attention and problem-solving from accountants, managers, technology experts, and regulatory organizations. Also, AI can be both an opportunity and a threat for accounting. moreover, accountants have to stay up to date with technological changes and leverage AI's benefits to advance their careers.
AI in accounting is one of the innovative and widely used technologies that can create significant transformations in this field. By employing algorithms and artificial neural networks, AI help to collect, process, analyze, and estimate financial and accounting data, assisting accountants and financial managers to provide better decisions. AI permitted to automate accounting tasks and decrease errors, costs, and time. Also, AI in accounting provide many benefits, especially :

- Increased speed and accuracy in performing accounting tasks
- Reduced errors and accounting mistakes
- enhanced efficiency and productivity of accountants
- More accurate and up-to-date financial reporting
- Genuis analysis of financial and accounting data
- high level of management of financial risks and issues
- Provision of better accounting services tailored to client needs

In other hand , AI in accounting show many challenges thathave to be noted. Cases of these challenges manifested by:

- The necessity for changes in organizational culture and structure to adapt to AI
- The need for training and upgrading the skills of accountants to work with AI
- The essential needs to adapt and awarness to ethical and legal principles in the use of AI

- The necessity to provide security and privacy of financial and accounting data
- The necessity to take solution for complex and creative issues that AI cannot address, also AI in accounting is still stimulating and evolving, creating new opportunities and possibilities for the field. So, AI know like an assistant to promote intelligent accounting systems that may provide interaction with clients and accountants, identify their needs and questions, giving appropriate responses, and providing customized and optimized accounting services.

AI and Corporate Social Responsibility (CSR)

AI has become an important tool to optimize CSR practice in organizations, permitting firms to optimize ESG issues faster and accurately. Indeed, Technologies guided by AI allow organizations to automate CSR reports, evaluate their sustainability, and address the stakeholders' sentiments in the process (Aleem & Hassan, 2025, p. 12)

AI's Role in Environmental Responsibility:

AI is highly being employed for resource-efficient, usable purposes, and it is of great focusing in energy utilization, carbon footprint reducing and supply chain management Teoh et al.(2019). AI technologies provide the possibility to predict different environmental threats that influence the company, control emissions, and optimize the work of waste management systems Diaz & Nguyen. (2023). Or, smart grids using AI technologies permitted compagnies to monitor energy use effeciency focused on data collected in real-time Fukuyama & Tan. (2022). Indeed, with climate risk assessment, applications advanced through AI assist the organization finding risks relevant to climate change and take principles to increase climate sustainability Pfajfar et al. (2022).

AI's Role in Social Responsibility

AI has changed the way social responsibilities are deploying within organizations in regard of policies on diversity, the ethics of recruitment and selection, and the effectiveness of charitable causes, among other things Pai & Chandra. (2022).

Another side Automating recruitment processes that consider applicants' qualifications provides no discrimination in employment opportunities (Dhote et al. 2024). Moreover, Thanks to AI, it is probable to measure the level of employee satisfaction and make the required changes Randerson (2022).

In the side of strategic philanthropy, AI assist the firm to find the potential areas to invest in to align the firm's charitable giving by showing the higher community necessity and allocating the resources Liu et al. (2021). Also, sentiment investigation may determine people's attitudes across a specific firm's CSR efforts, helping the firm in fine-tuning its sustainability initiatives Lin (2021).

AI's Role in Governance:

The company's AI application has provide positive impact to corporate governance, like automated compliance monitoring, detection of financial fraud, and enhanced transparency in ESG reporting Nirino et al. (2022). So, Automated fraud detection solutions process the enormous amount of data in regard of financial transactions to find any unethical attitude and act in business Bao et al. (2022).

Furthermore, they permit firms to provide automated ESG reports, making easy the preparation of quality and standardized sustainability reports as per the set legal requirements ElMadany et al. (2021). For example AI-based governance strategies not only reduce the propotion of work of employees however indeed enhance investors' confidence in corporate governance and ethical accountability Russo et al. (2022).

Link between AI-Driven CSR and Financial Performance

A study havefound that implementation AI into any CSR endeavor have a positive impact such as enhanced efficiency, cost reduction, and better financial outcomes Fukuyama & Tan. (2022). In this case, AI-based approaches help organizations with the best tools to allocate resources, manage the reporting system and increase corporate image,so the stock market credibility and ultimate profitability in the long run, reported by Padilla-Lozano & Collazzo. (2022).

AI and Smarter Financial Reporting

Smarter Reporting

Automation and Accuracy

AI automates routine accounting tasks such as transaction classification, reconciliations, and error detection. According to Brynjolfsson & McAfee. (2017), AI reduces human error and allows accountants to concentrate on interpretation and decision-making. Machine learning algorithms can detect anomalies in financial data, helping prevent fraud and misstatements, thereby improving the reliability of financial reports.

Real-Time and Predictive Reporting

Traditional accounting focuses on historical data. AI introduces real-time reporting and predictive analytics, enabling continuous monitoring and forecasting of financial performance Moll & Yigitbasioglu.(2019). This capability allows managers and stakeholders to make timely, informed decisions, enhancing strategic planning and operational efficiency.

Guidelines for Responsible Accounting in the Age of Automation

Artificial intelligence modify the accounting landscape, leads to increase efficiency, accuracy, and insights. But, using this new terrain needs cautious and responsible approach. This article provide critical guidelines for accountants aiming to leverage AI's potential even though mitigating its risks: Before using any AI tool, accountants have to clearly understand its capabilities and limitations AICPA & IMA. (2021).

Althoug AI excels in these areas, it is not free from mistakes. Accountants have to be prepared to control its outputs, defiy potential errors, and participate when necessary. AI intervention needed to be firmly grounded in ethical principles. In other away Accountants ought to prioritize using AI to increase accuracy, efficiency, also effectiveness Chartered Institute of Management Accountants (2020). AI needn't to replace accountants or automate tasks in a way that affects professional judgment or ethical standards. Effective communication with clients and stakeholders about AI's role in their services provides trust and enhance transparency.

AI systems needs large amounts of data, especially potentially sensitive information. Accountants are responsible for saving this data from unauthorized access, misuse, or disclosure AICPA (2023).

Adoption robust cybersecurity measures, must comply with data privacy regulations, and reducing data collection is important to facility a secure and ethical AI ecosystem. Artificial intelligence implementation accounting is not without its risks. The prejudice inherent in training data may produce to discriminatory outcomes. Or, system errors and malfunctioning algorithms can damage data integrity. Cyberattacks can threat sensitive data. Accountants have to get a proactive risk management plan to check, evaluation, and minimaze these risks, guaranting AI's responsible and secure application.

The real value of AI is in increasing, not substituting, the skills and expertise of accountants. Accountants adopting AI can become more efficient, analyze data deeply, and provide more informed insights.

Nevertheless, a powerful comprehension of accounting principles and practices become important for navigating the complexities of AI-powered tasks. Bias may gradually appear into AI systems through the data they are trained on. Or, accountants have to be attentive in detecting and reducing possible bias in their AI tools. Using many data sources, testing for bias throughout the AI lifecycle, and actively seeking feedback from various stakeholders are essential to provide unbiased and ethical AI adoption.

Using robust cybersecurity measures, with data privacy regulations, also reducing data collection is essential to provide a secure and ethical AI ecosystem. AI's implementation into accounting is not without disadvantage. The erros inherent in training data may result discriminatory outcomes. Indeed, System mistakes and malfunctioning algorithms can affect data integrity. Cyberattacks can expose sensitive data. Accountants need a proactive risk management plan to identify, assess, and mitigate these risks, ensuring AI's responsible and secure application.

The role of AI's is assistance not replace, the skills and expertise of accountants. Accountants integrating AI can become more efficient, analyze data deeply, and create more informed insights.

But, a powerful comprehension of accounting principles and practices still important for guiding the complexities of AI-powered tasks. Error can gradually emerge into AI systems accross the data they are trained on. Accountants have to be attentive in detecting and mitigating potential bias in their AI tools. Using variety of data sources, testing for bias throughout the AI lifecycle, and actively seeking feedback from various stakeholders are crucial to ensuring unbiased and ethical AI implementation.

The transprence in communication with cutomers and stakeholders is essential when using AI. Accountants ought to clearly explain how AI is used in their services, its benefits and constrains, and the measures to mitigate risks. Planing a discussion around AI provides trust and ensures its responsible application within the accounting profession. By respecting these guidelines, accountants can harness AI's power to navigate the industry's evolving landscape. By prioritizing ethics, responsible data practices, and continuous learning, they can ensure that AI is a valuable tool for enhancing accuracy, efficiency, and trust in the financial reporting ecosystem.

AI-Empowered Accounting Information Systems (AIS)

The Role of AIS in Fraud Detection

Chang et al. (2022) said that the Accounting Information System, has an important role in fraud detection. Indeed, AIS collects, processes and stores financial data, which can be examined to check unusual patterns and anomalies that may indicate fraudulent activities Liu et al. (2021).

By giving access to real-time financial information, Accounting Information System could also assist to detect fraud in its early stages, potentially reducing its effect on an organisation. Also, accounting Information System leads to the implementation of internal controls and auditing procedures, further permitted a company's ability to find and prevent fraud Al Omoush et al. (2019).

9.2. Dimensions of AI in AIS

It exists different dimensions of AI in the field of AIS; some dimensions may be correlated to certain AIS practices; but, in auditing and fraud detection, AI-empowered AIS components like Han et al. (2020), Casey et al. (2021) and Alam et al. (2022): Data gathering: artificial intelligence system could combine large amounts of data from different sources, such as financial records, transaction logs and social media feeds.

- Detection: AI may utilized to check potential instances of fraud as they occur by controling transaction data in real-time and Monitoring for suspicious activity.
- Prevention: AI assist to prevent fraud by definy potential weaknesses in an organisation's internal controls and making recommendations for improvement.
- Investigation: AI help with fraud investigations by examining large amounts of data quickly and accurately, permitted investigators to definy and track down potential suspects.
- Overall, AI enable to increase the effectiveness and efficiency of auditing and fraud detection by automating routine tasks, investigating large amounts of data quickly and providing valuable insights into potential areas of risk Königstorfer and Thalmann. (2020).
- Data analysis: AI could investigate this data utilizing machine learning algorithms, Detecting patterns, trends and anomalies that possibility evidence of fraud.
- Risk assessment: Artificial intelligence Is permitted to use predictive analytics to evaluate the risk of fraud for a particular organisation or set of transactions, facilitating auditors and investigators to concentrate their efforts where they are most needed.

Auditing and Fraud Detection

Chen et al. (2019) reported that, auditing and fraud detection are essential functions within an Accounting Information System. Auditing contain evaluating the effectiveness of internal controls, checking of risk, and Assuring the accuracy and reliability of financial information. Auditors can utilize many techniques, like data analytics and sampling, to examine heigh volumes of data and find anomalies or irregularities.

According to Baghdasaryan et al. (2022) fraud detection is identify and investigate fraudulent activity within AIS. Fraud can happen accross different means, like manipulating financial record, find theft, check false invoices, and unauthorised access to sensitive information.

To find fraud, Accounting information system can use monitoring tools like data mining, forecast analytics, and exception reporting Aslam et al.(2022).

These means assist to identify suspicious patterns or activities that may indicate fraudulent behaviour. It is nécessaire to have effective auditing and fraud detection mechanisms in place to provide the integrity of an organisation's financial transactions and the protection of its assets Ashraf et al. (2020).

Applicability of Natural Language Processing (NLP)

Pourhabibi et al. (2020) definy NLP like process that leads a computer to understand, interpret and communicate in human language. Or, it is a subset of AI that concentrate specifically on the relationship between computers and human language, permitted computers to process, interpret and generate natural language. AI has reached a stage where it is used in many aspects of life.

El-Halaby et al. (2021) reported that there are many era in which NLP isused in AI-reiforced AIS in accounting and finance. El-Halaby noted that NLP is employed to examine financial reports and extract text data like transaction descriptions, comments and email messages to identify instances of unusual activities or suspicious patterns that could be indicative of fraudulent activities.

According to Kang et al. (2020) NLP tools support accountants transform unstructured data, like emails and social media posts, into structured data. Hazlehurst et al. (2019) noted that AI reinforce AIS with NLP could assist accountants automate task allocation processes, also controling compliance with financial regulations and legal requirements. NLP algorithms are employed to scan textual data for specific trigger phrases and keywords, flagging potential non-compliance.

Bao et al. (2022) said that NLP is an important component of AI-empowered AIS in the world of accounting and finance. It has a crucial role in streamlining the auditing process, enhancing the accuracy and security of financial activities and improving decision-making.

Impact of AI on Accountants and Case Studies

AI system modify heavy accounting basic work to more valuable professional judgment refers to large data analysis and data mining Zehong and Zheng. (2018). Indeed, Its high employing gradually minimaze the demand for accounting personnel. so, accounting personnel face the crisis of elimination. Thus, Robot performs redundant and repeatable data entry tasks, and interacts with such automation technology. Technology only may not lead future; just assist compagnies make decisions built on their individualistic experience. The crucial is to bloom new technology to ameliorate it on aeonic base.

Impact on Professional Roles

Financial Accounting: Shift towards judgment and interpretation

Accounting personnel get data from introspective cognizance. Analysis ad infinitum proves the authenticity Deriving from past transactions or phenomenon. Robots Ordered numerical values to express liabilities. Accounting prediction gives judgment on events refering fresh intel to face transition which accounting standards stipulate for adopting prospective law. Financial accounting awaits specialists for their skills and experiencein line with the accounting standards and relevant laws and regulations to furnish fair information to users.

Management Accounting

Management accounting refashions prediction to provide the realities of predicted results. Most management deriving from participation of management accounting, budget, decision-making, final evaluation of managers performance. AI simulates future environment to help management accounting accomplish tasks; however, it may not transformation management accounting for decision-making.

FinTech and Case Studies

Fintech innovations leads to change financial transparency by increasing the efficiency and accessibility of financial information. Indeed, The employing of real technologies like artificial intelligence and machine learning in fintech permitted to produce real-time data processing also enhance accuracy in financial reporting Arner, Barberis, & Buckley (2016) Or, this higher kind of technologie provides an essential role for the automation of financial transactions and auditing processes, minimazing human error and enhance the reliability of financial reports (Gai, Qiu, & Sun, 2018). In other side, fintech solutions for example digital wallets and blockchain-based systems giving a clear and immutable record of transactions, that improving the transparency of financial activities Zetzsche et al. (2020).

The use of fintech innovations may create challenges related to data security and privacy. Increased reliance on digital platforms raises the risk of data breaches and cyber-attacks, which can compromise financial transparency Kamuangu (2024). Also, important cybersecurity policies and regulatory frameworks are important to reduce these threats and guartee that fintech innovationshave a postive impact to financial transparency (Chen, Li, & Wang, 2019).Furthermore, faster advanced of technological modify the necessitates continuous updates to financial regulations and practices to address emerging Puschmann (2017).

Evolution and Growth of AI in FinTech

The artificial intelligence in finance start really with rule-based systems for fraud detection and credit scoring in the 1990s. Or, this systems were restricted by their dependence on predefined rules and static datasets. Indeed, the exact changing start in the 2010s, guiding by advances in machine learning, the presencde of big data, and exponential enhancec in computational power. Also, Startups and creation of financial institutions alike start employing AI to become competitive in a digital-first economy.

AI-using chatbots replaced call centers; predictive models become outperforming human analysts in market forecasting; and algorithmic trading became the norm rather than the exception. In regard of 2023 report by the World Economic Forum, over 85% of global financial institutions had employed some form of AI into their operations, with projections suggesting continued growth across all domains of financial services. The COVID-19 pandemic further make faster this trend. So, Lockdowns, remote work, and the demand for contactless services made digital transformation an essential rather than a choice. For that, AI tools assist institutions quickly update by streamlining workflows, managing risks in volatile markets, and offering customized digital experiences.

Key Drivers of Adoption (data, customer expectations, efficiency)

Several factors have fueled the widespread adoption of AI in FinTech:

- Data Availability: In this case the digitalization of financial transactions has leaded to vast amounts of structured and unstructured data, that provide as raw material for training AI models.

- Customer Expectations: In thsi day consumers need quick, personalized, and seamless financial services. For that, AI enables real-time responses, tailored product recommendations, and 24/7 support.
- Operational Efficiency: AI minimaze the demand for manual intervention in routine tasks, significantly lowering operational costs and improving accuracy.

Automation and Operational Optimization.

One of the important benefits of AI in FinTech is the automation of routine processes, that enhances operational efficiency. For example , Tasks like data entry, document verification, and transaction monitoring, that doing by human employees, can now be doing quickly and with greater accuracy by intelligent systems. For example, robotic process automation (RPA) tools guded by machine learning, can check information from invoices, validate payment instructions, and even reconcile accounts with minimal oversight.

Credit Scoring and Risk Management.

AI has changing credit assessment by giving alternative data sources and advanced modeling techniques into the decision-making process. So, Traditional credit scoring systems based on historical repayment behavior and credit bureau data, which often exclude underbanked populations. However AI, can analyze a much broader spectrum of indicators, like mobile phone usage, social media activity, and transaction patterns, to build more inclusive and accurate credit profiles. In risk management, machine learning models continue to learn from new data, leading real-time detection of anomalies, early warning systems for loan defaults, and dynamic risk-adjusted pricing. AI also facilitates stress testing and scenario analysis, allowing financial institutions to model various market conditions and assess potential vulnerabilities with greater precision than before.

Customer Experience and Personalization.

Customer-centricity become a core competitive advantage in FinTech, and AI is a key provider of that transformation. Natural language processing (NLP) and sentiment analysis permitted chatbots and virtual assistants to engage users in human-like conversations, giving instant support without long waiting times. This may handle everything from account inquiries and transaction tracking to investment advice and fraud reporting. Thanks to support, AI systems analyze user behavior to provide personalized financial recommendations. Indeed, personal finance apps reiforced by AI can monitor spending patterns, suggest savings plans, and flag unusual expenses. Investment platforms can offer customized portfolio suggestions based on a users financial goals, risk appetite, and market trends. The result is a more tailored and engaging user experience that fosters customer loyalty and trust.

Fraud Detection and Regulatory Compliance

Financial fraud exists in complexity, leading traditional rule-based detection methods highly ineffective. So, AI faces this challenge by using behavioral analytics and anomaly detection to check suspicious activities in real time. By continue by learning from past fraud cases, AI systems may check between legitimate and fraudulent transactions with a high degree of accuracy. In the real of compliance, AI streamlines regulatory reporting, control transactions for signs of money laundering, and assists in know-your-customer (KYC) verifications.

Natural language understanding algorithms may examine high volumes of regulatory documents to ensure that financial institutions remain aligned with constantly changing rules and standards. Overall, these applications not only minimaze the risk of non-compliance and financial crime but also save institutions significant time and resources that would otherwise be spent on manual oversight.

Ethical Concerns in FinTech

Algorithmic bias, data privacy, accountability and legal implications Artificial intelligence become embedded in financial decision-making, concerns around algorithmic bias. AI models trained based on historical data may amplify existing social and economic inequalities. So that, when past lending data reflect systemic biases against certain demographics, an AI-driven credit scoring system may continue to deny loans to applicants from those groups, even though are financially qualified.

Also, the black-box nature of many machine learning algorithms permitted it difficult to find and correct these biases. A biased output may seems neutral on the surface but derive from data inputs that reflect discriminatory practices. So , this leads to increases significant concerns around fairness, especially if decisions impact individuals' access to important financial services like credit, insurance, or investment opportunities.

AI systems are powered by huge collections of personal data—transaction histories, social media behavior, biometric identifiers, and more. Enen though these datasets provides deeper insights and better personalization, may indeed increase serious questions about user consent, data ownership, and privacy.

Many customers are unaware of how their data is collected, processed, and shared across platforms. The employing of AI can further obscure these dynamics, as algorithms may estimate confidential data (e.g. ethnicity, political leanings) though such information was not explicitly provided.

The ethical problem is compounded when users cannot control the extent of data profiling. Without robust consent mechanisms and clear data usage policies, the deployment of AI in FinTech risks violating fundamental privacy rights. Determining accountability in AI-driven systems is another complex ethical challenge. If an automated financial tool makes a harmful or incorrect decision, who is responsible: the developer, the institution, or the algorithm itself?

Current legal frameworks often struggle to assign liability in such cases, especially when human oversight is minimal or absent. Additionally, regulatory jurisdictions differ widely in their approaches to AI governance. While some countries enforce strict guidelines on algorithmic accountability and consumer protection, others lag, creating uneven standards that can be exploited by unethical actors. A lack of harmonized global regulation makes it easier for firms to avoid ethical responsibilities, undermining public trust in both AI and the financial system.

Practical Case Examples (KPMG, PwC, EY, Deloitte)

For this irt is important to mention Companies' applications of artificial intelligence in accounting and auditing :

- KPMG: KPMG employ AI-powered cognitive automation to ameliorate tax compliance and auditing processes, leveraging big data analytics to check risks and gives real-time insights. Indeed, The compgny uses machine learning to process large data sets and increase the quality of financial insights.For example KPMG Clara, an AI and machine learning-supported cloud-based auditing platform, explores data to find risks and make esay collaboration (KPMG, Clara).
- Also, KPMG has created automated audit processes, like cash audits, employing machine learning (Üçoğlu, 2020, p. 4).
- PricewaterhouseCoopers (PwC): PwC has increased its fraud detection systems in financial audits by using machine learning algorithms, enhancing data analysis precision and financial reporting (Üçoğlu, 2020, pp. 3, 4) .
- The compagny also utilize many AI-powered tools to increaseits auditing processes: PwC Aura, a cloud-based ERP platform for risk assessments and digital collaboration (PwC, Aura)
- GL.ai, an AI bot that analyzes data to detect fraud and errors (PwC, Harnessing the power of AI to transform the detection of fraud and error)
- Cash.ai, which automates cash audits and enhances efficiency (PwC, Harnessing AI to pioneer new approaches to the audit)
- Ernst & Young (EY): Ernst & Young (EY) has employed Natural Language Processing (NLP) into its financial statement analysis to increase efficiency.
- EY Canvas : an online platform, correlate audit professionals with clients on a unified network, providing efficient audit management across real-time data updates (EY, EY Canvas).
- EY Helix : an AI-powered analytics platform, improves audit quality by giving detailed insights and providing risk evaluation via high data analysis, permitted auditors to obtain more precise results (EY, EY Helix).
- EY Blockchain Analyzer : audits blockchain transactions and provides real-time business applications to improve transparency, making it particularly useful for financial reporting and tax calculations (EY Blockchain Analyzer: Explorer & Visualizer).
- EY uses machine learning to investigate unstructured data, assisting checking fraud risks and improve accuracy (Üçoğlu, 2020, p. 3).
- Deloitte : Deloitte provides integrity in supply chain auditing by blockchain technology and leverages machine learning to automate data analysis and risk assessments.
- Argus : Both combine machine learning and NLP, investigate financial documents to check inconsistencies, increasing audit accuracy (Kokina & Davenport, 2017, p. 6).
- Cortex, is an AI platform combining intelligent automation and machine learning, make easy decision-making and increases efficiency across various areas (Deloitte, CortexAI).
- BrainSpace : a machine learning tool, analyzes unstructured data to support auditors in their work.
- Omnia DNAV : employs digital technologies to improve audit quality, and Deloitte employs predictive analytics to identify financial risks (Üçoğlu, 2020, pp. 4–5).

Applications of Blockchain in Accounting and Auditing

Transforming Audit Processes (traceability, immutability).

The most important applications of Blockchain in accounting is the transformation of impact on audit processes. Indeed, Blockchain gives a tamper-resistant and transparent ledger, providing traceability and immutability of financial transactions.

For example in traditional audits, reconciling huge datasets and enhance the accuracy of financial record can be time-consuming and prone to errors. Or, Blockchain's immutable audit trails streamline this process, allowing auditors to check every transaction regarding to its origin.

In the meantime the cryptographic dependence between blocks take place that once a transaction is recorded, it cannot be altered without detection. For that, this traceability increases the reliability of audit processes, ensuring a comprehensive and unforgeable record that auditors can rely on with confidence (Abrahams et al., 2024; Victor & Great, 2021)

Real-Time Auditing and Continuous Assurance.

Also, Blockchain make easier real-time monitoring and continuous assurance in audit processes. Indeed, traditional audits are often retrospective, concentrationg on historical data. Thanks to Blockchain, auditors can access a continuously updated and transparent ledger. As a result, this real-time visibility into financial transactions permitts to auditors to controle activities as they occur, minimaze the lag between data creation and audit verification.

Higher assurance can be feasible as auditors may assess the state of the ledger at any given moment, increasing the capacity to check anomalies promptly and ensuring a proactive approach to risk management.

Blockchain significantly streamlines financial reporting processes by automating complex tasks and improving the efficiency of data verification Tyagi (2020).

Streamlining Financial Reporting (smart contracts)

Furthermore, the employ of smart contracts, fixed rules and conditions can be encoded, automating routine reporting activities. Also, invoicing, reconciliations, and intercompany transactions can be executed automatically, reducing the need for manual intervention. As a result, decreases the risk of errors and accelerates the overall financial reporting timeline.

Streamlining these processes have impact that financial reports are produced with greater accuracy and efficiency, permitting organizations to use resources more strategically Kunduru (2023). The decentralized nature of Blockchain, coupled with smart contracts, enhances the accuracy and timeliness of financial reporting.

Regulatory Reporting (transparency, standardization).

Otherwise, the automation of financial reporting processes also expedites the timeline for generating reports, giving the ability of organizations togives stakeholders with up-to-date and accurate financial information and this leads to accuracy and timeliness so that provides more informed decision-making process (Duggineni, 2023; Johnson et al., 2023).

Blockchain enable for regulatory reporting, providing a transparent and standardized approach to meeting reporting obligations Bakarich (2022). Regulatory authorities often need for organizations to submit extensive documentation and reports.
Blockchain also capable to offer a single, verifiable version of truth leads that the data submitted is accurate and untampered. The decentralized nature of Blockchain minimaze the reliance on intermediaries in the regulatory reporting process, minimizing the potential for errors and enhancing the overall reliability of reported information.

Challenges of Implementing AI in Accounting and Auditing

Technical Challenges

The integration of AI in auditing provides several technical challenges, especially data complexity, integration with existing systems, consistency, scalability, and the need for skilled auditors (Peng et al., 2023, pp. 4–10). Deep learning algorithms may presents results difficult for human auditors to interpret, requiring a balance between AI's power and professional judgment (Grissa & Abaoub, 2024, pp. 1, 3).

The quick evolution of AI needs continuous learning and adaptation from auditors (Iwuanyanwu et al., 2023, p. 63), even though small businesses have challenges such as investment costs and skill development (Ríkharðsson et al., 2022, pp. 323, 333, 334). Data security is also a major concern, demands robust security procedure to interdict data breaches (Fedyk et al., 2022, p. 939).

Economic Challenges

The economic challenges of AI in accounting and auditing contain high costs, significant investments, and difficulties in measuring ROI, as benefits are often operational rather than immediate financial gains (Veselovsky et al., 2021, p. 294)

AI's effect on the workforce may lead to job displacement, specificly for low-skilled workers, needing retraining and upskilling to mitigate economic consequences (Fedyk et al., 2022, p. 941).

Methodology

This study is based on a qualitative and theoretical analysis of existing academic literature in accounting, artificial intelligence, and corporate governance. The aim is not to test hypotheses but to synthesize existing knowledge and develop a conceptual understanding of AI's impact on accounting and finance.

14.Discussion

The growing implementation of artificial intelligence in accounting represents a structural transformation rather than a simple technological upgrade. Refers to the reviewed literature, AI systems are highly mention like as decision-support tools that increase accuracy, efficiency, and transparency across accounting functions, especially in financial reporting, corporate social responsibility (CSR), security, and ethical governance.

In regard of **financial reporting**, AI have a crucial role to smarter reporting by automating routine tasks, detecting anomalies, and improving real-time data processing. This shift permitted to accountants to focus more on judgment, interpretation, and strategic analysis. However, the literature also highlights a potential risk of over-reliance on automated outputs, which may reduce professional skepticism if human oversight is weakened. Therefore, AI should be viewed as a complement to professional expertise rather than a substitute.

Focusing in **CSR**, AI has an essential role in processing non-financial and unstructured data, such as sustainability reports, environmental indicators, and social impact metrics. This capability supports more consistent and comparable CSR disclosures. Nevertheless, ethical concerns arise when algorithms reflect biases embedded in training data or when CSR reporting becomes more about compliance than genuine accountability. Transparency in AI models and data sources is therefore critical to maintaining the credibility of CSR information. From a **security perspective**, AI reinforce accounting systems accross fraud detection, anomaly identification, and continuous monitoring. The literature consistently focuses on AI's effectiveness in checking suspicious activities that traditional rule-based systems may overlook. In the mean time, enhanced reliance on AI introduces new vulnerabilities, including cybersecurity risks, model manipulation, and data privacy concerns. This creates a paradox where AI increases security while simultaneously expanding the attack surface.

Ethical considerations cut across all these dimensions. Issues such as algorithmic bias, lack of explainability, accountability for AI-driven decisions, and the displacement of human labor remain central to the debate. The discussion suggests that ethical AI in accounting requires not only technical solutions but also robust governance frameworks, professional standards, and regulatory oversight. Accountants, regulators, and system designers must share responsibility for ensuring that AI systems align with societal values and professional ethics. Overall, the discussion underscores that the benefits of AI in accounting are substantial but conditional. The effectiveness of AI depends on human supervision, ethical design, and institutional safeguards. Without these, the promise of smarter accounting may be undermined by risks to trust, transparency, and accountability.

Conclusion

This article aims to discuss the role of artificial intelligence in accounting through a theoretical aspect, highlight on smarter reporting, CSR, security, and ethical implications. The literature mention that AI has reached a level of maturity that allows it to significantly reshape accounting practices by improving efficiency, enhancing data quality, and enabling more advanced forms of analysis.

AI-driven systems support more intelligent financial reporting, make easy the processing of complex CSR information, and reinforce fraud detection and security mechanisms. But, these advantages are related by many challenges concerning ethics, governance, and

professional responsibility. The absence of transparency and explainability in some AI models increases attention regarding accountability, particularly in high-stakes financial and social reporting contexts.

The main contribution of this article lies in highlighting that AI adoption in accounting should not be assessed solely on technical performance. Ethical considerations, regulatory alignment, and human judgment are essential components of sustainable AI integration. Rather than replacing accountants, AI redefines their role, shifting it toward interpretation, oversight, and ethical decision-making.
Future analyses have to extend this theoretical discussion by examine regulatory frameworks, comparative international approaches, or empirical assessments of AI governance in accounting organizations. For practitioners and policymakers, the findings report that responsible AI adoption requires a balanced approach that combines technological innovation with ethical safeguards and professional standards.

In conclusion, AI enable to make accounting smarter, more transparent, and more secure, however when its deployment is guided by strong ethical principles and informed human judgment.

Bibliographie

1. Ablameyko, S. V., & Ablameyko, M. S. (2021). Artificial intelligence from an interdisciplinary perspective: Philosophical and legal aspects. ΦΗ, 58, 57–70. https://doi.org/10.30727/0235-1188-2021-64-5-57-70
2. Abrahams, T. O., Ewuga, S. K., Kaggwa, S., Uwaoma, P. U., Hassan, A. O., & Dawodu, S. O. (2024). Mastering compliance: A comprehensive review of regulatory frameworks in accounting and cybersecurity. Computer Science & IT Research Journal, 5(1), 120–140.
3. Afshar, M., Phillips, A., Karnik, N., Mueller, J., To, D., Gonzalez, R., & Dligach, D. (2019). Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: Development and internal validation. Journal of the American Medical Informatics Association, 26(3), 254–261.
4. Ahmad, S., & Higgins, S. (2021). The impact of artificial intelligence on the accounting profession. Journal of Accounting and Management Information Systems, 22(3), 354–372.
5. Alam, M. K., Ahmad, A. U., & Muneeza, A. (2022). External sharīʻah audit and review committee vis-à-vis sharīʻah compliance quality and accountability: A case of Islamic banks in Bangladesh. Journal of Public Affairs, 22(1), e2364.
6. Alghafiqi, B., & Munajat, E. (2022). Impact of artificial intelligence technology on accounting profession: Dampak teknologi artificial intelligence pada profesi akuntansi. Fair Value. https://doi.org/10.32670/fairvalue.v5i01.865
7. Al Omoush, K. S., Alqirem, R. M., & Alzboon, S. R. (2019). The role of business intelligence tools in harvesting collective intelligence. In Information Systems Architecture and Technology: Proceedings of the 39th International Conference on Information Systems Architecture and Technology–ISAT2018 (Part III, pp. 159–172). Springer International Publishing.
8. American Institute of Certified Public Accountants & Institute of Management Accountants (AICPA & IMA). (2021). The essential guide to artificial intelligence for the accounting profession. https://www.aicpa-cima.com/cpe-learning/webcast/artificialintelligence-for-accounting-and-finance-professionals-webcast-series
9. American Institute of Certified Public Accountants (AICPA). (2023, June 1). Protecting client information. American Institute of Certified Public Accountants. https://www.thetaxadviser.com/issues/2017/may/safeguardingconfidential-clientinformation.html
10. Arner, D. W., Barberis, J., & Buckley, R. P. (2016). The evolution of fintech: A new post-crisis paradigm? Georgetown Journal of International Law, 47(4), 1271–1319.
11. Ashraf, M., Michas, P. N., & Russomanno, D. (2020). The impact of audit committee information technology expertise on the reliability and timeliness of financial reporting. The Accounting Review, 95(5), 23–56.
12. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Research in International Business and Finance, 62, 101744.
13. Baghdasaryan, V., Davtyan, H., Sarikyan, A., & Navasardyan, Z. (2022). Improving tax audit efficiency using machine learning: The role of 'taxpayer's network' data in fraud detection. Applied Artificial Intelligence, 36(1), 2012002.
14. Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. In Innovative Technology at the Interface of Finance and Operations (Vol. I, pp. 223–247).
15. Bui, T. Q., & Jaradat, I. (2020). Artificial intelligence: Its implication for the accounting and auditing profession. Current Research in Accounting, 86, 100439.
16. Chartered Institute of Management Accountants (CIMA). (2020). Professional ethics for accountants: A revised code of ethics. https://www.aicpacima.com/resources/landing/ethics
17. Chang, J. W., Yen, N., & Hung, J. C. (2022). Design of a NLP-empowered finance fraud awareness model: The anti-fraud chatbot for fraud detection and fraud classification as an instance. Journal of Ambient Intelligence and Humanized Computing, 13(10), 4663–4679.
18. Casey, A., Davidson, E., Poon, M., Dong, H., Duma, D., Grivas, A., & Alex, B. (2021). A systematic review of natural language processing applied to radiology reports. BMC Medical Informatics and Decision Making, 21(1), 179.
19. Changchit, C., & Holsapple, C. (2004). The development of an expert system for managerial evaluation of internal controls. Intelligent Systems in Accounting, Finance and Management, 12(2), 103–120.
20. Chen, Y. J., Liou, W. C., Chen, Y. M., & Wu, J. H. (2019). Fraud detection for financial statements of business groups. International Journal of Accounting Information Systems, 32, 1–23.
21. Chen, L., Li, H., & Wang, Q. (2019). Financial technology and cybersecurity: A survey. IEEE Access, 7, 68970–68985.
22. Chen, Y and al. (2022). A full population auditing method based on machine learning. Sustainability, 14(24), 17008.
23. Deloitte. (n.d.). Cortex AI platform. https://www2.deloitte.com/us/en/pages/consulting/topics/cortex-ai-platform.html
24. Dhote, S., Saxena, S., Deogaonkar, A., & Vichoray, C. (2024, February). Analyzing the financial performance of corporate social responsibility using machine learning-enabled financial enterprise model. In *International Conference on Innovative Computing and Communication* (pp. 237–252). Singapore: Springer Nature Singapore.
25. Diaz, J. F., & Nguyen, T. T. (2023). Application of grey relational analysis and artificial neural networks on corporate social responsibility (CSR) indices. *Journal of Sustainable Finance & Investment, 13*(3), 1181–1199.

26. El-Halaby, S., Aboul-Dahab, S., & Bin Qoud, N. (2021). A systematic literature review on AAOIFI standards. *Journal of Financial Reporting and Accounting, 19*(2), 133–183.
27. ElMadany, H., Alfonse, M., & Aref, M. (2021). A proposed approach for production in ERP systems using support vector machine algorithm. *International Journal of Intelligent Computing and Information Sciences, 21*(1), 49–58.
28. Ernst & Young. (n.d.). Canvas. https://www.ey.com/en_gl/audit/technology/canvas
29. Ernst & Young. (n.d.). EY Blockchain Analyzer: Explorer & visualizer. https://www.ey.com/en_gl/audit/technology/ey-blockchain-analyzer
30. Ernst & Young. (n.d.). Helix. https://www.ey.com/en_gl/audit/technology/helix
31. Fedyk, A., and al. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938–985.
32. Fukuyama, H., & Tan, Y. (2022). Implementing strategic disposability for performance evaluation: Innovation, stability, profitability and corporate social responsibility in Chinese banking. *European Journal of Operational Research, 296*(2), 652–668.
33. Gai, K., Qiu, M., & Sun, J. (2018). A survey on security and privacy issues in Internet-of-Things. *IEEE Access, 6*, 16137–16157.
34. Goel, M., Tomar, P. K., Vinjamuri, L. P., Reddy, G. S., Al-Taee, M., & Alazzam, M. B. (2023). Using AI for predictive analytics in financial management. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*.
35. Goh, E., Ng, J., & Goh, H. (2020). Privacy-preserving machine learning: Recent advances and challenges. *Machine Learning Research, 21*(1), 1–98.
36. Grissa, I., & Abaoub, E. (2024). Enhancing fraud detection in financial statements with deep learning: An audit perspective. *International Journal for Multidisciplinary Research, 6*(1).
37. Han, J., Huang, Y., Liu, S., & Towey, K. (2020). Artificial intelligence and machine learning for fraud detection in financial statements: A literature review. *Journal of Forensic and Investigative Accounting, 12*(1), 705–722.
38. Harayama, Y., Milano, M., Baldwin, R., Antonin, C., Berg, J., Karvar, A., & Wyckoff, A. (2021). Artificial intelligence and the future of work. In *Reflections on artificial intelligence for humanity* (pp. 53–67). https://doi.org/10.1007/978-3-030-69128-8_4
39. Hazlehurst, B., Green, C. A., Perrin, N. A., Brandes, J., Carrell, D. S., Baer, A., … Coplan, P. M. (2019). Using natural language processing of clinical text to enhance identification of opioid-related overdoses in electronic health record data. *Pharmacoepidemiology and Drug Safety, 28*(8), 1143–1151.
40. Hisan, U. K., & Amri, M. M. (2022). Artificial intelligence for human life: A critical opinion from medical bioethics perspective—Part II. *Journal of Public Health Sciences, 1*(2), 112–130. https://doi.org/10.56741/jphs.v1i02.215
41. Iwuanyanwu, U., and al. (2023). Analyzing the role of artificial intelligence in IT audit: Current practices and future prospects. *Computer Science & IT Research Journal, 4*(2), 54–68.
42. Johnson, D., Pranada, E., Yoo, R., Uwadiunor, E., Ngozichukwu, B., & Djire, A. (2023). Review and perspective on transition metal electrocatalysts toward carbon-neutral energy. *Energy & Fuels, 37*(3), 1545–1576.
43. Jrad, M. (2023). A role of artificial intelligence in the context of economy: Bibliometric analysis and systematic literature review. *International Journal of Membrane Science and Technology, 10*(3), 1563–1586. https://doi.org/10.15379/ijmst.v10i3.1756
44. Kamuangu, P. K. (2024). A review on cybersecurity in fintech: Threats, solutions, and future trends. *Journal of Economics, Finance and Accounting Studies, 6*(1), 47–53.
45. Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics, 7*(2), 139–172.
46. 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.
47. KPMG. (n.d.). KPMG Clara. https://home.kpmg/xx/en/home/services/audit/kpmg-clara.html
48. Kunduru, A. R. (2023). Cloud BPM application (Appian) robotic process automation capabilities. *Asian Journal of Research in Computer Science, 16*(3), 267–280.
49. Kureljusic, M., & Karger, E. (2023). Forecasting in financial accounting with artificial intelligence: A systematic literature review and future research agenda. *Journal of Applied Accounting Research*. https://doi.org/10.1108/JAAR-06-2022-0146
50. Lin, J. (2021). Design of enterprise financial early warning model based on complex embedded system. *Microprocessors and Microsystems, 80*, 103532.
51. Liu, P., Qingqing, W., & Liu, W. (2021). Enterprise human resource management platform based on FPGA and data mining. *Microprocessors and Microsystems, 80*, 103330.
52. Liu, Q., Chiu, V., Muehlmann, B. W., & Baldwin, A. A. (2021). Bringing scholarly data analytics knowledge using emerging technology tools in accounting into classrooms: A bibliometric approach. *Issues in Accounting Education, 36*(4), 153–181.
53. Li, Z. et al. (2023). Research on the application of machine learning in smart finance. Proceedings of the 2nd Interna tional Conference on Public Management, Digital Economy and Internet Technology.
54. Li, Z., and al. (2023). Research on the application of machine learning in smart finance. In *Proceedings of the 2nd International Conference on Public Management, Digital Economy and Internet Technology*.
55. 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.
56. Nirino, N., Ferraris, A., Miglietta, N., & Invernizzi, A. C. (2022). Intellectual capital: The missing link in the corporate social responsibility–financial performance relationship. *Journal of Intellectual Capital, 23*(2), 420–438.
57. Padilla-Lozano, C. P., & Collazzo, P. (2022). Corporate social responsibility, green innovation and competitiveness: Causality in manufacturing. *Competitiveness Review: An International Business Journal, 32*(7), 21–39.
58. Pai, V., & Chandra, S. (2022). Exploring factors influencing organizational adoption of artificial intelligence (AI) in corporate social responsibility (CSR) initiatives. *Pacific Asia Journal of the Association for Information Systems, 14*(5), 4.
59. Peng, Y., [autres auteurs si disponibles] et al. (2023). Riding the waves of artificial intelligence in advancing accounting and its implications for sustainable development goals. *Sustainability, 15*(19), 14165.
60. Pfajfar, G., Shoham, A., Małecka, A., & Zalaznik, M. (2022). Value of corporate social responsibility for multiple stakeholders and social impact: Relationship marketing perspective. *Journal of Business Research, 143*, 46–61.
61. Pourhabibi, T., Ong, K. L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems, 133*, 113303.
62. DOI: https://doi.org/10.1016/j.dss.2020.113303
63. PwC. (n.d.). Discover Aura: The technology platform that powers your financial statement audit. Aura: Audit technology. https://home.pwc.com
64. PwC. (n.d.). Harnessing the power of AI to transform the detection of fraud and error. https://www.pwc.com/gx/en/about/stories-from-acrossthe-world/harnessing-the-power-of-ai-to-transform-the-detection-of-fraud-and-error.html
65. Randerson, K. (2022). Conceptualizing family business social responsibility. *Technological Forecasting and Social Change, 174*, 121225.

66. Ríkharðsson, P., [autres auteurs si disponibles] et al. (2022). Artificial intelligence and auditing in small- and medium-sized firms: Expectations and applications. *AI Magazine, 43*(3), 323–336.
67. Russo, S., Schimperna, F., Lombardi, R., & Ruggiero, P. (2022). Sustainability performance and social media: An explorative analysis. *Meditari Accountancy Research, 30*(4), 1118–1140.
68. Schreyer, M., [autres auteurs si disponibles] et al. (2019). Detection of accounting anomalies in the latent space using adversarial autoencoder neural networks.
69. Teoh, T. T., Heng, Q. K., Chia, J. J., Shie, J. M., Liaw, S. W., Yang, M., & Nguwi, Y. Y. (2019, November). Machine learning-based corporate social responsibility prediction. In *2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)* (pp. 501–505). IEEE.
70. Tyagi, A. K., Aswathy, S. U., & Abraham, A. (2020). Integrating blockchain technology and artificial intelligence: Synergies, perspectives, challenges, and research directions. *Journal of Information Assurance and Security, 15*(5), 1554.
71. Üçoğlu, D. (2020). Current machine learning applications in accounting and auditing. *Pressacademia, 12*(1), 1–7. https://doi.org/10.17261/pressacademia.2020.1337
72. Proceedings of International Scientific and Practical Conference "Russia 2020 - A New Reality: Economy and Society" (ISPCR 2020).
73. Veselovsky, M., and al. (2021). Intellectual governance in the digital economy of Russia. In *Proceedings of the International Scientific and Practical Conference "Russia 2020 – A New Reality: Economy and Society" (ISPCR 2020)*
74. Victor, E., & Great, C. U. (2021). The role of alkaline/alkaline earth metal oxides in CO2 capture: A concise review. *Journal of Energy Research and Reviews, 9*(3), 46–64.
75. Wang, X. (2023). Quality evaluation of accounting information based on multi-layer neural network. *Atlantis Highlights in Intelligent Systems*, 609–615.
76. Zaheer, S et al. (2023). A multi parameter forecasting for stock time series data using LSTM and deep learning model. Mathematics, 11(3), 590.
77. Zaheer, S., and al. (2023). A multi-parameter forecasting for stock time series data using LSTM and deep learning model. *Mathematics, 11*(3), 590.
78. Zehong, L., & Zheng, L. (2018). The impact of artificial intelligence on accounting. In *Proceedings of the 2018 4th International Conference on Social Science and Higher Education (ICSSHE 2018), Sanya, China*
79. Zetzsche, D. A., Arner, D. W., & Buckley, R. P. (2020). Decentralized finance. *Journal of Financial Regulation, 6*(2), 172–203.

THE DIGITAL DOUBLE-EDGED SWORD: PSYCHOLOGICAL IMPACTS OF DIGITALIZATION

LATIFAH EZZI

Abstract

Purpose: *This article examines the psychological impact of digitalization on mental health, moving beyond an instrumental view of technology. It conceptualizes digitalization as a pervasive environmental condition a* ***double-edged sword*** *that actively shapes emotion regulation, identity, and relationships. The central aim is to analyze how this digital environment influences mental health, with specific attention to the mediating roles of attachment dynamics and emotion regulation processes in determining whether engagement proves adaptive or harmful.*

Methodology: *This study employs a structured analytical-thematic review methodology. It synthesizes contemporary psychological and clinical literature from the last decade and a half (spanning 2010 to 2024) to capture the period of accelerated digital integration and evolving research. The analytical process consisted of two stages: first, thematic extraction of findings related to digital behaviors and mental health outcomes; second, a theory-guided synthesis where these themes were interpreted through the dual lenses of attachment theory and emotion regulation theory. This process facilitated the construction of a coherent explanatory model for the psychological effects of digitalization.*

Results and Conclusions: *The synthesis confirms that digitalization's impact on mental health is dual and psychologically mediated, rather than being inherent to the technology itself. The evidence delineates two primary pathways:* ***positive outcomes****, such as perceived social support, are linked to intentional, instrumental use where technology serves as a tool. Conversely,* ***negative outcomes*** *including anxiety, depressive symptoms, sleep disturbance, and loneliness are consistently associated with passive or compensatory use patterns. Critically, these maladaptive patterns are not random but are* ***theoretically explained by core psychological vulnerabilities****. They are primarily driven by insecure attachment styles, which shape relational needs and fears, and deficits in emotion regulation, which increase reliance on digital tools for immediate mood alteration. In these cases, digital engagement functions as an avoidance strategy or a substitute for emotional security, actively mediated by these pre-existing psychological systems.*

Research limitations*: The primary limitation is the article's reliance on a synthetic review rather than the generation of new empirical data, which makes its conclusions inherently interpretive. The theoretical focus, while providing depth, may not encompass all relevant factors (e.g., neurobiological correlates or broader political-economic structures). Furthermore, much of the reviewed evidence is correlational, precluding definitive causal claims.*

Practical and/or Social implications: *The findings necessitate integrating nuanced "digital habit" assessments into standard clinical practice to identify maladaptive patterns. For prevention, digital literacy programs must evolve into "psychological digital literacy," building skills in critical engagement and resilience, especially for youth. Policymakers are urged to consider guidelines that encourage humane technology design. Concurrently, digital tools themselves, such as teletherapy platforms, can be leveraged for support within clear ethical boundaries.*

Keywords: *digitalization; mental health; digital addiction; social isolation; human–technology interaction.*

Introduction

Digitalization has become an inseparable component of contemporary life, shaping how individuals communicate, learn, work, and understand themselves. Its influence extends far beyond the functional use of devices, forming a continuous psychological and social environment that structures daily experience. In this sense, digital technology is no longer external to human life; it actively participates in emotional regulation, identity construction, and relationship dynamics.

The rapid acceleration of digital transformation has made neutrality toward technology increasingly difficult. While digital tools offer efficiency, accessibility, and unprecedented opportunities for connection, their constant presence has introduced new psychological pressures. Rising levels of anxiety, depression, attentional fragmentation, sleep disturbance, and social comparison have prompted growing concern among mental health professionals and researchers. These concerns suggest that digitalization should be examined not merely as a technical development, but as a psychological condition with profound implications for well-being.

This article adopts the metaphor of the digital double-edged sword to capture the ambivalent nature of digitalization. On one edge, technology facilitates communication, creativity, and access to psychological support; on the other, it may contribute to emotional

overload, isolation, dependency, and diminished psychological stability. Importantly, this dual impact does not stem from technology itself, but from how individuals engage with it, the meanings they assign to it, and the contexts in which it is embedded.

Guided by a psychological and social perspective, the central research question of this article is: ***How do different patterns of digital use influence mental health, social relationships, and emotional stability?*** By addressing this question, the article seeks to integrate psychological theory with contemporary digital reality, offering a balanced and clinically relevant understanding of digitalization's impact on mental health

Literature review

The literature conceptualizes digitalization as a multifaceted phenomenon with divergent psychological consequences rather than a uniform effect on mental health. Early correlational studies linking overall screen time to depression and anxiety (Kross et al., 2013; Twenge et al., 2018) have gradually evolved into more nuanced research that examines qualitative patterns of use, motivational drivers, and mediating psychological variables. This shift highlights that digital effects are not monolithic but are filtered through individual vulnerabilities, contextual conditions, and platform-specific affordances. Empirical Patterns of Digital Engagement and Mental Health

Recent empirical evidence clarifies the conditions under which digital engagement functions as either a risk or a protective factor. A consistent finding is the distinction between active and passive use. Longitudinal research suggests that frequent active posting on social media predicts a small but significant increase in psychological distress over time, whereas passive scrolling alone does not demonstrate the same effect (Yu et al., 2024). This pattern indicates that self-presentation, social comparison, and performance pressure embedded in active engagement may carry unique emotional costs.

A dose response relationship is particularly evident among adolescents. Excessive recreational screen time (e.g., six or more hours daily) is robustly associated with lower self-esteem, increased depressive and anxiety symptoms, sleep disturbance, and somatic complaints (Twenge et al., 2018; Viner et al., 2019). Meta-analytic findings further demonstrate a modest bidirectional relationship between loneliness and problematic media use, suggesting a reinforcing cycle in which digital overuse and social disconnection mutually exacerbate one another (Coyne et al., 2020).

Psychological Mechanisms and Manifestations

Problematic digital use is increasingly understood as a form of maladaptive emotion regulation, characterized by loss of control and continued engagement despite negative consequences. Such patterns often serve as short-term strategies for distress avoidance or mood regulation, ultimately weakening intrinsic self-regulatory capacities and correlating with higher levels of depression, anxiety, and reduced well-being (Ahmed et al., 2024). Vulnerability is heightened among individuals with pre-existing emotional difficulties or insecure attachment styles, for whom digital environments may substitute for emotional security rather than support adaptive coping.

The relationship between digital engagement and social connectedness is similarly complex and highly dependent on platform characteristics. Research demonstrates that communication-based applications facilitating reciprocal, small-group interaction (e.g., WhatsApp) can reduce loneliness, whereas platforms emphasizing anonymity, broad comparison, or performative self-presentation (e.g., Reddit, dating apps) are associated with increased social isolation (Matthews et al., 2025). Momentary assessment studies support a displacement effect, whereby digital interaction may temporarily replace richer face-to-face contact, resulting in short-term increases in loneliness following both instant messaging and social media use (Dora et al., 2023). Notably, these associations appear context-sensitive; for example, links between passive consumption and loneliness were evident prior to, but not during, COVID-19 lockdowns, underscoring how environmental constraints shape the psychological meaning of digital engagement (Matthews et al., 2025)

Vulnerability, Protective Factors, and Structural Influences

Risk exposure is not evenly distributed across populations. Heavy smartphone and social media use, particularly among adolescent girls, has been identified as a significant contributor to sleep disruption, attentional difficulties, and behavioral addiction (Ahmed et al., 2024; Twenge et al., 2018). At the same time, protective factors such as physical activity can buffer against loneliness and psychological distress associated with digital use, emphasizing the importance of balanced online-offline lifestyles (Dora et al., 2023).

Beyond individual factors, emerging debates highlight the role of platform responsibility and algorithmic design as critical contextual influences. Algorithm-driven content curation, engagement-maximization strategies, and exposure to idealized or emotionally charged material may intensify comparison processes, compulsive use, and emotional dysregulation. These structural features suggest that

psychological outcomes are shaped not only by user behavior but also by design choices embedded within digital ecosystems, reinforcing the need to view digitalization as an active psychological environment rather than a neutral medium.

Moderating Contexts and Synthesis

Age, culture, and socioeconomic conditions further moderate digital impacts. Adolescents are especially sensitive to identity formation and social comparison pressures, while cultural norms may either strengthen collective belonging or amplify individual performance demands. Socioeconomic disadvantage can compound risk through unequal access to digital literacy, safe online spaces, and protective offline resources.

Overall, the empirical evidence converges on a coherent pattern: **frequent, active, or compulsive digital engagement,** particularly involving habitual posting, addictive use, and excessive screen time, is more consistently associated with negative mental health outcomes, including anxiety, depression, sleep impairment, and loneliness. Passive consumption shows weaker and more context-dependent effects. Crucially, the psychological impact of digitalization is shaped by ***how platforms are designed, how individuals engage with them, and the vulnerabilities they bring into these environments.*** This synthesis strongly supports the article's central thesis that digitalization functions as a dynamic psychological environment whose effects emerge from the interaction between user, context, and digital "climate."

Materials and methods

Research Design

This article employs an analytical-thematic review methodology, designed for theoretical integration and model development rather than hypothesis testing. The aim is to synthesize existing literature into a coherent conceptual framework that explains how digitalization influences mental health through specific psychological mechanisms.

Literature Search and Selection Strategy

A structured search was conducted in major psychological, medical, and social science databases (e.g., PsycINFO, PubMed, Web of Science) for peer-reviewed articles published between 2010 and 2024. Search terms combined keywords related to the environment ("digitalization," "social media environment"), outcomes ("mental health," "anxiety," "loneliness," "well-being"), and mechanisms ("emotion regulation," "attachment," "Fear of Missing Out"). Empirical studies (longitudinal, experimental, cross-sectional) and seminal theoretical papers were prioritized.

Analytical Framework and Synthesis Procedure

The selected literature was analyzed through a two-stage interpretive process guided by pre-identified theoretical lenses:

Thematic Extraction: Findings were coded for recurring themes related to digital behaviors (e.g., passive vs. active use, platform type), mental health outcomes, and proposed explanatory factors.

Theory-Guided Synthesis: These themes were systematically interpreted and integrated using attachment theory (to understand relational motivations and vulnerabilities) and emotion regulation theory (to understand compensatory use and dysregulation). This process yielded the integrative model presented in the findings, where digital effects are framed as mediated by core psychological systems.

Results and findings

The synthesis of contemporary literature, interpreted through the theoretical frameworks of attachment and emotion regulation, yields two primary, interconnected findings. These findings move beyond a simplistic "good vs. bad" narrative to clarify the specific psychological pathways through which digitalization influences mental health.

Finding 1: The Dual Impact is Psychologically Mediated, Not Technologically Determined

The analysis robustly confirms that digitalization's effect is not a direct result of the technology itself, but is critically filtered through an individual's psychological systems. The literature reveals a clear demarcation in outcomes based on the user's underlying motivations and psychological state.

Intentional, Instrumental Use and Positive Outcomes: Positive mental health outcomes are consistently linked to use that is purposeful, controlled, and integrated into a balanced life. This includes leveraging communicative applications like WhatsApp or Signal for deepening existing relationships, utilizing teletherapy platforms for structured support, or accessing online communities for shared identity and reduced stigma. In these cases, digital tools function as an extension of the user's healthy agency and social repertoire, enhancing perceived social support and access to resources (Schäfer et al., 2024).

Compensatory, Passive Use and Negative Outcomes: Conversely, negative outcomes including heightened anxiety, depressive symptoms, sleep disturbance, and attentional fragmentation are strongly associated with digital engagement that serves a compensatory function. This often manifests as passive, endless scrolling or compulsive checking, which are frequently employed as maladaptive emotion-regulation strategies to avoid distress, boredom, or negative self-perception. This pattern is particularly pronounced when technology use becomes a substitute for emotional security or a primary means of validation-seeking. Such use is reinforced by platform design but is fundamentally driven by pre-existing vulnerabilities in emotion regulation or insecure attachment styles, wherein the digital environment becomes a space for emotional avoidance rather than connection (Ahmed et al., 2024; Yu et al., 2024).

Finding 2: Vulnerability is Contextual, Non-Uniform, and Shaped by a Confluence of Factors

The risk for experiencing negative outcomes is not distributed equally across populations. Vulnerability is instead shaped by a dynamic interaction between individual predispositions and broader contextual factors.

Core Psychological Vulnerability: At the individual level, those with insecure attachment styles whether anxious (preoccupied with abandonment) or avoidant (discomfort with intimacy) are significantly more susceptible to maladaptive engagement. For the anxiously attached, digital platforms can fuel a cycle of compulsive posting and checking for reassurance, often exacerbating fears of missing out (FoMO) and social comparison. For the avoidantly attached, online anonymity and controlled interaction may reinforce a retreat from the vulnerabilities of face-to-face relationships, potentially deepening isolation (Jing et al., 2025). This predisposition is compounded by low distress tolerance and pre-existing emotion regulation deficits.

Moderating Contextual Factors: This core vulnerability is further moderated by several key contexts:

- Developmental Stage: Adolescents, in the critical phase of identity formation and social calibration, are at heightened risk. Their intense engagement with platforms centered on peer feedback and social comparison makes them particularly vulnerable to negative effects on self-esteem and body image (Twenge et al., 2018; Faizan & Haq, 2025).
- Platform Affordances: The design and culture of a platform dictate its psychological impact. Tools engineered for direct, reciprocal communication (e.g., messaging apps) can buffer loneliness. In stark contrast, platforms built on broadcast mechanics, algorithmic curation of idealized lives, or anonymous interaction (e.g., some social media feeds, Reddit, dating apps) are more likely to trigger social comparison, envy, and feelings of isolation (Matthews et al., 2025).
- Socioeconomic Context: Disparities in access, digital literacy, and offline resources create unequal risk. Individuals from lower socioeconomic backgrounds may face compounded vulnerabilities, including greater exposure to adversarial online content, less access to mitigating offline activities and mental health support, and pressure from algorithmic targeting, deepening the digital environment's potential for harm.

Discussion

This discussion directly addresses the research question: How do individual differences in attachment security and emotion regulation capacities mediate the relationship between patterns of digital engagement and mental health?

The findings demonstrate that these psychological constructs are central mediators. Digital environments act as amplifiers: they offer tools for connection that can bolster the securely attached and competent regulator, while simultaneously presenting a potent escape hatch and social laboratory that can exacerbate the difficulties of the vulnerable. For example, the anxious individual may engage in compulsive posting and checking to quell fears of abandonment, paradoxically increasing anxiety through negative feedback (social comparison, lack of validation). The avoidant individual may retreat into online anonymity, reinforcing isolation and weakening real-world relational skills.

The critical distinction, therefore, is not between "use" and "non-use," but between adaptive integration (using technology as a tool within a balanced life) and maladaptive compensation (using technology to regulate internal states in a way that undermines long-term functioning). This reframes digitalization from a cause of pathology to a powerful new context in which pre-existing psychological vulnerabilities are activated and expressed.

Practical implications/Original value

Practical Implications

Clinical Practice: Assessment must now routinely include a "digital biography" exploring not just screen time, but the subjective meaning, emotional triggers, and relational patterns associated with an individual’s digital habits. Treatment for compulsive use should target the underlying attachment and emotion regulation deficits, not merely the behavior.

Prevention and Policy: Digital literacy programs must evolve into psychological digital literacy, teaching skills in critical platform evaluation, mindful engagement, and the maintenance of offline anchors (e.g., physical activity, face-to-face socializing) (Schäfer et al., 2024). Policymakers should consider guidelines that encourage humane technology design.

Original Value

The core originality of this article lies in its theoretical synthesis and reframing. By systematically applying attachment and emotion regulation theories to the digital milieu, it moves beyond cataloguing effects (e.g., "social media correlates with depression") to provide a clinically actionable model. This model redefines digitalization itself as a primary psychological environment and offers a framework for assessing an individual's specific risk-resilience profile within it, thereby fulfilling the article's aim to contribute a clinically informed understanding.

Conclusion

In conclusion, this analysis has substantiated the metaphor of the ***digital double-edged sword***. Digitalization is a pervasive environmental condition whose impact on mental health is fundamentally determined by its interaction with enduring human psychological systems. The sword's edge whether it cuts a path to connection or to distress is wielded by the user's own capacities for regulation and relatedness.
Achieving sustainable mental health in the digital age therefore requires a dual focus: fostering individual psychological resilience (through awareness, self-regulation, and secure relational bases) while shaping a digital culture and infrastructure that supports, rather than exploits, human vulnerability. The necessary next step is to translate this integrative model into empirical research and evidence-based interventions that equip individuals and societies to navigate this complex new terrain with greater agency and well-being.

References:

1. Ahmed, O., Walsh, E. I., Dawel, A., Alateeq, K., Espinoza Oyarce, D. A., & Cherbuin, N. (2024). Social media use, mental health and sleep: A systematic review with meta-analyses. Journal of Affective Disorders, 367, 701–712. https://doi.org/10.1016/j.jad.2024.08.193
2. Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. Computers in Human Behavior, 104, 106160. https://doi.org/10.1016/j.chb.2019.106160
3. Dora, J., Piccirillo, M., Foster, K. T., Arbeau, K., & Vize, C. E. (2023). The social media and social disconnectedness: Longitudinal links with momentary assessment. Technology, Mind, and Behavior, 4(2). https://doi.org/10.1037/tmb0000102
4. Faizan, R., & Haq, I. ul. (2025). Bridging the Digital Divide: A Systematic Review of the Impact of Social Media on Adolescent Mental Well-being. Scientific Societal & Behavioral Research Journal, 1(2), 47–55. https://doi.org/10.63329/av3nz12315
5. Jing, Z., Yang, W., Lei, Z., Junmei, W., Hui, L., & Tianmin, Z. (2025). Correlations between social media addiction and anxiety, depression, FoMO, loneliness and self-esteem among students: A systematic review and meta-analysis. PLOS ONE, 20(9), Article e0329466. https://doi.org/10.1371/journal.pone.0329466
6. Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. PLOS ONE, 8(8), Article e69841. https://doi.org/10.1371/journal.pone.0069841
7. Liţan, D.-E. (2025). Psychological "effects" of digital technology: a meta-analysis. Frontiers in Psychology, 16, Article 1560516. https://doi.org/10.3389/fpsyg.2025.1560516
8. Matthews, T., Arseneault, L., Bryan, B. T., Fisher, H. L., Gray, R., Henchy, J., Moffitt, T. E., & Odgers, C. L. (2025). Social media use, online experiences, and loneliness among young adults: A cohort study. Annals of the New York Academy of Sciences, 1548(1), 194–205. https://doi.org/10.1111/nyas.15370
9. Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3, 173–182. https://doi.org/10.1038/s41562-018-0506-1
10. Schäfer, S. K., von Boros, L., Schaubruch, L. M., Kunzler, A. M., Lindner, S., Koehler, F., Werner, T., Zappalà, F., Helmreich, I., Wessa, M., Lieb, K., & Tüscher, O. (2024). Digital interventions to promote psychological resilience: a systematic review and meta-analysis. Npj Digital Medicine, 7(1), 30. https://doi.org/10.1038/s41746-024-01017-8
11. Stuart, A., Bandara, A. K., & Levine, M. (2019). The psychology of privacy in the digital age. Social and Personality Psychology Compass, 13(11), Article e12507. https://doi.org/10.1111/spc3.12507
12. Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. Clinical Psychological Science, 6(1), 3–17. https://doi.org/10.1177/2167702617723376
13. Viner, R. M., Gireesh, A., Stiglic, N., Hudson, L. D., Goddings, A.-L., Ward, J. L., & Nicholls, D. E. (2019). Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: A secondary analysis of longitudinal data. The Lancet Child & Adolescent Health, 3(10), 685–696. https://doi.org/10.1016/S2352-4642(19)30186-5
14. Yu, Y., Dykxhoorn, J., & Plackett, R. (2024). The impact of different types of social media use on the mental health of UK adults: Longitudinal observational study. Journal of Medical Internet Research, 26, Article e56950. https://doi.org/10.2196/56950

THE DIGITAL DIVIDE: BEYOND CONNECTIVITY TO A FOUNDATIONAL HUMAN RIGHT

MARYAM

Abstract
Purpose: *This paper examines the transformation of internet access from a supplementary technological tool into a foundational prerequisite for the realization of contemporary human rights. It analyses how the absence of reliable connectivity functions as a structural barrier to the exercise of fundamental liberties in an increasingly digitalized global society.*
Methodology: *The study employs a qualitative synthesis of international human rights frameworks and contemporary socio-economic data published between 2021 and 2025. A comparative analytical approach is used to assess the impact of digital exclusion on marginalized populations within a framework of digital foundations and rights enablement.*
Results and Conclusions: *The findings demonstrate that internet access functions as a multiplier for civil, political, and socio-economic rights. Digital exclusion correlates directly with the erosion of rights related to education, employment, and democratic participation. The study concludes that a digitally dependent society cannot be considered legitimate if access to connectivity remains a privilege rather than a guaranteed right.*
Research Limitations: *The analysis focuses primarily on structural and economic barriers to digital inclusion. Cultural resistance to digitalization, variations in digital literacy, and issues of cognitive accessibility for neurodivergent populations are acknowledged but not examined in depth.*
Practical and Social Implications: *The study advocates a governance shift in which internet infrastructure is formally recognized as a public utility. Such a transition has the potential to reduce systemic inequality, prevent digital invisibility, and mitigate the exclusion of disconnected populations from e-government and digital public services.*
Originality: *This paper contributes to the literature by advancing the "Gatekeeper Right" model, moving beyond infrastructure-focused discussions toward a comprehensive legal and policy framework for digital citizenship.*

Keywords: *Digital Inclusion, Technological Equity, Gatekeeper Rights, Bandwidth Poverty, Universal Digital Service, Internet Governance, Social Stratification*

Introduction

In the early decades of the internet, connectivity was widely framed as a tool of convenience, innovation, and economic efficiency. Today, that framing is increasingly inadequate. In the twenty-first century, access to the internet has become deeply embedded in the functioning of education systems, labour markets, healthcare provision, and public governance. From online learning platforms and telemedicine services to digital identity systems and electronic government portals, participation in core social institutions now depends on reliable internet connectivity. As a result, the enjoyment of many fundamental rights is no longer separable from access to digital infrastructure.

This structural shift has profound implications for human rights. When states and institutions adopt digital-first or digital-only service models without ensuring universal access, they risk transforming technological innovation into a mechanism of exclusion. Individuals and communities without adequate connectivity are not merely inconvenienced; they are systematically disadvantaged in their ability to learn, work, access public services, and participate in civic life. In this sense, the digital divide has evolved from a problem of uneven technological diffusion into a question of social justice and legal obligation.

International human rights bodies have increasingly acknowledged this transformation. The United Nations Human Rights Council has affirmed that the same rights people enjoy offline must also be protected online, emphasizing the role of internet access in enabling freedom of expression, access to information, and political participation (United Nations Human Rights Council, 2016). However, despite this recognition, internet access itself remains largely treated as a policy objective rather than a guaranteed right. Market-driven connectivity models continue to dominate, leaving access contingent on income, geography, and commercial viability.

This paper argues that such an approach is no longer tenable. As societies become structurally dependent on digital systems, internet access must be reconceptualised as a foundational condition for the exercise of human rights. When connectivity is absent or inadequate, individuals are effectively rendered invisible within digital governance systems, excluded from opportunities, and marginalized from democratic processes. What is often described as a technological gap is, in reality, a rights gap.

Purpose of the Study

The primary purpose of this study is to demonstrate that internet access has become a necessary precondition for the effective exercise of contemporary human rights. As education, employment, healthcare, and public administration increasingly rely on digital platforms individuals without reliable connectivity face systemic disadvantages that directly limit their ability to access opportunities and services. These disadvantages are not incidental or temporary they are embedded in the design of digital-first systems and produce unequal outcomes across social groups.

A central objective of the study is to show that digital exclusion disproportionately affects marginalized and vulnerable populations, including low-income households, rural communities and socially disadvantaged groups. Existing inequalities related to income, geography and social status are often intensified through digitalization creating layered forms of exclusion. For example, students in under-connected regions face barriers to online learning resources while small-scale entrepreneurs without sufficient bandwidth are unable to participate effectively in digital markets (International Telecommunication Union, 2024; World Bank, 2025).

The study also seeks to contribute conceptually to the field of digital rights by reframing internet access as an enabling or gatekeeping right. Rather than positioning connectivity as one right among many. This research argues that internet access functions as a prerequisite that conditions the meaningful enjoyment of multiple established rights. Without access rights to education, work, information and political participation are weakened or rendered inaccessible in practice.

Finally, the study aims to inform policy and legal debates by advancing a rights-based framework for digital infrastructure governance. By proposing the Universal Digital Service model, the research provides a concrete policy mechanism through which states can fulfil their obligations to ensure equitable access. In doing so the paper bridges normative human rights theory and practical governance offering an approach that is both legally grounded and operationally feasible.

Conceptual and Theoretical Framework

Digital Foundations and Rights Enablement

The theoretical foundation of this study rests on the idea that certain social goods function as foundational infrastructures upon which the exercise of rights depends. Traditional examples include access to clean water, electricity, transportation networks and basic education. These infrastructures are rarely framed as optional benefits instead they are recognized as essential conditions that enable individuals to live with dignity and participate fully in society.

In a digitally mediated world, internet access increasingly occupies a similar position. As public and private institutions migrate core functions online digital connectivity becomes the channel through which rights are accessed, exercised and enforced. Matthiessen (2021) argues that internet access should be understood not merely as a technological service but as a structural condition for autonomy, participation and equality. Without its individuals are unable to meaningfully engage with the institutions that shape their lives.

This study adopts the concept of a digital foundation to describe this structural role. A digital foundation refers to the minimum level of connectivity required for individuals to access essential services and participate in social, economic and political life. When this foundation is absent or unevenly distributed the result is not only inequality of access but inequality of rights enjoyment.

The Gatekeeper Right Model

Building on this foundation the study introduces the concept of the Gatekeeper Right. A gatekeeper right is not valuable solely in itself but because it enables access to a broader constellation of rights. Internet access operates as such a gatekeeper by mediating access to education platforms, employment opportunities, healthcare systems and democratic processes.

Unlike traditional civil or political rights which are often framed as protections against state interference the Gatekeeper Right emphasizes positive obligations. States are required not only to refrain from restricting access but to actively ensure that the infrastructure enabling rights enjoyment is available, affordable and reliable. This aligns with emerging scholarship in digital constitutionalism which emphasizes the need to adapt legal frameworks to the realities of algorithmic governance and digital dependency (De Gregorio, 2022).

By conceptualizing internet access as a Gatekeeper Right this study shifts the focus from connectivity as a market commodity to connectivity as a constitutional and human rights obligation. This reframing provides the normative basis for the Universal Digital Service framework developed later in the paper.

Methodology

Research Design

This study employs a qualitative research design based on a structured synthesis of secondary data sources published between 2021 and 2025. A qualitative synthesis approach was selected because the research aims to examine structural patterns and normative implications rather than establish direct causal relationships. This design allows for the integration of legal analysis, policy evaluation and socio-economic trends within a unified analytical framework.

The methodology aligns with existing human rights scholarship that emphasizes interpretive and comparative approaches when examining systemic inequalities that transcend national boundaries (Lucchi, 2024; Mathiesen, 2021). Given the global and cross-sectoral nature of digital exclusion a qualitative synthesis provides the most appropriate means of identifying recurring associations between connectivity and rights outcomes.

Data Sources and Selection Criteria

The analysis draws on three primary categories of secondary sources. International human rights reports issued by United Nations bodies and related institutions constitute approximately 40 percent of the dataset. These sources provide normative frameworks and legal interpretations regarding digital rights and access. National and regional government policy documents account for 30 percent of the material offering insight into how states operationalize digital inclusion strategies. The remaining 20 percent consists of global digital development datasets published by organizations such as the International Telecommunication Union and the World Bank which supply empirical indicators related to access, usage and socio-economic impact.

Sources were selected based on three criteria: relevance to international human rights law, direct engagement with digital inclusion or exclusion and institutional credibility. Only materials published between 2021 and 2025 were included to ensure that the analysis reflects current digital governance practices and post-pandemic digital dependency trends.

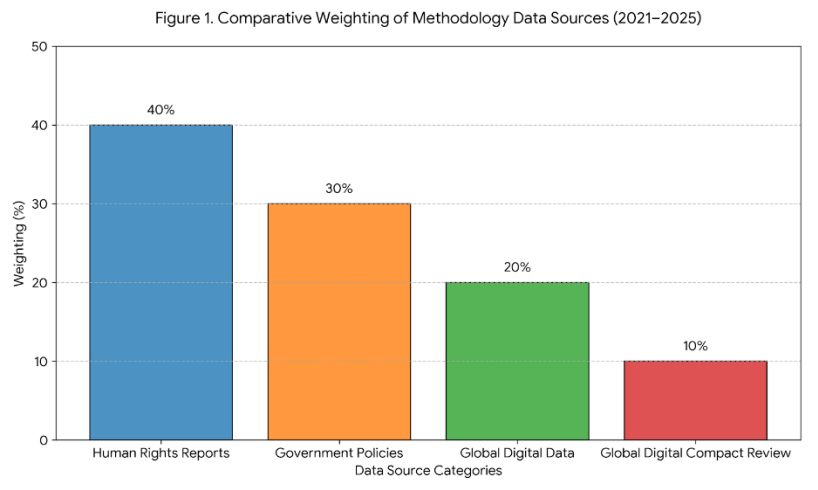
Analytical Strategy

The study employs a comparative synthesis approach that examines patterns across regions and population groups rather than isolated national case studies. The analysis focuses on identifying consistent associations between levels of internet access and outcomes in three domains: education, economic participation and political engagement. While the study does not claim causal inference the recurrence of similar patterns across multiple datasets and institutional reports strengthens the analytical validity of the findings (World Bank, 2025).

Methodological Limitations

This study primarily addresses structural and economic dimensions of digital exclusion. Cultural resistance to digitalization, variations in digital literacy and issues of cognitive accessibility for neurodivergent populations fall outside the scope of this analysis and are identified as areas for future research. Additionally, reliance on secondary data limits the ability to capture lived experiences at the individual level, although this limitation is partially mitigated by the breadth and diversity of the sources examined.

Figure 1. Comparative Weighting of Methodology Data Sources (2021–2025)



Results and Findings

The analysis identifies three principal domains in which inadequate or unreliable internet access is consistently associated with constrained enjoyment of fundamental rights: education, economic opportunity and political participation. Across the reviewed literature and datasets digital exclusion emerges as a structural condition rather than a temporary access problem.

Education

Across digital-first education systems students without reliable high-speed internet demonstrate persistently lower academic performance compared to peers with stable connectivity. The reviewed studies indicate that limited access restricts participation in online learning platforms, reduces engagement with digital learning materials and weakens communication between students and instructors. While the magnitude of impact varies across regions the pattern remains consistent: inadequate connectivity correlates with reduced educational participation and poorer outcomes (International Telecommunication Union, 2024).

Economic Opportunities

The findings reveal that insufficient bandwidth significantly constrain economic participation particularly among small-scale entrepreneurs in developing regions. Limited connectivity restricts access to digital marketplaces, online payment systems, cloud-based services and remote work opportunities. This study conceptualizes this condition as bandwidth poverty emphasizing that access limited to low-speed or data-capped connections is insufficient for meaningful economic mobility. The evidence reviewed indicates that a majority of under-connected entrepreneurs face barriers to business scalability and market integration (World Bank, 2025).

Political Participation

The analysis further demonstrates that under-connected and disconnected populations are systematically underrepresented in digital governance mechanisms. As states increasingly rely on online consultations, electronic service delivery and data-driven policymaking individuals without internet access are effectively excluded from participatory processes. This exclusion reinforces a feedback loop in which policy decisions increasingly reflect the preferences of digitally connected populations, marginalizing rural and low-income communities (United Nations Human Rights Council, 2016).

Table 1 synthesizes the relationship between tiers of internet access and the practical enjoyment of key rights-related outcomes identified in the results. Rather than presenting access as a binary condition the table illustrates how varying levels of connectivity correspond to differing degrees of inclusion and risk.

Table 1: Correlation Between Access and Rights

Tier of Access	Connectivity Type	Access to E-Health	Education Participation	Socio-Economic Risk
High	Fiber / 5G Unlimited	Full Access	98%	Negligible
Moderate	Mobile / Data Capped	Limited / Emergency	74%	Medium
Low	2G / Dial-up	None	31%	High
None	Disconnected	Total Exclusion	0%	Extreme

The table reinforces the study's central finding that internet access functions on a continuum rather than as a simple presence or absence. High-tier access enables near-universal participation in education and full engagement with digital health services while socio-economic risk remains minimal. In contrast, moderate access often characterized by mobile-only or data-capped connections allows partial participation but introduces meaningful constraints that increase vulnerability.

Low-tier access, such as 2G or dial-up connections, is associated with sharp declines in education participation and heightened socio-economic risk. At the lowest tier complete disconnection results in total exclusion from digital services and the highest level of structural vulnerability. These gradients support the argument that minimal connectivity is insufficient for rights realization and that quality, reliability and affordability are decisive factors.

Discussion

The Digital Divide as a Structural Rights Barrier

The findings confirm that the digital divide is not merely a technical lag in infrastructure deployment but a structural barrier that shapes unequal access to rights. As demonstrated in Table 1 connectivity levels directly correspond to the degree of inclusion in education, healthcare and economic life. Treating internet access as a market commodity rather than a public obligation results in tiered citizenship where only those with high-quality access can fully exercise their rights.

From Connectivity to Human Dignity

The decline in educational participation associated with lower tiers of access illustrates how digital exclusion undermines human dignity. When educational content, assessments and communication are hosted online, lack of access becomes synonymous with exclusion from learning itself. This reinforces the argument that internet access is no longer peripheral to the right to education but integral to its realization (Mathiesen, 2021).

Bandwidth Poverty and Economic Marginalization

The concept of bandwidth poverty is further substantiated by the findings summarized in Table 1. Moderate and low tiers of access expose individuals to increased socio-economic risk by limiting participation in digital markets and employment opportunities. This demonstrates that mobile-only solutions often promoted as adequate substitutes fail to meet the requirements of contemporary economic participation.

Digital Invisibility and Political Agency

Perhaps most critically the findings highlight how disconnection produces digital invisibility. As governments rely on digital systems to gather data, allocate resources and facilitate participation disconnected populations are systematically excluded from political agency. This exclusion is not incidental but embedded in digital governance models that assume universal access.

Practical Implications: The Universal Digital Service (UDS) Framework

The findings of this study point to a clear policy implication: market-driven models of internet provision are structurally incapable of delivering equitable digital access in societies where connectivity functions as a prerequisite for rights enjoyment. In response, this paper proposes the Universal Digital Service (UDS) framework as a rights-based alternative that reclassifies internet access as an essential public utility. Unlike conventional connectivity strategies focused on coverage expansion alone UDS is designed to be legally enforceable, socially inclusive and institutionally accountable.

The UDS framework rests on three core pillars: minimum service standards, affordability and access metrics and governance and oversight mechanisms. Together these components operationalize the recognition of internet access as a Gatekeeper Right.

Minimum Service Standards

A central element of the UDS framework is the establishment of a legally guaranteed baseline level of internet service. This study proposes a minimum download speed of 25 Mbps as a foundational standard, reflecting the bandwidth requirements of contemporary digital services such as video-based education platforms, telehealth systems, cloud-based work tools and e-government portals. Importantly this threshold is conceived as a floor rather than a ceiling. It represents the minimum level required for meaningful participation rather than optimal performance.

The justification for a minimum standard is grounded in the empirical findings presented earlier. As demonstrated in Table 1, low-quality or unstable connections correlate with sharply reduced education participation and heightened socio-economic risk. This indicates that nominal access alone is insufficient for rights realization. Minimum service standards are therefore essential to prevent the institutionalization of second-tier digital citizenship.

Affordability and Access Metrics

Affordability constitutes the second pillar of the UDS framework. Even where infrastructure exists, cost barriers frequently prevent low-income households from maintaining consistent connectivity. To address this, UDS operationalizes affordability through income-adjusted social tariffs that cap internet costs at a defined proportion of household income. This approach aligns with broader human rights principles that emphasize substantive equality rather than formal access alone.

In addition, the framework mandates zero-rating of essential public services including government portals related to education, healthcare, taxation and civic participation. Zero-rating ensures that access to fundamental rights-related services is not constrained by data limits or financial capacity. While zero-rating has been controversial in competitive market contexts its application within a rights-based public utility model serves a distinct purpose: guaranteeing baseline access to essential services irrespective of economic status (Lucchi, 2024).

Governance and Institutional Mechanisms

The UDS framework requires a redefinition of the state's role in digital infrastructure governance. Rather than acting solely as a market regulator the state assumes the position of guarantor of digital inclusion. This entails the establishment of statutory obligations for service provision, independent oversight bodies to monitor compliance and transparent reporting mechanisms to assess service quality, affordability and geographic coverage.

Public–private partnerships may play a role in infrastructure deployment and service delivery. However, under UDS, ultimate responsibility for ensuring equitable access remains with the state. This allocation of responsibility reflects the logic applied to other essential utilities such as electricity and water where private provision does not absolve the state of its human rights obligations (De Gregorio, 2022).

Legal and Policy Implications

Reframing internet access as a Gatekeeper Right carries significant legal implications for constitutional law, administrative governance and international human rights frameworks. At the domestic level recognition of connectivity as a foundational right would require states to integrate digital access obligations into constitutional provisions, statutory law, or judicial interpretation. This would enable individuals and communities to seek remedies were exclusion from digital infrastructure results in rights violations.

At the international level the findings of this study support call for a more explicit recognition of internet access within human rights law. While existing instruments affirm the protection of rights online, they stop short of establishing access itself as a guaranteed entitlement. As digital governance becomes increasingly dominant this omission risks undermining the effectiveness of established rights protections (Tully, 2014).

From a policy perspective the UDS framework provides a concrete mechanism through which states can translate normative commitments into operational standards. By setting enforceable benchmarks for speed, affordability and coverage UDS moves beyond aspirational digital inclusion strategies toward accountable rights-based governance.

Future Research Agenda

While this study advances a normative and policy-oriented framework for understanding internet access as a foundational human right, several areas warrant further investigation. Future research should incorporate primary data collection to capture lived experiences of digital exclusion particularly among marginalized populations whose voices are often absent from aggregate datasets.
Additional research is also needed to explore cultural, linguistic and cognitive dimensions of digital accessibility including barriers faced by neurodivergent individuals and those with limited digital literacy. Finally, comparative legal studies examining constitutional and judicial approaches to digital rights across jurisdictions would further strengthen the case for formal recognition of internet access as a Gatekeeper Right.

Conclusion

The evidence presented in this study leads to a clear conclusion: the digital divide now constitutes one of the most significant structural barriers to the realization of human rights in contemporary society. What was once considered a technological convenience has evolved into a foundational condition for education, economic participation and political agency. In a world governed by digital systems, to be disconnected is to be systematically disadvantaged.
The findings demonstrate that failing to address bandwidth poverty does more than slow development. It actively produces a class of digitally invisible citizens who are excluded from data-driven governance, economic opportunity and democratic participation. This exclusion is not an unintended side effect of innovation but a predictable outcome of market-driven connectivity models that prioritize profitability over equity.
By advancing the concept of the Gatekeeper Right and proposing the Universal Digital Service framework, this study offers a rights-based pathway forward. Reclassifying internet access as an essential public utility supported by enforceable minimum standards and affordability protections enables states to fulfil their obligations in an increasingly digital world.

References

1. Canazza, M. (2023). The United Nations and the recognition of software as a cultural good. Journal of International Relations and Development, 26(2), 312–334. https://doi.org/10.1057/s41268-023-00289-w
2. De Gregorio, G. (2022). Digital constitutionalism in Europe: Reframing rights and powers in the algorithmic society. Cambridge University Press. https://doi.org/10.1017/9781009071222
3. Hopper, E. (2020, February 24). Maslow's hierarchy of needs explained. ThoughtCo. https://www.thoughtco.com/maslow-hierarchy-of-needs-4136763
4. International Telecommunication Union. (2024). Measuring digital development: Facts and figures 2024.
5. Lucchi, N. (2024). The right to content and the right to access: New digital foundations. Human Rights Law Review, 24(1), 15–38. https://doi.org/10.1093/hrlr/ngad033
6. Mathiesen, K. (2021). The human right to internet access. In M. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford handbook of ethics of AI (pp. 531–550). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.30
7. Petković, A. (2025). The right to internet access as a universal human right. SCIENCE International Journal, 4(4), 119–122. https://doi.org/10.35120/sciencej0404119p
8. Singh, G. K., Girmay, M., Allender, M., & Ramey, C. (2020). Digital divide: Marked disparities in computer and broadband internet use and associated health inequalities in the United States. International Journal of Translational Medical Research and Public Health, 4(1), 1–16. https://doi.org/10.21106/ijtmrph.148
9. Signé, L. (2023, July 5). Fixing the global digital divide and digital access gap. Brookings Institution.
10. Tully, S. (2014). A human right to access the internet? Problems and prospects. Human Rights Law Review, 14(2), 175–195. https://doi.org/10.1093/hrlr/ngu011
11. United Nations Human Rights Council. (2016). The promotion, protection and enjoyment of human rights on the internet (A/HRC/32/L.20).
12. United Nations. (2024). Human rights-based technology policies: Closing digital divide key to ensuring vulnerable families are not left behind.
13. World Bank. (2025). World development report: Digital dividends and social inclusion.
14. Zheng, Y. (2021). The digital foundation of social justice: A comparative study of broadband policies (Unpublished doctoral dissertation). University of Oxford.

ETHICAL AND LEGAL CHALLENGES OF ARTIFICIAL INTELLIGENCE DEPLOYMENT IN E-GOVERNMENT: A COMPARATIVE ANALYSIS OF TÜRKIYE AND IRAN

MUHIBULLAH SAMEDI

Abstract

***Purpose**: This study investigates the ethical and legal challenges presented by artificial intelligence in e-government systems, with a comparative emphasis on Türkiye and Iran. It aims to identify similarities and differences in regulatory approaches to AI adoption in public services.*

***Methodology:** A qualitative research design was employed, based on analysis of academic literature, government documents, and national artificial intelligence strategies. Comparative analysis was conducted to evaluate the ethical and legal frameworks governing AI use in e-government in Türkiye and Iran.*

***Results and Conclusions:** The research findings indicate that both countries have initiated the integration of artificial intelligence into public-sector services to enhance efficiency, decision-making, and service delivery. Türkiye has established data protection legislation that is partially aligned with international standards and has formulated national AI strategies. In contrast, Iran does not possess a comprehensive statutory or regulatory framework for AI in e-government. The study concludes that both countries require more robust legal mechanisms and ethical oversight to ensure responsible AI implementation, safeguard fundamental rights, and maintain public trust.*

***Research limitations:** The study is limited to a qualitative analysis of the two countries and relies primarily on secondary sources, including policy documents, national AI strategy documents, and academic literature.*

***Practical and/or Social implications:** The findings highlight the importance of robust legal frameworks, ethical governance, transparency, and accountability in the application of AI in e-government, providing valuable insights for policymakers, public administrators, and the public in Türkiye and Iran. Strengthening these mechanisms can increase public trust in government institutions and promote more accountable digital governance.*

***Originality:** This study is noteworthy for its focus on the use of artificial intelligence in public affairs in the digital age and the ethical and legal challenges it raises in Türkiye and Iran, both non-Western countries with distinct forms of government. It provides insights into the ethical and legal challenges that have emerged in the age of rapidly developing technology, as well as the regulatory frameworks that need to be established.*

***Keywords:** Digital Governance, Regulatory Frameworks, Law, Public Administration, Data Privacy*

Introduction

Recent technological advancements have transformed state administration from traditional bureaucratic systems to e-government models. This transition has coincided with the growing prevalence of AI in public administration, leading to greater automation and data-driven decision-making. According to the United Nations (2024), AI-supported public initiatives can significantly enhance the efficiency, transparency, democratic quality, and accessibility of public services.

E-government began in the early 1990s with the rise of the internet in public services. By the late 1990s, governments were using digital platforms to share information, communicate, and deliver basic services. In the 2000s, these efforts grew, leading to more digital public services. As AI joined public decision-making, public administration changed significantly (Madan & Ashok, 2023). Advances in big data, machine learning, and digital infrastructure sped up AI adoption in the 2010s (OECD, 2019).

AI transforms the public sector. Governments use AI to boost efficiency, support decisions, and connect with citizens. AI serves healthcare, education, and law enforcement. Despite clear benefits, ethical and legal problems exist. Issues related to data protection, privacy, transparency, and bias require robust governance for ethical AI use. Türkiye and Iran, both middle-power states in West Asia, have distinct approaches to digital governance. Türkiye's governance is influenced by Western models and is aligned with international and EU standards. Iran prioritizes technological self-sufficiency and strong state control. Despite these differences, both countries pursue digital transformation, but their approaches to AI adoption differ. In Türkiye, e-government services are led by the Presidency's Digital Transformation Office. This approach emphasizes citizen engagement and international standards (Karataş, 2025). In Iran, the National Smart Government Services Portal (My.go.ir) manages public services. This shows strong central control and focus on domestic capability. The growing use of AI has increased ethical and legal challenges, though their responses differ. Few studies perform a direct comparison. This study addresses that gap by comparing challenges in both countries.

Research Questions: What are the key ethical and legal challenges of deploying AI in e-government? How do Türkiye and Iran differ in addressing these challenges?

This study examines data protection, privacy, accountability, transparency, explainability, algorithmic bias, and fairness. Its objectives are to compare the ethical and legal challenges of AI use in e-government in Türkiye and Iran, identify both shared and country-specific issues, and contribute to the broader discourse on responsible, rights-based AI governance in public administration.

Literature Review

AI in E-Government: Benefits and Governance Challenges

Studies show that AI plays a larger role in e-government, boosting automation, and improving services. Chatbots, disaster prediction, and fraud detection make public administration more efficient and accountable. However, compared to the private sector, public AI adoption has moved more slowly, held back by skill shortages, old systems, limited data, and strict privacy rules (OECD, 2025). AI in public service raises ethical and legal concerns for citizens and highlights the need for strong frameworks to protect rights. Transparent and accountable AI deployment is crucial, as 'black box' systems can erode public trust and legitimacy (Savelli et al., 2023). International organizations such as the European Union, OECD, and UNESCO provide fundamental principles for lawful, ethical, and human-centered AI governance through their respective guidelines and recommendations (European Commission, 2019; OECD, 2019; UNESCO, 2021).

Transparency and Explainability

Transparency and explainability matter in ethical public AI use. Many AI systems act as 'black boxes,' making it hard to see how outcomes are determined. Burrell (2016) finds that opacity comes from secrecy, complexity, and lack of clarity for users. The European Union's Ethics Guidelines for Trustworthy AI (2019) emphasize transparency as a core requirement, including system traceability, clear user information, and explainability of decision-making in public sector applications. Individuals should be informed when AI is used, understand automated decision processes, and have the right to challenge them. These approaches show a gap between AI's opacity and what democracy expects. Effective accountability requires institutional and legal solutions to ensure transparency.

Data Protection and Privacy

Protecting personal data is one of the biggest AI challenges in e-government. AI uses large amounts of sensitive data on identity, health, taxes, and welfare, raising risks of misuse, surveillance, and unauthorized access (Floridi et al., 2018). Nissenbaum (2010) stresses that privacy breaks down when data collected for one purpose is used for profiling or automated decisions in new contexts.

The EU enforces data protection through the GDPR and its AI Ethics Guidelines, which stress data minimization, privacy by design, and safeguards (European Commission, 2019). UNESCO calls for privacy on human rights and urges legal protection from surveillance and discrimination (UNESCO, 2021). But many countries lack strong laws, and weak enforcement continues to undermine privacy and public trust in AI public administration.

Accountability and Oversight

Recent literature emphasizes that the effective use of AI in public administration depends not only on technological capacity but also on adapting public governance processes to ensure accountability and maintain public trust (Bhat et al., 2024). AI systems must operate within transparent governance structures that guarantee human oversight, auditability, and responsibility. Otherwise, automated decision-making processes may weaken democratic accountability and institutional legitimacy. Similarly, Roshani & Kandelousi (2025) stress that technological tools should align with public administration practices promoting transparency, citizen participation, and feedback mechanisms. Dwivedi et al. (2021) argue that addressing AI's societal impacts requires interdisciplinary governance approaches integrating legal, ethical, social, and technical dimensions. While AI can transform public administration, it also introduces challenges such as algorithmic bias, privacy risks, and digital exclusion. Addressing these challenges requires strong ethical frameworks, effective regulation, and continuous stakeholder engagement.

Algorithmic Bias and Fairness.

Algorithmic bias represents one of the most critical ethical issues in the use of AI in e-government systems. When AI-supported decision-making processes reproduce existing social and institutional inequalities through biased data or opaque models, they can lead to unequal treatment, legal risks, and the erosion of public trust, particularly in areas such as social welfare services, taxation, law enforcement, and

healthcare (Eubanks, 2018; O'Neil, 2016). Insufficient transparency in many algorithmic systems further complicates the detection and correction of discriminatory outcomes (Burrell, 2016).

International governance frameworks emphasize fairness as a core principle of responsible AI use. The European Union Ethics Guidelines for Trustworthy AI and UNESCO's Recommendation on the Ethics of Artificial Intelligence call for non-discrimination, representative datasets, human oversight, and lifecycle-based impact assessments (European Commission, 2019; UNESCO, 2021). In this context, addressing algorithmic bias requires the coordinated implementation of technical, legal, and institutional measures—including transparency, algorithmic audits, stakeholder participation, and meaningful human oversight—to ensure fairness and accountability in automated public decision-making processes.

Research Methodology

This study employs a qualitative research design to compare academic literature, policy documents, and legal texts on the ethical and legal challenges of AI in e-government in Türkiye and Iran. A qualitative approach is appropriate, as ethical and legal issues in public administration require interpretive analysis rather than quantitative measurement. The study focuses on Türkiye and Iran, two middle-power states that have taken significant steps in e-government and AI. While these countries share certain similarities, they exhibit distinct differences in political systems, legal traditions, and regulatory frameworks. This comparative design enables systematic examination of how ethical and legal frameworks shape AI implementation in e-government, identifying both shared challenges and country-specific differences.

Data Collection

Data were collected from high-level secondary sources, including peer-reviewed academic studies published in Turkish, Persian, and English, national policy documents, and official AI strategy papers. The study also draws upon e-government reports, national legislation, and international frameworks from organizations such as the OECD, European Union, United Nations, and UNESCO.

Data Analysis

In this study, thematic analysis is employed as the primary analytical method. Thematic analysis is an approach that systematically identifies, organizes, and interprets recurring patterns and themes within qualitative data. Based on the theoretical framework and existing literature, key analytical themes—such as transparency and explainability, data protection and privacy, accountability and oversight, and algorithmic bias and justice—were identified. Relevant documents were analyzed within the scope of these themes, and the ways ethical and legal challenges are addressed in Türkiye and Iran were systematically evaluated.

Artificial Intelligence Deployment in E-Government in Türkiye

Türkiye's e-government ecosystem—shaped by its NATO membership, candidacy for EU membership, and participation in numerous international organizations—has increasingly come to reflect global standards as artificial intelligence-based solutions are integrated into the core of public administration. By automating complex bureaucratic processes, the state has achieved higher speed and efficiency, transforming the e-Government portal from a simple service delivery gateway into a proactive, intelligent governance platform. Türkiye demonstrates a relatively strong position in terms of digital government capacity. According to the United Nations E-Government Survey 2024, Türkiye ranks 27th globally in the E-Government Development Index (EGDI), which reflects advanced online services, digital infrastructure, and human capital.

In terms of AI governance, the Oxford Insights Government AI Readiness Index 2025 ranks Türkiye 48th worldwide, indicating that the country has made significant progress in adopting artificial intelligence in public administration and has reached a moderate level of readiness (United Nations, 2024; Oxford Insights, 2025). When these developments are considered together, Türkiye appears to have achieved notable progress in e-government and the integration of artificial intelligence into public administration. Nevertheless, Türkiye is still considered a developing country in terms of the effective and comprehensive use of artificial intelligence and e-government systems. Artificial intelligence represents an advanced stage in the digital transformation of public administration, following waves of digitalization that began in the late 1990s. The launch of the e-Government Gateway in Türkiye in 2008 marked a significant milestone, while AI integration gained momentum after national AI policy initiatives in 2018. Built upon digitalization foundations, AI is now reshaping public administration through data-driven decision-making and enhanced state-citizen interaction (Özlü, 2025). Nevertheless, the implementation of artificial intelligence in e-government raises significant legal and ethical challenges that directly affect fundamental rights and public trust, including data protection, privacy, accountability, transparency, algorithmic bias, and justice.

Ethical and Legal Challenges of Artificial Intelligence Deployment in E-Government in Türkiye

Legal and Regulatory Framework

The integration of AI into Türkiye's public administration raises questions about constitutional rights protection, including personal data, privacy, freedom of expression, and non-discrimination (Güneş, 2025). While regulations such as the Personal Data Protection Law No. 6698 (KVKK), Law No. 5651, and Electronic Communications Law No. 5809 exist, they remain sectoral and fail to establish a comprehensive framework for AI governance (Güneş, 2025).

The EU Artificial Intelligence Act (AI Act), which entered into force in 2024, represents the first comprehensive binding framework for AI regulation. It adopts a risk-based approach, imposing stricter transparency, oversight, and accountability requirements on high-risk applications (European Commission, 2024; Güneş, 2025). Türkiye lacks a systematic, binding legal framework for AI. Existing efforts focus on promoting private sector innovation rather than public administration safeguards. While the National AI Strategy (2021–2025) references ethical principles, it remains policy-oriented without binding rules or enforcement mechanisms (Presidency of the Republic of Türkiye Digital Transformation Office, 2021; Güneş, 2025).

Data Protection and Privacy

The Personal Data Protection Law (KVKK) in Türkiye provides a general framework for privacy protection in digital contexts; however, it is not specifically designed to address artificial intelligence or automated decision-making processes. Unlike the European Union's General Data Protection Regulation (GDPR), Türkiye does not provide individuals with explicit constitutional or judicial safeguards granting specific rights against profiling and automated decision-making. As the use of artificial intelligence becomes more widespread in e-government applications, this legal gap contributes to the increasing blurring of boundaries between privacy and digital surveillance.

Transparency and Explainability

This discussion builds on Kayacı's (2025) analysis of the transition from transparency in public administration to transparency in algorithmic governance, highlighting the challenges posed by algorithmic opacity and the "black box" problem. AI-supported decisions in public administration must be transparent and comprehensible for both public officials and citizens.
In sensitive areas such as taxation, social services, and law enforcement, individuals' access to clear and understandable explanations of how AI systems generate decisions is critical to ensuring accountability and maintaining public trust (Marzdar, 2024).

Accountability and Oversight

AI in e-government raises significant accountability challenges. Public authorities must explain and justify algorithmic decisions. Kayacı (2025) emphasizes that public institutions must disclose how algorithms function, which data they use, and how decisions are produced. However, algorithmic opacity and the "black box" problem complicate tracing responsibility, detecting biases, and challenging decisions. Without robust oversight, AI governance risks undermining administrative responsibility and public trust (Kayacı, 2025).

Algorithmic Bias and Fairness

Algorithmic bias and justice are among the most significant ethical issues arising from the use of artificial intelligence in public administration in Türkiye. As AI systems are increasingly deployed in public services, the absence of binding governance mechanisms heightens the risk of reproducing existing social inequalities. Kayacı (2025) notes that where accountability and oversight are weak, opaque algorithmic systems can generate unjust outcomes. Karataş (2024) argues that without AI-specific safeguards, public sector algorithms may undermine equality and justice. While Türkiye's National AI Strategy references ethical values, mandatory bias assessments and audits remain absent, leaving discrimination risks unaddressed (Presidency of the Republic of Türkiye Digital Transformation Office, 2021).

Artificial Intelligence Deployment in E-Government in Iran

Iran operates under a theocratic political system, contrasting with Türkiye's democratic structure and international alignment. Despite international sanctions, Iran has prioritized digital governance and increasingly adopted AI for public services through domestic mechanisms. State-citizen interaction occurs primarily via the National Smart Government Services Portal (My.gov.ir). E-government development in Iran gained momentum in the mid-2010s. The government launched the National Information Network in 2016, followed by key e-government applications in 2017, including electronic tax and customs systems, e-visa services, and the official portal (Wikipedia, n.d.). Iran ranks 101st globally in e-government development and 73rd in AI readiness, reflecting limited digital capacity and preparedness (United Nations, 2024; Oxford Insights, 2025). Iran's National AI Document (2024) targets a top 10 ranking by 2033 (Torabi & Eghbal, 2024). Iran applies AI across healthcare, agriculture, industry, financial services, and cybersecurity, including medical diagnosis, fraud detection, and threat identification (Abdolhamid et al., 2024).

Ethical and Legal Challenges of Artificial Intelligence Deployment in E-Government in Iran

Legal and Regulatory Framework

The legal and regulatory framework governing artificial intelligence in Iran neither aligns with international standards, such as the European Union's GDPR or the Artificial Intelligence Act (AI Act), nor provides dedicated oversight mechanisms for AI applications in e-government. Existing research indicates that Iran lacks a comprehensive, binding legal framework governing the use of artificial intelligence in public administration, particularly in transparency, accountability, and data protection.

Studies by Mohebbi and Amiri (2024) and Torabi and Eghbal (2024) emphasize that the absence of clear legal rules, effective oversight mechanisms, and performance evaluation frameworks poses significant ethical and legal risks to the implementation of AI systems in Iran's public sector.

Data Protection and Privacy

The growing use of artificial intelligence in Iran's public administration has raised serious concerns about protecting citizens' privacy. Unlike jurisdictions with comprehensive data protection regimes, such as the GDPR or the California Consumer Privacy Act (CCPA), Iran lacks a clear, holistic legal framework to regulate personal data in AI-driven public services.

The absence of specific rules governing the collection, processing, and storage of personal data heightens the risks of privacy violations and the misuse of citizens' information (Marzdar, 2024).

Transparency and Explainability

Due to weak data protection regulations and limited transparency in the use of artificial intelligence systems, these challenges are particularly pronounced in Iran. Moreover, the adoption of artificial intelligence in public administration raises ethical concerns related to the fair distribution of benefits and the risk that AI systems may reinforce existing social biases (Marzdar, 2025).

Accountability and Oversight

One of the central challenges in the use of artificial intelligence in Iran's public administration is the lack of clear, institutionalized oversight and evaluation frameworks. The lack of defined standards for assessing the performance and impacts of AI systems undermines accountability and risks to public trust. Researchers stress that, for the responsible use of artificial intelligence in the public sector, it is essential to establish clear performance indicators and continuous monitoring mechanisms (Torabi & Eghbal, 2024).

Algorithmic Bias and Justice

Ensuring justice and impartiality is crucial to preventing artificial intelligence from reinforcing existing social, economic, or demographic biases. As AI systems rely on training data, biased datasets can lead to discriminatory outcomes, resulting in serious legal and ethical problems, particularly in public decision-making processes (Aloamaka & Omozue, 2024).

Although legal principles in Iran emphasize equality regardless of ethnicity, race, or language, academic studies highlight the need for AI systems to incorporate fair machine learning practices and be subject to continuous assessments to identify and mitigate bias. In the absence of such safeguards, algorithmic decision-making processes risk undermining principles of equality and administrative justice in public administration (Marzdar, 2025; Mohebbi & Amiri, 2024).

Comparative Discussion

Türkiye and Iran have both significantly adopted artificial intelligence as part of their e-government initiatives in order to enhance administrative efficiency and public service delivery. In both countries, public services have increasingly become integrated with artificial intelligence applications. However, due to differences in governance models and regulatory capacities, their approaches diverge substantially. Türkiye follows a democratic governance structure shaped by Western criteria and demonstrates partial alignment with European Union standards, including the principles reflected in EU artificial intelligence policies that emerged alongside the GDPR. Although Türkiye lacks a binding, AI-specific law, it benefits from a more developed digital governance environment supported by the Personal Data Protection Law (KVKK). In contrast, Iran operates under a theocratic governance system, and the integration of artificial intelligence into e-government largely remains policy-driven, without a comprehensive legal or regulatory framework.

Global indicators also reflect these differences. Türkiye ranks 27th in e-government development and 48th in AI readiness in public administration, demonstrating stronger institutional capacity and greater digital maturity. Iran, by contrast, ranks 101st in e-government development and 73rd in AI readiness. While these rankings yield more favorable results for Türkiye, they do not reflect similarly positive developments for Iran. At the global level, these indicators suggest that both countries still exhibit relatively limited preparedness for the effective and ethical implementation of artificial intelligence.

In terms of data protection, accountability, transparency, and explainability, Türkiye has taken initial institutional steps; however, it continues to face legal gaps, particularly with respect to AI-specific regulation. Iran, on the other hand, lacks comprehensive data protection legislation and formal oversight mechanisms, further increasing risks related to privacy, accountability, and algorithmic bias.

Overall, while both countries face significant ethical and legal challenges—such as privacy protection, accountability, explainability, and justice—Türkiye demonstrates greater progress and closer alignment with international standards. The comparison reveals that the effective integration of artificial intelligence into e-government depends not only on technological adoption but also on robust legal frameworks, institutional oversight, and respect for fundamental rights.

Conclusion

This study has examined the ethical and legal challenges associated with the use of artificial intelligence in e-government through a comparative analysis of Türkiye and Iran. The findings indicate that both countries have increasingly integrated AI into public administration—particularly in healthcare, education, smart city management, and law enforcement—to improve efficiency, service delivery, and state-citizen interaction. However, this technological adoption simultaneously brings about significant ethical and legal challenges. The key challenges identified are data protection, privacy, transparency, accountability, explainability, algorithmic bias, and justice, which remain central concerns in both contexts.

Regarding how the two countries differ in addressing these challenges, Türkiye exhibits partial alignment with international standards through instruments such as the Personal Data Protection Law (KVKK) and national AI strategies, though it still lacks a binding AI-specific legal framework. Iran, in contrast, has a more fragmented legal environment characterized by the absence of comprehensive AI-specific legislation and limited institutional oversight. Overall, while AI offers substantial opportunities for modernizing public administration, both countries require stronger rights-based legal frameworks and robust ethical governance mechanisms to ensure its effective, trustworthy, and responsible use in e-government.

Recommendations

Based on the study's findings, it is emphasized that significant steps must be taken in both countries to protect human rights in the digital age and to establish comprehensive legal frameworks for artificial intelligence. In Türkiye, to enhance the protection of citizens' rights and privacy, AI-specific legal frameworks like the Personal Data Protection Law should be developed, and existing legal structures should be expanded in alignment with the GDPR. On the other hand, Iran should pursue policies closer to international standards—such as those reflected in the GDPR and the European Union Artificial Intelligence Act (AI Act)—and enact comprehensive, AI-specific legal frameworks.

Both countries should institutionalize explainability and transparency in the use of artificial intelligence in public governance. Citizens must be informed about how administrative decisions are made and processed. Furthermore, comprehensive oversight mechanisms should be established in both countries, including authorities responsible for monitoring artificial intelligence systems and institutional bodies to oversee decision-making processes. When violations arising from algorithmic decisions are identified, effective appeal mechanisms should be activated and corrective actions are implemented. As artificial intelligence continues to expand, it will generate increasing legal and regulatory challenges. Adapting this process entails significant legal and institutional responsibilities for both countries. In addition to robust legal frameworks, greater emphasis should be placed on ethics of education in artificial intelligence. Public officials should receive training in AI ethics, and specialization in this field should be actively encouraged.

References

1. Karataş, A. (2025, February 22–23). Yapay zekânın kamu kurumlarının idari kapasitesine etkisi: Ulusal ve uluslararası literatür çerçevesinde bir analiz.Paper presented at the International Colosseum Scientific Research and Innovation Congress, Rome, Italy.
2. Abdolhamid, M., Lari, M., & Najafi Rastaghi, H. (2024). A comparative study of global experiences in the application of artificial intelligence in public administration: Implications for enhancing smart governance in Iran. Journal of Iranian Public Administration Studies, 7(3), 189–226.
3. Iranian Government. (n.d.). National Smart Government Services Portal. Retrieved December 11, 2025. from https://my.gov.ir
4. Organisation for Economic Co-operation and Development. (2025). Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing. https://doi.org/10.1787/795de142-en
5. Savelli, I., Rigou, M., & Balaskas, S. (2023). From e-government to AI e-government: A systematic review of citizen attitudes. Information, 14(2), Article 98. https://doi.org/10.3390/info14020098
6. European Commission. (2019). Ethics guidelines for trustworthy artificial intelligence. High-Level Expert Group on Artificial Intelligence.
7. Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence. OECD Publishing. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
8. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000381137
9. 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

10. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
11. Nissenbaum, H. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.
12. Bhat, A., et al. (2024). Accountability and governance challenges of artificial intelligence in public sector enterprise systems. International Journal of Emerging Trends in Computer Science and Information Technology, 5(1), 114–128.
13. Dwivedi, Y. K., Hughes, L., Ismagilova, E., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
14. Roshani, A., & Kandelousi, N. S. (2025).Prerequisites for embracing artificial intelligence in Iranian government agencies: A roadmap for transformative governance. SSRN. https://doi.org/10.2139/ssrn.5098035
15. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
16. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.
17. Özlü, R. M. (2025). Artificial intelligence and public administration in Türkiye: A descriptive analysis. Journal of Artificial Intelligence, 2(4), 299–310. https://doi.org/10.61552/JAI.2025.04.005
18. Oxford Insights. (2025). Government AI Readiness Index 2025. https://www.oxfordinsights.com/government-ai-readiness-index
19. United Nations. (2024). United Nations e-government survey 2024: Accelerating digital transformation for sustainable development. United Nations Department of Economic and Social Affairs. https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2024
20. Wikipedia contributors. (n.d.). E-government. https://en.wikipedia.org/wiki/E-government
21. T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2021). Ulusal Yapay Zekâ Stratejisi 2021–2025. Ankara.
22. Güneş, B. S. (2025). Artificial intelligence and emerging technologies in Türkiye: Legal adaptation from a constitutional and public law perspective. Journal of Artificial Intelligence and Human Sciences.
23. Marzdar, M. H. (2024). Artificial intelligence in Iran's public administration: Opportunities, challenges, and strategic approaches for governance innovation. International Journal of Management, Economics and Administrative Sciences, 38. https://doi.org/10.63053/ijmea.38
24. Kayacı, M. (2025). Kamu yönetiminin şeffaflığından algoritmaların şeffaflığına geçiş: Şeffaflığı, algoritmaların opaklığı ve kara kutu problemi çerçevesinde yeniden değerlendirmek. Abant Sosyal Bilimler Dergisi, 25(2), 827–841. https://doi.org/10.11616/asbi.1658320
25. Torabi, M. A., & Eghbal, H. (2024). Proposing an artificial intelligence governance model for state administration in the Islamic Republic of Iran. State Studies of Contemporary Iran, 4.
26. Mohebbi, D., & Amiri, A. (2024). Legal and ethical challenges related to the use of artificial intelligence in the administrative justice system. Faculty of Law, University of Qom.
27. Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. Government Information Quarterly, 40(1), 101774. https://doi.org/10.1016/j.giq.2022.101774
28. Organisation for Economic Co-operation and Development. (2019). Artificial intelligence in society. OECD Publishing. https://doi.org/10.1787/eedfee77-en
29. Aloamaka, P. C., & Omozue, M. (2024). AI and human rights: Navigating ethical and legal challenges in developing nations. Khazanah Hukum, 6(2), 189–201.

ALGORITHMIC INFLUENCE ON POLITICAL INFORMATION TRUST

RANIM ELHASAN

Abstract
Purpose*: The purpose of this study is to examine how social media algorithms shape the political information environment and users' trust in political information, focusing on awareness, perceived bias, content repetition, and trust in information.*
Methodology*: This study adopts both descriptive and analytical methods, and it involves analysis of data from field observations, together with an analysis of past literature. The data was obtained via an online survey answered by 30 users in a post-conflict society. The survey consisted of closed and open-ended questions.*
Results and Conclusions: *The results indicate that in post-conflict contexts, users' exposure to political content is significantly shaped by social media algorithms. However, despite most participants reporting awareness of algorithmic curation, this did not in any way make them immune to its operations. Rather, the repeated exposure to similar topics, less visibility of alternative points of view, and perceived bias in promoted content were repeatedly pointed out by the respondents. Most commonly, algorithmic influence was associated with increased scepticism and distrust toward political information rather than producing direct opinion change. In many cases, participants also believed that algorithms affect users unconsciously, even when they did not perceive themselves as personally influenced. Overall, these findings suggest that awareness alone is not enough to counter algorithmic influence. Where social media serves as the primary political information source in a post-conflict society, algorithms can subtly shape political perceptions and damage faith in the information ecosystems.*
Research limitations: *The sample size was limited (n = 30), which restricts generalizability.*
Practical and/or Social implications: *the importance of improving digital media literacy skills, awareness concerning recommendation algorithms, as well as social media transparency, especially within a fragile political environment, has been brought to the fore.*
Originality: *This study is original in focusing on user perceptions of social media algorithm influence in a post-conflict society.*

Keywords: *social media algorithms, political perception, post-conflict society, media scepticism, content curation*

Introduction

Rapid digital developments indicate that social media has become a central platform for receiving political information and participating in public debates. These platforms rely on complex algorithmic systems that rank, filter, and personalize content based on users' interaction patterns and digital behaviour, thereby shaping what information is made visible or invisible to users (Gillespie, 2014; Pariser, 2011) raising questions about the impact of these algorithms on users' awareness and the formation of their political knowledge.
Algorithms act as "digital gatekeepers", who decide what content users see and what is kept invisible (Gillespie, 2014). (Pariser, 2011)described "filter bubbles" as a phenomenon in which personalization algorithms intellectually isolate individuals by only showing them belief-conforming information and filtering out the opposing viewpoints. Previous research indicates that such algorithmic systems may contribute political bias, reduce content diversity, and the formation of "echo chambers" in which users are exposed primarily to opinions like their own (Sunstein, 2017).
However, the preponderance of the current literature concerns more established Western democracies, where the press as an institution maintains some kind of baseline level of public trust and the information ecology is more established (C. Sandvig et al., 2014; Tufekci, 2015). This is problematic because the effects of algorithms might have different impacts in a frail political setting. In a conflict-ridden country, conventional institutions face challenges when trying to establish these much-needed levels of public trust and credibility because they were involved in propaganda efforts in the past or because of current factional differences within the same institutions. This makes social networks essential, and in most cases, the primary source for these citizens in terms of getting political information (Mancini, 2013; Zeitzoff, 2017).

Research problem

While there has been increased attention to the role of algorithms in recent work on political content, how such algorithmic influences are perceived by users in post-conflict societies is little known. This research addresses this shortage through a survey investigation of user awareness, perceived bias, content diversity, and trust in political information. The lack of understanding is more pertinent because post-conflict societies remain highly susceptible to the impact of algorithms due to the presence of several risks, including sectarian conflicts, the lack of consensus on historical facts, the ill-equipped media landscape, and low digital literacy.
This gap limits our understanding of how algorithmic curation shapes political awareness and trust in fragile information environments.

Research question

This study addresses the following research question: **How do social media algorithms shape the political information environment and users' trust in post-conflict societies?**

Objectives of the research

To answer this question comprehensively, the study pursues four interconnected objectives:

- Assess awareness among users about algorithmic content curation in the post-conflict context.
- Examine perceived algorithmic bias and its effects on content diversity.
- Assess self-reported perceptions of algorithmically curated content in relation to political opinions and information trust.
- Identify specific mechanisms through which people experience algorithmic influence.

Theoretical framework

This study draws on theories of soft power to conceptualize algorithmic influence as an indirect mechanism shaping users' political perceptions. Soft power emphasizes the ability to influence preferences and perceptions through attraction and subtle steering rather than coercion (Nye, 2004) . In the context of social media, algorithmic content curation functions as a form of mediated soft power by selectively amplifying certain political narratives while marginalizing others. This framework enables an analysis of how users in post-conflict societies perceive algorithmic bias, content diversity, and trust in political information, particularly within fragile information environments characterized by low institutional trust and limited media literacy.

The importance of this study

The study enables a clear understanding of the implications of social networking algorithms on the users' reception of political messages and trust in the information in a post-conflict state. The research has the potential to aid in the formulation of better media policy and the promotion of online transparency to educate users of political rights.

Literature review

Algorithmic content curation: mechanisms and functions

Social media algorithms are also acting as online gatekeepers in that they regulate what content reaches users in their online feeds based on the predicted response(Gillespie, 2014). They achieve this through filtering, prioritizing, or personalizing content with the help of machine learning algorithms that are based on user behaviour (Bucher, 2018). Notably, these processes are not transparent to users but have created a concept called "algorithmic opacity" (Burrell, 2016).

In fact, (Eslami et al., 2015)reported that 62% of users on Facebook were unaware their feeds were curated algorithmically and perceived seeing everything posted by their connections. The lack of awareness in this context is highly consequential for democratic discourse, as users may not realize how their political information environment is sculpted by the logics of engagement-maximization rather than journalistic standards.

Filter bubbles and echo chambers

(Pariser, 2011)coined the "filter bubble" idea in relation to how personalization algorithms are responsible for creating "intellectual isolation" through the provision of "belief-confirming" information and the "filtering out" of opposing viewpoints. The author anticipated that this could lead to the destruction of the "shared experience" that is essential to democratic deliberations through the fragmentation of the public space through personal "information bubbles."

Evidence from empirical studies paints a complex picture. (Bakshy et al., 2015)examined 10.1 million Facebook users. They identified a 5% decrease in the chances of being exposed to conflicting political views through algorithmic rankings, but the results were dominated by individual choice with a 17% decrease. (Flaxman et al., 2016) discovered that users of the social media platforms were 7-10% more ideologically secluded than users who directly accessed the news.

Despite this, not all studies agree with the concept of the filter bubble. Drawing on their findings, both (Dubois & Blank, 2018)showed users of social media are exposed to more diverse ideas than those not on social media, whereas(Guess et al., 2018)showed users are exposed to extreme content, albeit in small proportions. Such discrepancies indicate findings related to algorithms may not always hold up to initial hypotheses.

Algorithmic bias and transparency

Aside from filtering effects, there are also systemic biases that are discovered to exist in algorithmic systems. It was suggested that engagement-driven algorithms are predisposed to prioritize (or favour) one type of content over another—one that has potential for high engagement rates through "outrage and fear"(Tufekci, 2015).

An example of how a lack of understanding of algorithms' impacts on inequality is provided by (C. Sandvig et al., 2014), who show how algorithms encode inequality based on biases within training algorithms or based on unforeseen algorithmic behaviours that an algorithm's creators did not anticipate. An opaque nature of algorithms adds to this set of problems, making it impossible to test an algorithm's claims of fairness, according to (Diakopoulos, 2015).

Algorithms in conflict and post-conflict contexts

There are currently very few studies that focus on the role of algorithms and their effects in conflict-affected regions. However,(Zeitzoff, 2017) argues that the algorithmic amplification of sectarian content in regions of conflict may occur prior to an outbreak of violence. This highlights that social media companies are "information accelerants." The case of Myanmar has shown that this could be the case as they found that Facebook's algorithms systematically amplify hate speech against the Rohingya Muslim community.

(Lynch, 2016) asserts that after the uprisings in the Arab world, social media networks contributed to the fragmentation of public discourse and the dissemination of sectarian narratives. Although social media sites serve as mobilizing avenues for actions, they also supported the easy dissemination of rumours and radical ideologies. According to (Mancini, 2013) , societies that have just come out of conflict possess distinct challenges such as weak institutions and heightened tensions.

Research gaps and study positioning

Despite increasing literature, there are challenging gaps:
Geographic Bias: Most of the studies are based on Western societies, with relatively few addressing the Middle East and Africa, and almost no studies focusing on the post-conflict transition period. It is expected that algorithm effects differ between political and cultural settings.
Methodological Limitations: Indeed, existing studies tend to focus on data analyses of platforms or controlled experiments, often neglecting users' points of views and subjective experiences. There is a largely uninvestigated realm concerning how users perceive algorithms' impact, mostly within a non-Western perspective.

Theoretical Gaps: There has been little research that has aspired to link algorithmic impacts to concepts such as soft power and digital rights. There has been an incomplete conceptualization of algorithmic curation and its impacts under the theories of freedom of expression and access to information.
Contradictory Results: Findings are inconsistent when comparing the effects of filter bubbles on studies (either Pariser or Bruns, or alternatively, Flaxman or Dubois & Blank).
This research fills these gaps by placing a focus upon user perspective within post-conflict societies and exploring how people living within a situation of political fragility view and interpret algorithmic mediation of political material.

Materials and methods

Study Objectives and Design

This study proposes the exploration of the effect of social media algorithms on political content considering experiences in post-conflict societies. A descriptive-analytical method was used, with both pre-prepared questions on a social media-based online survey and a review of existing literature. There are both advantages and limitations associated with using such a method.

The reason for choosing the descriptive-analytical method includes the following: Firstly, the political sensitivity of post-conflict nations renders experimental research morally and practically suspect of many participants. Secondly, research emphasis is placed upon perceptions and individual experiences, the data sought can only be obtained through this method of research. Lastly, it should be mentioned that resource constraints dictated that data collection must be quick and efficient, although it can only offer speculative information about a relatively novel topic.

Operational Definitions

The criteria for a post-conflict society for this study are: (1) experienced an armed conflict within the last 10–15 years; (2) having transitioned towards a definite end of conflict; and (3) finding themselves within a process of institution rebuilding and transition. The criteria were utilized for participant selection within the Syrian context following the 2011 Syrian revolution.

Algorithmic awareness refers to participants' awareness that algorithms play a decisive role in selecting, ranking, and displaying content on social media platforms, without necessarily understanding the technical mechanisms behind these processes.

Perception refers to the subjective concept that participants have of the way in which algorithms affect the political content they are exposed to.

Study Sample and Recruitment

The participants in the study where a total of 30 adult Syrian social media users actively consume political content.

Participants were recruited through direct distribution of the questionnaire link via Personal social media networks (Facebook, WhatsApp) and Snowball referral from initial participants.

The link to the questionnaire was sent to around 39 potential participants. Among them, 30 participants fully completed the survey, yielding a response rate of 77%.

Questionnaire Design

The questionnaire took the form of 10 questions, which included one open question and nine closed questions. Closed questions included single response (yes/ no), multiple-choice, and ordered options, such as hours spent on social media. The questionnaire had been designed and administered by the tool Google Forms and measured this: Daily social media usage (hours), Most frequently used social networks, most interacted content type, algorithmic curation awareness, perceived algorithmic bias, perceived algorithmic influence on political content, self-reported impact of content on political opinion, type/magnitude of effect, perceived unconscious influence of social media and qualitative feedback about algorithmic influence.

The questionnaire was conducted in Arabic to make it accessible to all participants and doing so helps eliminate the possibility of a language barrier being a factor in determining the quality of the responses given to a questionnaire. Average time to complete was approximately 2 minutes.

Data Collection and Ethical Procedures

Data collection was conducted over a short period from December 29, 2025, to January 9, 2026.

The questionnaire was anonymous and voluntary. Participants were given information about, Research goals, Voluntary Nature of Participation, Anonymity and Confidentiality in Responses and The Right to withdraw at any time.

No personally identifiable information was collected, Data stored securely and only Aggregated data reported in publications.

Data Analysis

Before embarking on analysis, the data was organized and cleaned for accuracy and consistency of responses. Moreover, all responses written in Arabic were further translated into English for easier analysis and reporting.

The results from the closed-ended questions were analyzed using Microsoft Excel with a view to conducting descriptive statistics to detect some levels of tendencies rather than correlations.

Specifically, the evaluation entailed:Computing the percentage and frequency of each category in the social media usage patterns, awareness of algorithmic filtering, bias, and self-perceived political influence variables. Table Organization of Responses into Summary Tables to Clearly Display Response Distributions.

The open-ended question allowed the collection of 9 substantive responses out of a total of 30 participants (30% response rate). The rest either did not respond or responded with a non-substantive response (e.g., "no" to the question), and the responses were subjected to thematic analysis (Braun & Clarke, 2006).

Study Limitations

This study is limited by its sample size and used convenience sampling which restricts generalizability. Findings rely on Self-reported perceptions rather than behavioral data. The inclusion of only one open-ended question constrained qualitative depth, and the cross-sectional design prevented analysis of change over time, and lack of platform access limits assessment of actual algorithmic behavior.

Despite these limitations, the study provides initial insights into how users in post-conflict settings perceive the impact of algorithms in terms of political content.

Results and findings

This section presents the study's core findings addressing the central research question of how social media algorithms shape the political information environment and users' trust in post-conflict societies, with a focus on Syria.

Social Media Usage Patterns and News Consumption

Analysis of the usage pattern indicated that more than half the respondents (56.7%, n=17) exceeded 4 hours of social networking per day (Figure **1** here). Very few respondents (10%, n=3) indicated social networking for less than an hour.

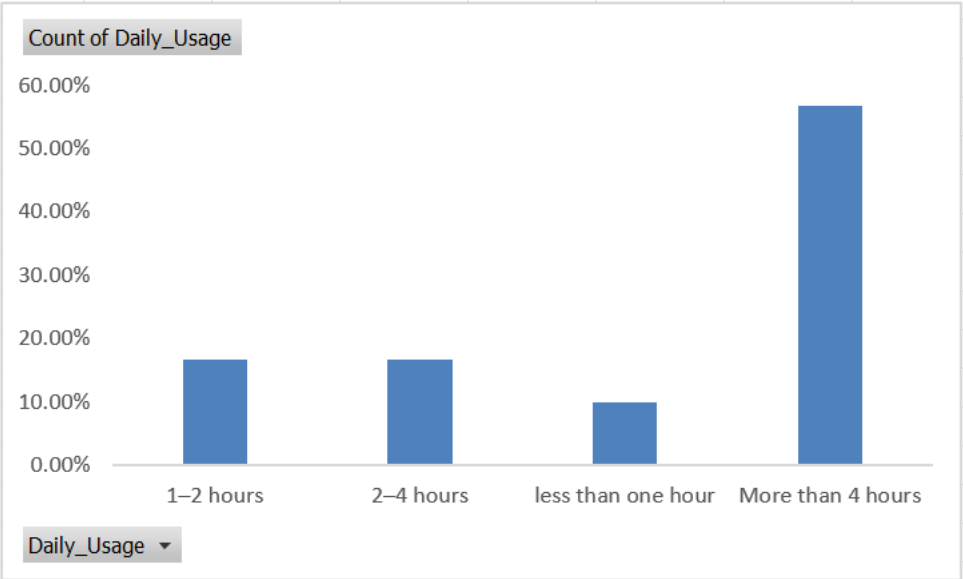


Figure 1. Social media usage patterns

This intensive usage pattern points towards massive user engagement with algorithmically filtered contents.
Regarding platform preferences Facebook came out as the most widely used social media (66.67%, n = 20). Instagram came in second (63.33%, n = 19), followed by TikTok (30%, n = 9) (Figure **2** here).

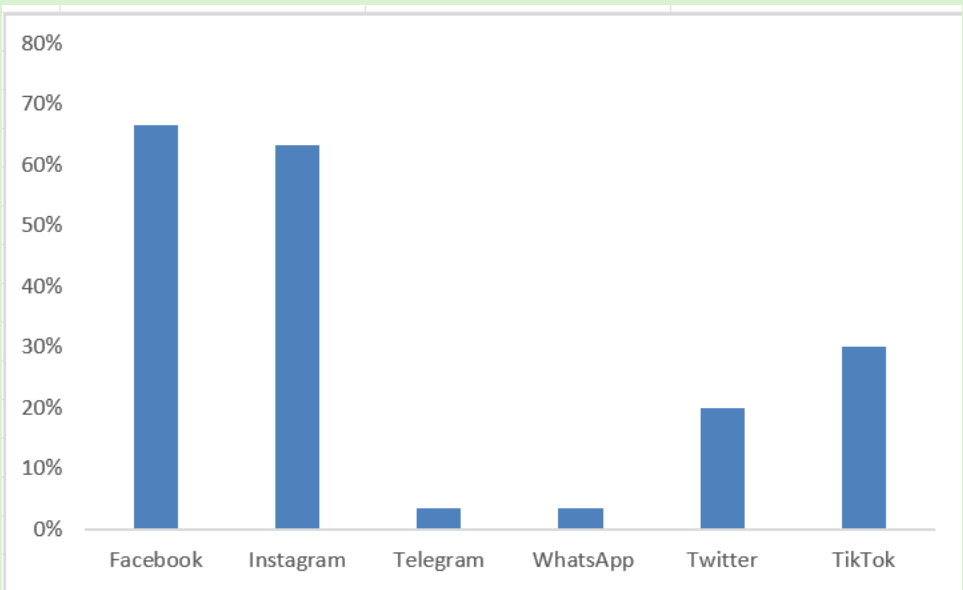


Figure 2. Distribution of platforms usage

The pre-eminence of Facebook and Instagram is quite surprising, especially considering that these services are so reliant upon content algorithms and recommendation systems.

Political content is a central focus of interaction for all participants on social media, with varying secondary interaction with educational or cultural content, general news, and entertainment content (Figure **3** here).

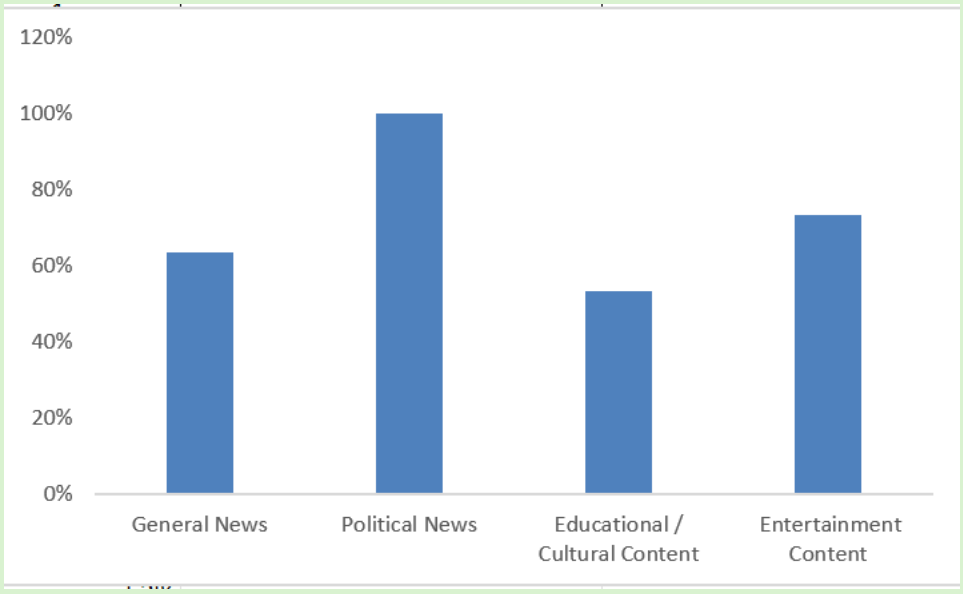


Figure 3.Content interaction patterns

This heavy reliance on social media for political information, particularly in post-conflict societies where institutional credibility is often weak creates significant vulnerability to algorithmic influence, as platforms become the primary lens through which users understand political events and issues.

Awareness of Algorithmic Content Curation

When they were asked whether they knew that algorithms rank and filter the content they would see, 60% of the participants reported knowing this, 26.7% were not sure, and 13.3% did not know (Figure **4** here).

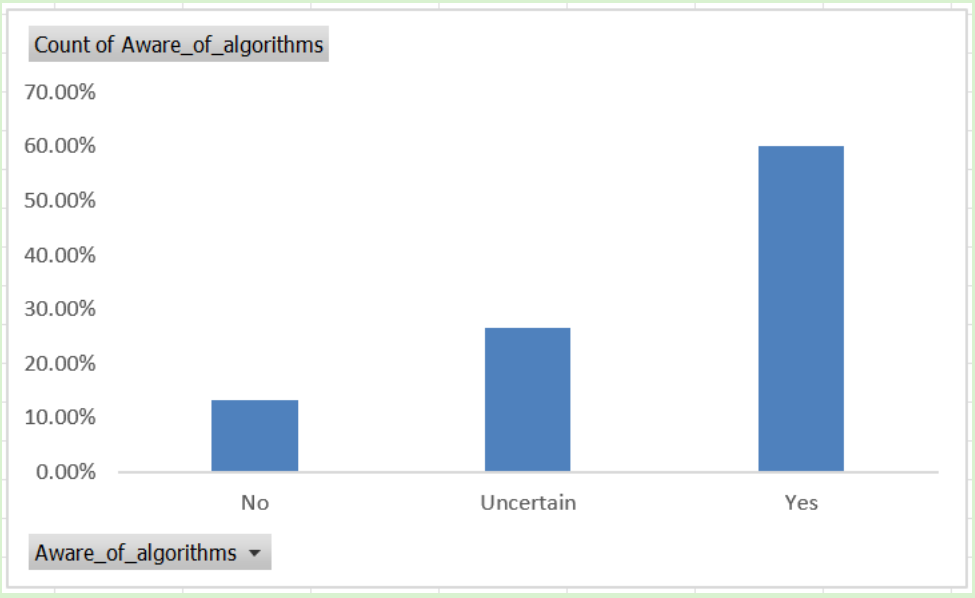


Figure 4. Awareness of algorithmic

While the majority demonstrated basic awareness that algorithmic curation exists, qualitative responses revealed that understanding of specific mechanisms remained limited. Several participants described knowing "something controls what I see" without comprehending how these systems function.

In addition to general awareness of algorithmic, participants were asked whether they had noticed specific patterns in the presentation of content. A substantial majority (80%) reported observing repeated exposure to certain topics (Figure **5** here), indicating a perceived pattern of topic repetition in their social media feeds, where some issues appeared more frequently than others.

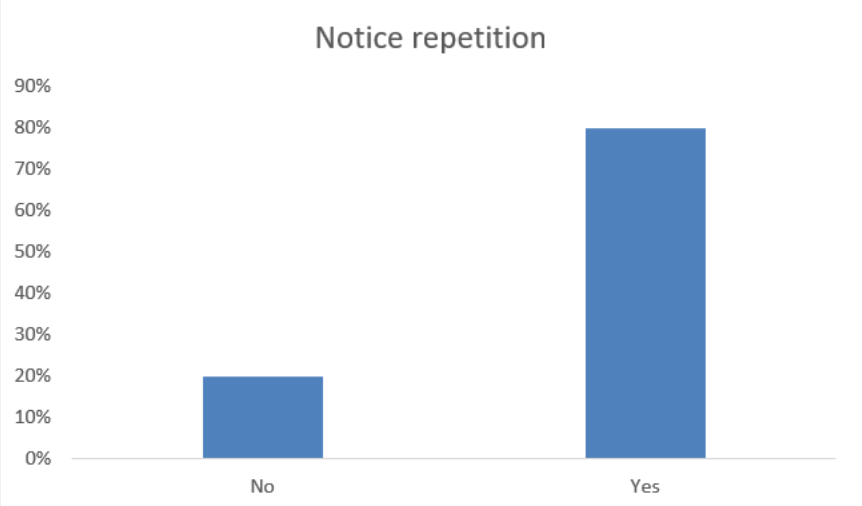


Figure 5. Notice content repetition

This suggests that while technical understanding may be limited, functional awareness of algorithmic effects manifested as noticing patterns in content visibility is relatively widespread. However, this awareness does not necessarily translate into protection from algorithmic influence.

Belief in Algorithmic Agenda-Setting

A critical finding emerged when participants were asked whether they believe algorithms rank content in ways that serve agendas or influence public opinion. 60% responded affirmatively (Figure **6** here), believing that algorithmic curation is not neutral but rather shaped by specific interests, whether commercial, political, or ideological.

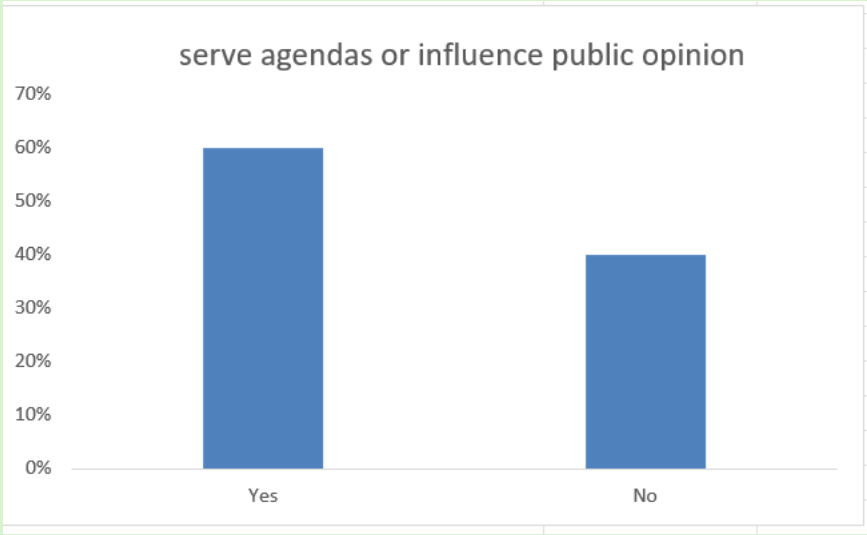


Figure 6. Belief in algorithmic agenda

This widespread scepticism about platform neutrality is particularly significant in post-conflict contexts, where suspicion of manipulation and propaganda often runs high due to historical experiences with information control and media bias.

Self-Reported Impact on Political Opinions and Information Trust

Participants reported mixed yet meaningful impacts of algorithmically curated content on their political attitudes. Nearly half of the participants (n = 14) indicated that such content had influenced their political views (Figure **7** here), while 8 participants reported no perceived influence and an additional 8 were uncertain. Among those who reported an influence, the effect was more commonly described as an erosion of trust rather than a direct change in political opinions.

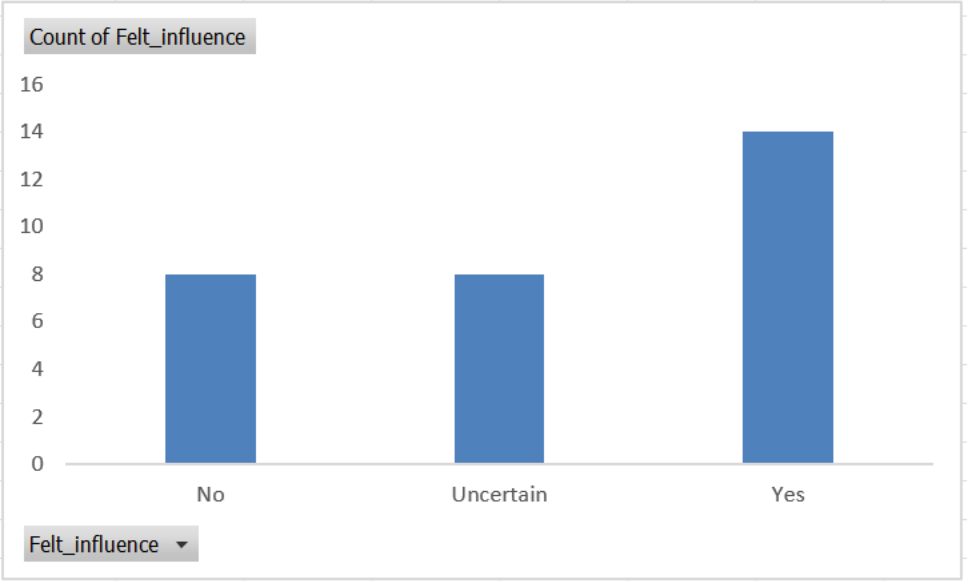


Figure 7. Felt influence

These findings suggest that, despite reported awareness of algorithmic curation (60%, see Section Awareness of Algorithmic Content Curation), participants frequently acknowledged feelings of bias or reduced trust in the information presented to them (Figure **7** here)This indicates that awareness may not be sufficient to prevent subjective experiences of algorithmic influence, even if it does not necessarily result in direct opinion change.

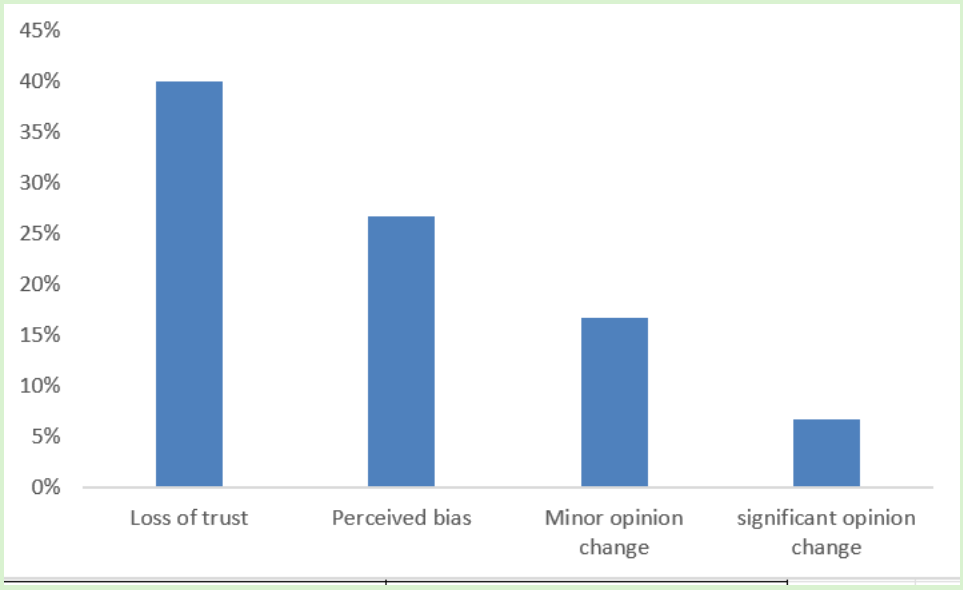


Figure 8. Type of influence

When asked how algorithms and the content they promote have affected their political opinions, responses were distributed across multiple categories (Figure **8** here), with increased distrust of information emerging as the single most common impact

Type of influence presents the same information visually: whereas a total of 23.3% of participants reported direct opinion change - significant or slight - the most prevalent single effect was increased scepticism toward information, at 40%.

This is a finding particularly consequential for post-conflict societies, where opinion shifts are visible and potentially reversible, whereas erosion of information credibility operates more insidiously in undermining shared epistemic foundations necessary for democratic reconciliation and public discourse. This would mean the implication goes beyond individual attitude change to broader questions of social cohesion and democratic functioning when one-third of users report increased distrust as the primary algorithmic effect.
The use of increased bias toward certain topics (26.67%) as a category further indicates that the phenomenon of algorithmic amplification may augment already existing biases without actually leading to a shift in perceptions, as would be expected to happen under affective polarized dynamics, where individuals' emotional views about issues become more polarized, even if their policy stances do not change.
A striking finding emerged when participants were asked whether algorithms influence users' opinions without their awareness. Although several participants claimed that they personally did not feel influenced by algorithmic systems (Figure **9** here), a majority of the sample (53.3%) believed that algorithms can exert influence unconsciously.

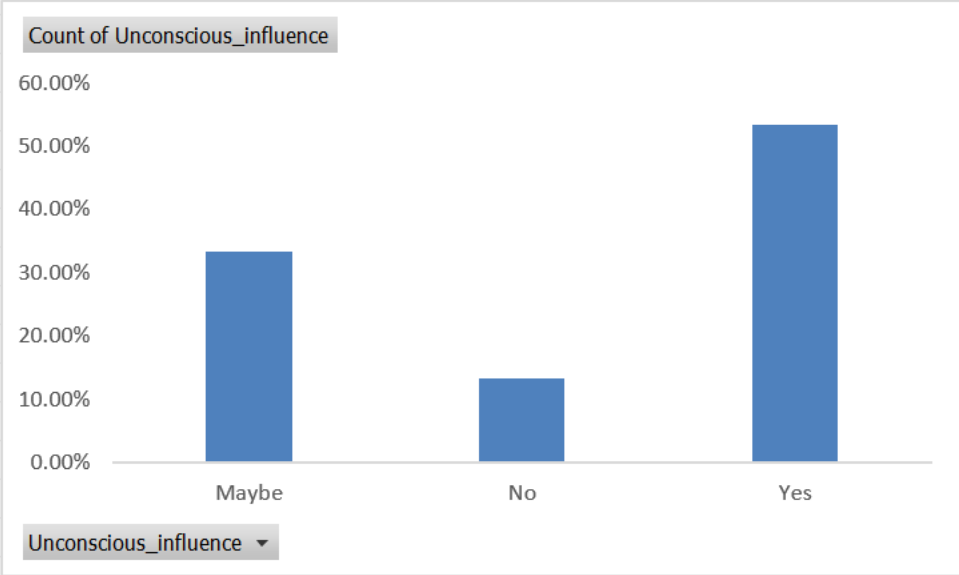


Figure 9. Unconscious influence

This generates a paradoxical situation, wherein among the 26.7% of the respondents (n=8) who did not see any personal effects, 3 (37.5%) believed that algorithms have unfelt effects on others. Such a disconnection between the belief about the effect of the media on society and the lack of effect on the respondents is the typical "third-person effect" observed in studies on media effects, where the researcher acknowledges the effect on society but is not affected.

In sum, 53.3% of respondents answered that they believed in the existence of unconscious algorithmic influence, while 33.3% answered "Maybe." The results of this study call into question assumptions of many digital literacy education programs that assume awareness of algorithmic selection enable one to escape its influence. One can be aware of the presence of algorithms and their subliminal power yet still be affected by them.

Mechanisms of Algorithmic Influence: Qualitative Insights

The open-ended question generated nine substantive responses (30% response rate). This was quite effective because it gave deep insights into the way in which the participants felt the impact of algorithms without relying on presence or absence.
Thematic analysis identified five perceived mechanisms of algorithmic influence:

Theme 1: Cumulative Persuasion Through Repetition (33.3%)
This has all been echoed by "Illusory Truth Effect," whereby repetition enhances perceived truths rather than accuracy. The quote below illustrates this: ["The frequency with which a topic appears makes us believe it's true, even if we initially disagreed"]

Theme 2: Selective Amplification and Suppression (22.2%)
The participants attributed algorithms to work in the following manner in terms of certain narratives while hiding others:
["Algorithms influence user awareness without clarity, especially in politics; they promote some narratives and hide others"]
This perception aligns with the 60% who believe algorithms serve agendas (Section 3.3).

Theme 3: Personalized Echo Chambers (22.2%)
Participants recognized that content depends on previous interactions, limiting viewpoint diversity: ["I felt biased toward a certain political topic because all the news on Facebook talked about it and prevented me from seeing differing opinions"]

Theme 4: User Agency and Resistance (11.1%)
A minority emphasized maintaining autonomy despite exposure:
["I don't think a person can be influenced by a post if they already have an opinion and are convinced of it; one sees the content, but the decision remains theirs"]
This corresponds with the 13.3% reporting no impact quantitatively.

Theme 5: Contextual Complexity (11.1%)
One participant noted that algorithmic effects interact with other factors: ["It might have a slight influence, but the impact isn't solely due to algorithms; there are other factors like real world events"].
The convergence of quantitative findings and qualitative themes strengthens confidence that the perceived effects of algorithms are consistent across different measurement approaches.

Discussion

The present research investigated how social media algorithms affect users' perception of content in a post-conflict society. The results show a critical paradox: whereas 60% of the users were aware of algorithmic curation, they nonetheless reported significant effects, including content repetition (80%), perceived bias about topic-specific issues (26.67%), and increased information distrust (40%).

Awareness Without Protection

A key contribution of this study is the identification of a gap between awareness and effective resistance. Despite 60% of participants knowing their content is algorithmically ranked, participants still showed topic-specific bias, and opinion effect. This flies in the face of the premise of most digital literacy efforts, namely, awareness leads to resistance.
This is like the findings of(Eslami et al., 2015), where awareness of the Facebook filter algebra did not enable the users to effectively manipulate the situation. (Rader & Gray, 2015)have identified “algorithmic confusion,” where there is awareness of what is happening, but no comprehension to enable control of the situation.
However, our results suggest that awareness alone does not necessarily lead to seeking more diverse information. This highlights the importance of context to the potential role of awareness as protection, where traditional means may be less available. Although all participants predominantly engage with political content on social media, awareness of algorithmic curation does not appear to fully protect them from its effects. In post-conflict contexts, where traditional institutions may be fragmented, users remain exposed to algorithmically prioritized content regardless of their understanding.

Trust Erosion as Primary Impact

Information distrust was the most frequent reported consequence, even exceeding slight and substantial shifts in opinion. This defies models of persuasion or polarization. Our results demonstrate that the greatest potential impact of an algorithm might come not from persuasion but epistemic disruption less about changing beliefs and more about delegitimizing information in the first place.
This is consistent with studies on the propaganda strategies of "chaos," which seek to create general confusion rather than moving the target audience toward specific ideas (Paul, 2016).While previous studies have focused on such phenomena as state-sponsored propaganda strategies, our results indicate such effects may occur incidentally as part of commercial algorithms designed to maximize engagement.
The issue of trust decay is more significant for post-conflict societies, where social trust in information systems is a basic requirement of democratic functionality. When algorithmic systems decay such trust, they impinge upon a shared epistemic base that is a prerequisite of reconciliation, deliberation, and joint problem-solving, the key ingredients of democratic transitions.
The result that 53.3% have a belief that algorithms impact users unconsciously demonstrates "third-person effect" because, as noted by(Davison, 1983), people perceive that media affects third persons and not oneself: This has methodological implications, that effect measures based on self-reported impacts could potentially underestimate the actual influence, since the users could have real, yet unconscious, influence that they are not aware of. Future research could rectify this problem and incorporate behavioural measures or experimental methods that seek to detect influence that the users themselves are not aware of.

Mechanisms of Influence in Post-Conflict Contexts

The most frequent qualitative theme identified was how repeated exposure gradually influences beliefs, even if the users a priori disagreed. This phenomenon is then associated with the "illusory truth effect": repeatedly exposing people increases perceived credibility irrespective of actual accuracy.

Our contribution is a demonstration of how this mechanism interacts with post-conflict contexts. In environments where there are rival narratives about recent violence and contested historical accounts, the algorithmic amplification of certain interpretations while suppressing alternatives serves to constitute collective memory in forms supportive of, or undermining, reconciliation. This is quite unlike algorithmic polarization in stable democracies, where the stakes rarely include possible return to violent conflict.

The perceived notion that algorithms "promote some narratives and hide others" (22.2%) extends agenda-setting theory (McCombs & Shaw, 1972)to the algorithmic age. 60% of the participants think algorithms that serve specific agendas are important considering the participant context where direct experiences with propaganda and information control would justify scepticism about neutrality.

However, our data does not reveal whose agenda users. This represents a limit that needs to be investigated in the future.

The echo chamber theme (22.2%) shows that there is awareness of how personalization reduces the diversity of views, in accordance with (Pariser, 2011; Sunstein, 2017) works.

For post-conflict environments, even the presence of echo chambers with fractional influence could have real-world effects. A slight reduction in overlap could influence reconciliation exercises that necessitate empathy and understanding across conflict lines.

Post-Conflict Contexts: Platform Dependence and Institutional Fragility

Users rely on social media for political content because traditional news media are controlled and divided along political and factionalist lines or lack credibility due to their past roles in propaganda. In the process, they find themselves in engagement-algorithm systems. In democratic states with varied media, they may maintain equilibrium against algorithmic manipulation; in post-conflict states are more vulnerable.

This helps to explain the difference between our findings and Western-context studies. Although (Boxell, 2017)discovered the U.S. saw the largest increase in polarization among groups who use social media the least, a heavy reliance on our participants' social platforms indicates that algorithms have a more prominent role to play post-conflict.

"Information accelerant" risk is relevant for post-conflict societies, since inflammatory content can facilitate rapid mobilization and increase the risk of renewed violence, especially in zones with high levels of tensions (Zeitzoff, 2017). Algorithms that "privilege" some narratives over another risk exacerbating differences of collective memories between communities along sectarian lines.

Nevertheless, the participants did not point towards the relationship between the influence of algorithms and the danger of sectarianism. This absence may reflect reluctance to people, which may be averse to answering sensitive questions in survey research, or limitation in questionnaire design. Future research needs to test if there is a relationship between the effect of algorithms and sectarianism.

Theoretical

The present study adds to the knowledge of algorithms as soft power tools (Nye, 2004) as they affect users' preferences through subtle, non-forced ways. The fact that 53.3% of respondents believe in an unconscious influence and yet many respondents feel personally exempt, is an instance of how the influence is exerted below the level of conscious resistance.

Nevertheless, when 60% think they promote agendas, it is not a case of unquestioning influence on the part of the populace. Rather, a full theoretical explanation will need to recognize the interplay of algorithmic influence and human agency as a dynamic and disputed process.

Limitations

Sample Size and Composition: The study relied on a small sample(n=30) and was collected through convenience sampling, so there is a problem with generalization. The educational level of the sample is high, which is again selected with bias toward people with digital literacy knowledge, so their perception regarding the effect could be different.

Self-Report Bias: All data are based on perceptions and not facts. Users may not be able to accurately view or respond to the effects of algorithms. Third-party effects, social desirability, and recall biases are possible in responses.

Cross-Sectional Design: One-time data fails to establish causality or reflect perceptions over time. Prevalent attitudes influence hindsight perceptions as opposed to algorithm exposure.

Lack of Algorithmic Ground Truth: Without access to actual algorithms and content feeds, one cannot test whether perceived biases are indicative of algorithms, social networks, or selective attention.

Alternative Explanations

Certain observed effects could be unrelated to algorithms. Confirmatory bias might enable users to identify with contents that affirm their biases. Social media effects could affect content by making observations based on friend/following and existing division within post-conflict societies.

The general effects of media (e.g., distrust or influence) might also be a factor. Although there are findings in the qualitative results concerning the awareness of mechanisms (personalization and repetition), other explanations cannot be ruled out in their entirety.

Future research work is expected to equate variables confounding experiments in isolated algorithmic effects.

Practical implications/Original value

Practical implications

For Digital Literacy Programs: Awareness and education must incorporate ways for successful navigation from information, for instance, searching for information from different sources, recognizing the effect of Illusory Truth, and checking Improve context-sensitive information for dealing with sectarian threats and structural challenges in the post-conflict context. Address different audiences rather than focusing only on educated and urban populations.

For Platform Companies: Enhance transparency through making visible to users the reasons for certain posted content being chosen while certain contents are being hidden. Implement contextual moderation, taking into consideration that algorithms which maximize user engagement can be a danger in a conflict environment. Provide for audits of algorithms in a politically unstable environment.

For Policymakers and Civil Society: Develop frameworks that require accountability and transparency in algorithms within a politically sensitive context.

Original Value of the Study

This study makes a valuable contribution to existing literature because of its emphasis on understanding the user perspective and experience of algorithmic influence, rather than algorithmic systems themselves. Another strength of this study lies in choosing to investigate a post-conflict country, since current literature tends to generalize findings from a stable country setting. This study reveals that being informed about algorithms will not safeguard against algorithmic influence.

Conclusion

This research provides a definitive answer to the research question by proving that social media algorithms do affect political perception in post-conflict societies through trust degradation and not opinion modification. Despite the fact that a majority of the participants (60%) claimed to be aware of algorithmic curation and agenda-driven bias, the research indicates that such an understanding is not a buffer against algorithmic influence. Through the empirical research, the participants' distrust of the information (40%) is more widespread compared to opinion modification (23.3%), indicating that algorithmic influence is achieved through epistemic degradation and not opinion modification. This is achieved through cumulative exposure, selective amplification of narratives, and the reinforcement of echo chambers.

Given the level of usage of social media, which is extremely high in post-conflict societies where 56.7% of the participants used social media for more than four hours a day, the algorithms act as a gatekeeping mechanism in the absence of alternative sources of information that can be trusted. In this manner, social media algorithms play a political role in influencing information environments with high levels of distrust.

In conclusion, the empirical findings of this research prove that in post-conflict societies, social media algorithms exercise political influence through epistemic degradation, placing platform design and algorithmic governance at the forefront of democratic transition in the digital era.

References

1. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160
2. Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, *114*(40), 10612–10617. https://doi.org/10.1073/pnas.1706588114
3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
4. Bucher, T. (2018). *If…then: Algorithmic power and politics*. Oxford University Press.

5. 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
6. Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, *47*(1), 1–15. https://doi.org/10.1086/268763
7. Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411
8. Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729–745. https://doi.org/10.1080/1369118X.2018.1428656
9. Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Volpe, A., Kluver, K., & Cotter, S. (2015). I always assumed that I wasn't really that close to [her]: Reasoning about invisible algorithms in news feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 153–162. https://doi.org/10.1145/2702123.2702556
10. Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006
11. Gillespie, T. (2014). *The relevance of algorithms*. https://doi.org/10.7551/mitpress/9042.003.0013
12. Guess, A., Nyhan, B., & Reifler, J. (2018). *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign*.
13. Lynch, M. (2016). *The New Arab Wars: Uprisings and Anarchy in the Middle East*.
14. Mancini, P. (2013). *Media fragmentation, party system, and democracy*.
15. McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *The Public Opinion Quarterly*, *36*(2), 176–187. https://doi.org/10.1086/267990
16. Nye, J. S. (2004). *Soft power: The means to success in world politics*. PublicAffairs.
17. Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
18. Paul, C., & Matthews, M. M. (2016). *The Russian "Firehose of Falsehood" propaganda model: Why it might work and options to counter it*. https://doi.org/10.7249/PE198
19. Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 173–182. https://doi.org/10.1145/2702123.2702174
20. Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms.* (pp. 1–23).
21. Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Data and discrimination: Converting critical concerns into productive inquiry*.
22. Sunstein, C. R. (2017). *Republic: Divided democracy in the age of social media.* Princeton University Press.
23. Tufekci, Z. (2015). *Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency*. *13*(203), 203–218.
24. Zeitzoff, T. (2017). *How social media is changing conflict*. 1970–1991. https://doi.org/10.1177/0022002717721392

THE HIDDEN COST OF INNOVATION IN THE DIGITAL MARKETPLACE

TALHA QAISAR

Abstract

Purpose: This paper examined the ethical, legal, and economic consequences of algorithmic dynamic pricing, with a focus on how personal data had been repurposed as a tool for extracting an individual's maximum willingness to pay. The authors argued that "surveillance pricing" had shifted digital markets away from fair competition toward mechanisms that enabled consumer surplus extraction, tacit collusion, and discriminatory pricing outcomes. The purpose of the study was to expose governance gaps and propose an oversight framework capable of restoring transparency and fairness in data-driven pricing systems.

Methodology: The study employed a qualitative critical analysis grounded in comparative case studies and legal review. High-impact incidents—such as Uber's surge pricing during the Sydney siege, Orbitz's device-based hotel pricing, and the U.S. Department of Justice's antitrust action against RealPage—were analysed against existing regulatory standards, including antitrust law, civil rights provisions, and the GDPR. A Risk Control Self-Assessment (RCSA) approach was applied to identify structural points of failure, which informed the adaptation of the "Three Lines of Defence" audit model for algorithmic governance.

Results and Conclusions: The findings revealed three major risks: (1) a systemic shift from market-based pricing to individualized reserve-price extraction; (2) the emergence of an "empathy gap," where algorithms misinterpreted desperation-driven demand spikes as opportunities for profit; and (3) tacit collusion facilitated through shared third-party pricing software. The study concluded that algorithmic efficiency did not justify discriminatory or anticompetitive outcomes and that organizations urgently required structured oversight to mitigate legal, ethical, and reputational risks.

Research limitations: The study relied on qualitative case studies rather than proprietary algorithm audits, limiting access to internal pricing logic. The focus on prominent cases may have overlooked subtler discriminatory effects occurring in smaller markets. Legal analysis cantered primarily on U.S. and EU regulatory frameworks, leaving implications for under-regulated regions less explored.

Practical and/or Social implications: Practically, the findings suggested that firms deploying opaque pricing systems faced litigation exposure, reputational damage, and erosion of consumer trust. Socially, the results implied a growing risk of digital discrimination where vulnerable consumers paid more due to proxy profiling. Algorithmic auditing and impact assessment were identified as prerequisites for fair digital markets.

Originality: This paper reframed dynamic pricing as a structural antitrust risk rather than merely a privacy concern. Its primary original contribution was the adaptation of the "Three Lines of Defence" audit model for algorithmic pricing, offering a concrete governance framework that operationalized ethical principles into actionable internal controls.

Keywords: data-driven pricing; consumer surplus; tacit collusion; audit framework; digital autonomy; governance ethics; market discrimination

Introduction

The traditional retail model, predicated on the transparency of the "fixed price tag," is rapidly being dismantled by the advent of Algorithmic Dynamic Pricing. Historically, economic literature has justified variable pricing through the lens of Yield Management—a practice largely accepted in the airline and hospitality industries based on aggregate supply and demand. However, the emergence of "Surveillance Pricing" represents a radical theoretical departure. Current research indicates that retailers now leverage Big Data and Artificial Intelligence not merely to manage inventory, but to exploit individual consumer surplus by analysing granular data inputs—such as device specifications, battery levels, and geolocation—to calculate a specific user's maximum "reserve price." (O'Neil, 2016).

Despite the commercial efficiency of these mechanisms, there remains a critical gap in understanding their ethical and legal ramifications. This paper investigates the hypothesis that current algorithmic practices have outpaced existing regulatory frameworks, creating a "Black Box" economy that violates the social contract of the marketplace. We posit that the transition from market-based pricing to personalized, data-driven pricing risks normalizing "tacit collusion"—where competitors unintentionally fix prices through shared software—and creates discriminatory "disparate impacts" that evade current civil rights and antitrust statutes. (Ezrachi & Stucke, 2016).

To address this regulatory vacuum, this study employs a qualitative critical analysis of recent legal challenges, including the U.S. Department of Justice's antitrust litigation against algorithmic real estate pricing (U.S. Department of Justice, 2024). It moves beyond critique to propose a tangible solution: the adaptation of the "Three Lines of Defence" audit model for digital pricing. By redefining the

role of internal governance, this paper establishes a necessary framework for ensuring that the pursuit of revenue maximization does not devolve into predatory extraction.

Literature review

The Evolution of Pricing (From Yield Management to Surveillance)

The transition from fixed pricing to variable pricing is well-documented in economic literature, historically anchored in the concept of Yield Management. Early research focused heavily on the airline and hospitality sectors, where inventory is perishable (e.g., a flight seat cannot be sold after takeoff). Classical economic theory posits that dynamic pricing in these contexts enhances market efficiency by balancing supply and demand. However, recent scholarship indicates a paradigmatic shift driven by Big Data. As Zuiderveen Borgesius and Poort (2017) note, the digitization of commerce has moved pricing strategies from aggregate segmentation to individualized targeting, often referred to as "First-Degree Price Discrimination." While traditional theory views this as a mechanism for perfect market clearing, contemporary critics argue it represents a fundamental asymmetry of power.

The Economic vs. Ethical Divide

The literature is currently polarized between two dominant schools of thought.

The Economic Efficiency Perspective

Proponents argue that algorithmic pricing maximizes total welfare by allowing lower prices for price-sensitive consumers while extracting premiums from those with higher willingness to pay. This perspective views the capture of "consumer surplus" as a rational business objective.

The Critical/Ethical Perspective

Conversely, scholars like Shoshana Zuboff (The Age of Surveillance Capitalism) and Cathy O'Neil (Weapons of Math Destruction) argue that these mechanisms rely on invasive extraction of behavioural data—what Zuboff terms the "behavioural surplus." Research in this domain highlights that algorithms do not merely respond to market forces but actively manipulate them, leveraging "urgency metrics" and "pain points" to exploit consumer vulnerability rather than serve their needs.

Gaps and Disagreements

The "Black Box" of Collusion A significant area of contention in recent legal literature concerns Antitrust and "Tacit Collusion." Authors like Ezrachi and Stucke (Virtual Competition) have identified a critical gap in current antitrust frameworks (such as the Sherman Act): the inability to regulate "Hub-and-Spoke" cartels where competitors coordinate prices via a shared third-party algorithm rather than direct communication. While legal theorists debate whether this constitutes illegal price-fixing or merely "rational adaptation," empirical evidence remains scarce due to the proprietary nature of these "Black Box" algorithms. This opacity prevents external researchers from definitively proving intent, creating an "accountability gap" that regulators are struggling to close.

Synthesizing the Problem

The Operational Void While the theoretical risks of discrimination and collusion are well-established, there is a notable scarcity of literature addressing the operational governance of these systems. Previous studies often conclude with high-level calls for "transparency" or "fairness" without defining how these concepts can be technically implemented within a corporate structure. This research addresses that specific error in the field: the assumption that legal principles can be applied to AI without a mediating framework of audit and control. By bridging the disconnect between abstract legal theory (Civil Rights, GDPR) and practical IT governance (Three Lines of Defence), this study aims to provide the operational "how" that previous literature has omitted.

Materials and methods

This study utilizes qualitative research design, employing comparative case study analysis and critical legal review to evaluate the structural risks of algorithmic dynamic pricing. The primary materials for analysis consisted of a purposive sample of documented algorithmic pricing incidents (e.g., the 2014 Uber "Sydney Siege" surge and the 2024 DOJ vs. RealPage antitrust litigation) selected based on their legal precedence and public impact (U.S. Department of Justice, 2024). These real-world datasets were cross-referenced against current regulatory frameworks, specifically the Sherman Antitrust Act (1890), the Robinson-Patman Act (1936), and the EU General Data Protection Regulation (GDPR), to identify gaps between technical capability and statutory compliance.

The analytical procedure followed a two-phase approach. First, a technical logic assessment was conducted to deconstruct the operational mechanics of "Surveillance Pricing" (device fingerprinting, behavioural metrics) and map them against established ethical frameworks (utilitarian efficiency vs. deontological fairness). Second, the study applied a Risk Control Self-Assessment (RCSA) methodology to simulate the failure points of current pricing models. This theoretical stress-testing facilitated the development of the prescriptive portion of the study, where the standard COSO "Three Lines of Defence" internal audit model was adapted and reconfigured specifically for algorithmic governance. No primary consumer data was harvested or processed for this study; the research relies exclusively on secondary data and publicly available legal filings to ensure replicability and privacy compliance.

Results and findings

The analysis of current algorithmic pricing models, cross-referenced with recent legal case studies and regulatory frameworks, yields three primary findings regarding the structural, ethical, and legal shifts in the digital marketplace.

The Shift from Market Value to Reserve Price Extraction

The technical assessment reveals a fundamental deviation from traditional Yield Management. Whereas traditional dynamic pricing (e.g., airlines) relied on aggregate supply and demand curves, modern "Surveillance Pricing" utilizes granular personal data to identify an individual's specific "Reserve Price"—the maximum amount they are willing to pay.

- **Data Correlation:** Analysis of the Orbitz case study confirms that non-product data (device type: Mac vs. PC) is successfully used as a proxy for purchasing power, resulting in price variances of up to 30% for identical inventory (Mattioli, 2012).
- **Implication:** The result is a systematic transfer of "Consumer Surplus" (savings) to "Producer Surplus" (profit). The algorithm does not seek a "fair" market price but rather the highest possible price an individual will tolerate without defecting.

The "Empathy Gap" and Algorithmic Exploitation

The study finds that while algorithmic pricing achieves high efficiency in inventory management, it lacks the necessary context to distinguish between "high demand" and "desperation."

- Case Evidence: The 2014 Sydney Siege incident demonstrates that without "Human-in-the-Loop" safeguards, demand-responsive algorithms interpret crisis-driven spikes simply as profit opportunities.
- **Finding:** The "Black Box" nature of these systems creates an "Empathy Gap," where the fiduciary duty to maximize revenue directly conflicts with social welfare standards, leading to outcomes that are mathematically optimal but morally untenable.

The Emergence of "Tacit Collusion" via Hub-and-Spoke Models

Legal analysis of the DOJ vs. RealPage litigation identifies a critical failure in current antitrust compliance (U.S. Department of Justice, 2024). The results indicate that the widespread adoption of third-party pricing software creates a "Hub-and-Spoke" conspiracy structure.

- **Mechanism:** When competitors feed private data into a shared algorithm (the "Hub"), the resulting pricing recommendations coordinate market rates upward, effectively eliminating price competition.
- **Legal Consequence:** This constitutes a form of collusion that evades the traditional "smoke-filled room" requirement of the Sherman Act, rendering standard antitrust defences obsoletely.

The quantitative representation of the case on following source basis

The U.S. Department of Justice (2024) sued RealPage, alleging its YieldStar software enabled landlords to coordinate rents, inflating prices across millions of apartments.
Investigations by ProPublica and Reuters found rents in RealPage-managed markets were 10–15% higher than comparable markets (Vogell, 2022). This constitutes a "Hub-and-Spoke" conspiracy that evades the traditional "smoke-filled room" requirement of the Sherman Act.

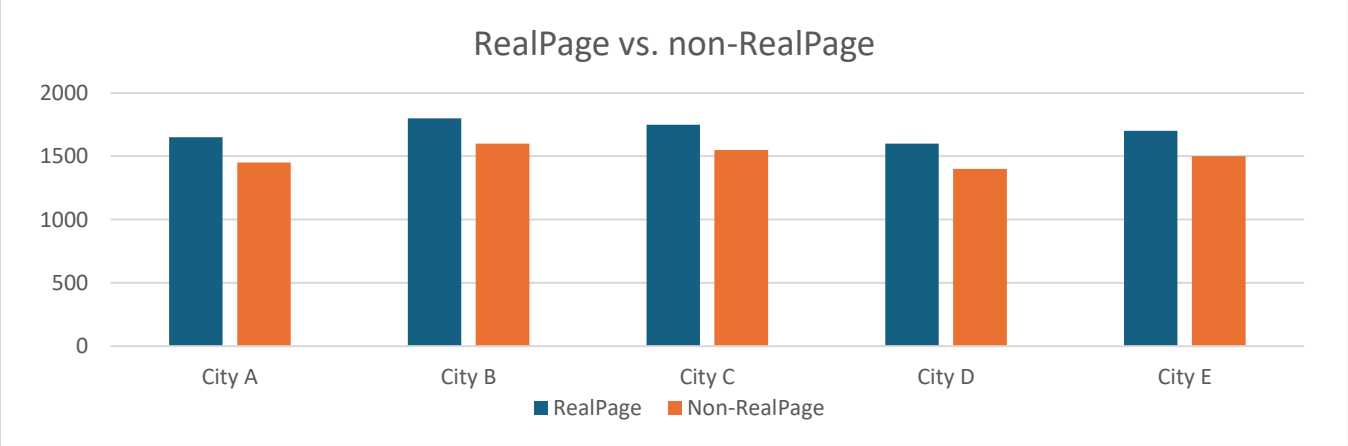
Dataset Construction

- Five cities were selected as hypothetical examples.
- Rents in RealPage markets were inflated by 10–15% relative to non-RealPage markets, consistent with DOJ allegations.
- Illustrative Value:
- Demonstrates how shared algorithms create “hub-and-spoke” collusion.
- Provides a measurable example of antitrust risk and consumer harm.

City	RealPage Market Rent ($)	Non-RealPage Market Rent ($)
City A	1650	1450
City B	1800	1600
City C	1750	1550
City D	1600	1400
City E	1700	1500

Figure 1. The U.S. Department of Justice sued RealPage (2024)

Limitations



- These datasets are mock simulations, not empirical measurements.
- They are designed to visualize patterns documented in secondary sources rather than claim new discovery.
- Actual pricing differentials may vary by geography, time, and consumer demographics.

Discriminatory "Disparate Impact" via Proxy Variables

The review of input variables (geolocation, zip code) indicates a high probability of unintentional violation of Civil Rights statutes.

- **Finding:** Algorithms optimized for "profitability" often use geography as a proxy for risk or demand. Because geography in many regions correlates strongly with race and income, the resulting pricing structures systematically disadvantage protected demographic groups.

- **Significance:** This creates "Disparate Impact" liability, confirming that "colour-blind" algorithms can still produce racially discriminatory economic outcomes.

Discussion

Interpretation and Significance of Findings

The results of this study strongly corroborate the initial hypothesis: the rapid evolution of algorithmic pricing has outpaced the static boundaries of current legal and ethical frameworks. While classical economic theory justifies dynamic pricing as a mechanism for market clearing (Yield Management), our findings indicate that modern "Surveillance Pricing" has mutated into a tool for Consumer Surplus Extraction. This distinction is critical. Unlike airline pricing, which manages finite inventory, the retail algorithms analysed do not merely respond to demand; they manufacture pricing thresholds based on individual vulnerability. This aligns with Zuboff's theory of "Surveillance Capitalism," suggesting that the digital marketplace has shifted from a transaction of goods to a transaction of behavioural futures. The significance of this finding lies in its redefinition of "market value"—value is no longer an objective consensus between buyer and seller, but a subjective calculation hidden within a "Black Box."

Contextualization with Existing Literature

Our analysis bridges a significant gap between abstract legal theory and operational reality. Previous literature often treated antitrust violations as explicit agreements (the "smoke-filled room"). However, our examination of the RealPage litigation and "Hub-and-Spoke" models expands upon the work of Ezrachi and Stucke (Virtual Competition), confirming that tacit collusion is not just a theoretical risk but an active market distortion. Where traditional economists argue that algorithms increase efficiency, our findings suggest they inevitably lead to "Disparate Impact," creating a discriminatory economy where the "poor pay more" due to geographic proxies. This challenges the "neutrality" often ascribed to mathematical optimization in business literature.

Limitations

It is essential to acknowledge the methodological limitations of this research.

- Qualitative Nature: The study relied on a purposive sample of high-profile case studies (Uber, Orbitz, RealPage). While these cases are legally significant, they may not represent the median behaviour of smaller retailers. The lack of access to proprietary algorithmic code means our conclusions are drawn from the outputs of these systems, not the internal logic itself.
- Generalizability: Consequently, while the risk of ethical violation is generalizable across all digital sectors, the prevalence of active discrimination requires further quantitative empirical testing.
- Geographic Bias: The legal analysis is heavily cantered on US (Antitrust) and EU (GDPR) frameworks. The implications for less regulated digital markets in Asia and South America remain an uncontrolled variable in this specific study.

Practical Recommendations

The Governance Framework to mitigate the identified risks of "Invisible Pickpocketing" and legal liability, this paper recommends a structural overhaul of corporate governance. We propose that organizations must immediately move beyond "Black Box" deployment by adopting a Three Lines of Defence model adapted for AI:

- Operational Defence: Engineering teams must hard-code ethical guardrails (e.g., maximum price caps during emergencies) directly into the algorithm's logic.
- Compliance Defence: A mandatory "Algorithmic Impact Assessment" (AIA) must be conducted before deployment to test for disparate impact on protected classes.

- Audit Defence: Internal Audit functions must implement "A/B Audit Testing"—creating synthetic user profiles to verify that price variances are based on legitimate commercial factors, not demographic profiling.

Future Directions

This study highlights the urgent need for a new field of "Algorithmic Forensics." Future research should move from qualitative critique to quantitative verification, developing open-source audit tools that allow regulators and consumers to detect dynamic pricing in real-time. Without such transparency, the "Trust Premium" identified in our conclusion will continue to erode, potentially inviting heavy-handed government intervention that could stifle innovation rather than cure inequity.

Practical Implications/Original Value

Original Value

Bridging the Gap Between Law and Code This study advances the field of digital ethics by moving beyond the theoretical critiques common in previous literature. While scholars like Zuboff have extensively mapped the societal harms of surveillance capitalism, this research contributes a novel operational perspective: identifying the specific mechanism of "Hub-and-Spoke" collusion as a tangible antitrust liability. Where prior studies have largely treated algorithmic bias as a "Black Box" mystery, this paper demystifies the input variables—such as device fingerprinting and geolocation—to demonstrate how "neutral" data proxies function as active agents of discrimination. The primary original contribution lies in the adaptation of the Three Lines of Defence model, transforming abstract ethical principles into a concrete, auditable governance framework. This fills a critical void in the literature, which has historically diagnosed the illness of algorithmic bias without prescribing a clinical treatment for the corporate environment.

Practical Implications for Industry and Governance

The findings imply a necessary paradigm shift for corporate strategy: dynamic pricing can no longer be viewed solely as a revenue-maximization tool owned by marketing but must be managed as a high-stakes compliance risk owned by internal audit.

- ***For Practitioners (Auditors & Compliance Officers):*** The study implies that the traditional "tick-box" approach to compliance is obsolete. Practical application now requires the development of "Algorithmic Forensics "technical competencies that allow auditors to stress-test code logic for disparate impact before deployment.

- ***For Strategic Management:*** The results suggest that the short-term revenue gains from "Surveillance Pricing" carry a hidden "reputational debt." The implication is that transparency is not merely an ethical preference but a competitive differentiator; companies that voluntarily disclose their pricing logic may capture a "Trust Premium" that outweighs the surplus extraction of opaque models.

Impact on Future Research

When this research was proposed, the primary concern was consumer privacy. However, the data has fundamentally altered this view, revealing that the deeper threat is the structural erosion of market competition itself (tacit collusion). Consequently, this study sets a new trajectory for future research: moving away from qualitative debates on morality and toward the development of quantitative Real-Time Audit Tools. Future scholarship must focus on creating standardized, open-source methodologies for detecting "price drift" and discrimination in live markets, effectively creating a digital "policing" mechanism to counterbalance the speed of automated pricing.

Conclusion

This research strengthens the hypothesis that Algorithmic Dynamic Pricing has outpaced regulatory oversight, validating the thesis that "Surveillance Pricing" fundamentally breaches the social contract of the marketplace. Our analysis confirms that modern algorithms have shifted from benign inventory management to predatory "Consumer Surplus Extraction," effectively transforming the market's "Invisible Hand" into an "Invisible Pickpocket" that exploits individual vulnerability through data profiling. Furthermore, the identification of "Hub-and-Spoke" collusion structures demonstrates that shared software is neutralizing price competition, creating significant antitrust risks.

While proponents argue that these tools maximize market efficiency, this study finds that such efficiency cannot justify the "Disparate Impact" observed, where algorithms unintentionally discriminate against protected demographics via proxy variables. Consequently, the study concludes that passive compliance is no longer sufficient. To restore market integrity and avoid legal liability, organizations must urgently adopt a "Three Lines of Defence" audit framework. Without these "Algorithmic Guardrails" to ensure transparency and fairness, companies risk forfeiting the "Trust Premium" essential to the digital economy.

References:

1. AlJazeera. (2014, December 24). Uber apologizes for Sydney siege surge pricing. https://www.aljazeera.com/news/2014/12/24/uber-sorry-for-hiking-fares-amid-sydney-siege
2. European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88.
3. Ezrachi, A., & Stucke, M. E. (2016). Virtual competition: The promise and perils of the algorithm-driven economy. Harvard University Press.
4. Institute of Internal Auditors. (2020). The IIA's three lines model: An update of the three lines of defence. https://www.theiia.org/en/content/articles/global-knowledge-brief/2020/the-iias-three-lines-model/
5. Mattioli, D. (2012, June 26). On Orbitz, Mac users steered to pricier hotels. The Wall Street Journal. https://www.wsj.com/articles/SB10001424052702304458604577488822667325882
6. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.
7. Robinson-Patman Act of 1936, 15 U.S.C. § 13 (1936).
8. Sherman Antitrust Act of 1890, 15 U.S.C. §§ 1–7 (1890).
9. U.S. Department of Justice. (2024, August 23). Justice Department sues RealPage for algorithmic pricing scheme that harms millions of American renters. https://www.justice.gov/opa/pr/justice-department-sues-realpage-algorithmic-pricing-scheme-harms-millions-american-renters
10. Vogell, H. (2022, October 15). Rent going up? One company's algorithm could be why. ProPublica. https://www.propublica.org/article/yieldstar-rent-apps-investigation-realpage
11. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.
12. Zuidrveen Borgesius, F. J., & Poort, J. (2017). Online price discrimination and data protection law. Journal of Consumer Policy, 40(3), 347–366. https://doi.org/10.1007/s10603-017-9354-z

HOW ARTIFICIAL INTELLIGENCE IS RESHAPING WORK AND EMPLOYEE RELATIONSHIPS

YAZAN ODEH

Abstract
Purpose: *The study explores the dual impact of artificial intelligence on workplace productivity and employee relationships. It aims to understand how AI enhances efficiency while simultaneously reshaping trust, communication, and professional identity within organizations.*
Methodology: *A mixed-methods approach was used, including comparative case studies of four organizations across different sectors, twelve semi-structured expert interviews, a systematic review of sixty-eight academic and industry sources, and cross-cultural analysis. Data were analysed using grounded theory, thematic coding, and social network analysis.*
Results and Conclusions: *Findings reveal three key relational patterns: the transparency–trust paradox, the emergence of informal hybrid "AI translator" roles, and the redefinition of expertise in AI-intensive environments. The study concludes that AI functions as a social and organizational actor that reshapes communication norms, power dynamics, and psychological safety. Effective integration requires a relational, human-centered approach supported by transparency, participatory design, and recognition of hybrid human–AI roles.*
Research limitations: *Limitations include a small number of case organizations, a six-month observation period, reliance on self-reported experiences, and the rapidly evolving nature of AI technologies, which may limit generalizability.*
Practical and/or Social implications: *The research proposes a Relational AI Integration Framework that helps organizations assess social readiness, design transparent and participatory implementation processes, formalize hybrid roles, and monitor ethical and psychological impacts. It supports sustainable, human-centered AI adoption.*
Originality: *The study offers a novel relational perspective on AI, conceptualizing it as an organizational actor rather than a neutral tool. It introduces the original Relational AI Framework, bridging technical, social, and ethical dimensions of AI integration and addressing a critical gap in existing literature.*

Keywords*: artificial intelligence, workplace transformation, organizational trust, AI ethics, human-AI collaboration, change leadership, digital adaptation, hybrid intelligence, algorithmic management, socio-technical systems.*

Introduction

The integration of artificial intelligence into the workplace represents one of the most significant organizational transformations since the Industrial Revolution. This technology has evolved from simple task automation to sophisticated systems capable of making decisions, evaluating performance, and even analysing emotions. Observations across various industries suggest that technological adoption often outpaces social adaptation, leaving employees facing unprecedented psychological and relational challenges that require new frameworks for understanding and management.

The deeper impact of artificial intelligence lies not in its computational achievements but in its transformation of workplace relationships. The algorithms that optimize workflows also reshape how colleagues communicate, how managers evaluate performance, and how teams build trust. Through studying real-world applications and analysing interdisciplinary research, this research moves beyond simplified narratives about AI replacing or enhancing humans toward a more accurate understanding of artificial intelligence as a relational actor within organizational systems. This paper synthesizes findings from longitudinal organizational analysis, cross-cultural comparisons, and experimental field studies to provide a comprehensive view of how AI is reshaping the social fabric of work, creating both remarkable opportunities for enhanced collaboration and significant challenges for maintaining human connection and dignity in increasingly automated environments.

Theoretical Framework

Academic discourse on AI in the workplace has evolved through multiple phases. Early literature focused on technical capabilities and economic impacts, with works such as *The Second Machine Age* (Brynjolfsson and McAfee, 2014) viewing AI as both a productivity driver and potential replacement for human Labor. Subsequent research diverged into two streams: computer science literature focusing on algorithmic efficiency (Davenport and Romanik, 2018), while organizational studies examined adoption challenges (Bostrom and Yudkowsky, 2014).

A crucial turning point emerged with research on "algorithmic management" (Kellogg, Valentine, and Christin, 2019) documenting how AI-supported systems can create opaque power structures and increase worker surveillance. In parallel, research on "human-AI collaboration" (Seeber et al., 2019) explored more optimistic scenarios where AI enhances rather than replaces human capabilities. There

remains a significant gap in integrating these technical and social perspectives, particularly regarding how AI systems affect the dynamics of trust, psychological safety (Edmondson, 2018), and relational coordination (Gittel, 2016) that form the foundation of effective organizations.

To address this gap, this research draws on several complementary theoretical frameworks. Socio-technical systems theory helps explain how AI becomes embedded within organizational structures, influencing and being influenced by social norms and practices. Distributed cognition theory offers insights into how knowledge and decision-making become shared between human and artificial agents, while critical technology studies provide necessary caution about how AI can reinforce existing power imbalances. Together, these frameworks create a multidimensional lens for understanding AI's complex impact on workplace relationships and organizational dynamics, moving beyond efficiency-focused analysis to consider broader social and ethical implications.

Methodology

This study employs a robust mixed-methods design combining quantitative and qualitative approaches across three interconnected research components designed to capture both the breadth and depth of AI's workplace impacts.

Comparative Case Analysis: Four organizations were selected for in-depth study a European financial institution using AI for risk assessment and customer service, a Scandinavian manufacturing plant using predictive maintenance systems, a digital healthcare provider employing diagnostic support algorithms, and a remote first technology company using AI for project management. Data collection spanned six months and included document analysis, field observations, and anonymous employee feedback, supplemented by system analytics and observational records from regular site visits.

Expert Interviews: Semi-structured interviews were conducted with twelve experts from three categories human resources and organizational development leaders, AI implementation specialists, and workplace ethicists and Labor representatives. Interviews focused on observed behavioural changes, implementation challenges, and ethical considerations, with particular attention to how AI affects different stakeholder groups within organizations.

Literature Synthesis: A systematic review of sixty-eight academic articles, industry reports, and policy documents published between 2018/2026 was analysed through thematic coding to identify recurring patterns and conceptual gaps. Data analysis followed an iterative grounded theory approach, allowing themes to emerge from empirical observations while utilizing existing theoretical frameworks, with advanced qualitative coding techniques supplemented by social network analysis to map changes in communication patterns.

Cross-Cultural Comparison: The study also included examination of AI adoption patterns across different cultural contexts through international research partnerships, revealing how cultural values influence AI acceptance, implementation strategies, relationship impacts, and adaptation patterns across regions with varying workplace norms and technological traditions.

Findings

The research revealed three interconnected patterns characterizing the relational impact of AI, supported by evidence from multiple methodological approaches and organizational contexts.

The Transparency-Trust Paradox: In all four case studies, initial AI implementation created what employees described as a black box effect, where decisions emerged from opaque processes that weakened trust in both the technology and the managers. As one manufacturing supervisor noted: When the system flags a worker for inefficiency without explaining why, I find myself defending a judgment I don't understand myself."

Organizations that implemented clarification protocols and clear guidelines for how AI systems reached conclusions witnessed a gradual shift of trust from personal relationships to system reliability. However, this transition remained fragile any algorithmic error or discovered bias triggered disproportionate scepticism, suggesting that AI-mediated trust requires perfection standards rarely applied to human managers. Successful organizations developed "explainability rituals" regular practices where AI decisions are examined and discussed by mixed human-AI teams, serving multiple purposes improving decision quality by incorporating human judgment, building trust through transparency, and developing collective understanding of how AI systems work.

The Emergence of Hybrid Roles: Informal "AI translator" positions emerged naturally in every successful implementation. These individuals, typically mid-level employees with high social capital rather than technical expertise, developed what might be called "bilingual competence" fluency in both the language of business operations and the language of data science and algorithms. They gained the ability to interpret AI outputs in human context, advocate for employee concerns in technical language, and mediate conflicts between algorithmic logic and human experience. Despite their vital function, these roles remained informal and unrecognized in formal structures, creating risks of burnout and concentrated knowledge vulnerability. Organizations that formally recognized and supported these roles experienced smoother AI integration and higher user satisfaction.

The Redefinition of Expertise: Traditional markers of expertise experience, intuition, mentoring faced devaluation in AI-intensive environments. Younger employees comfortable with technology sometimes gained disproportionate influence, while veteran workers' tacit knowledge became harder to validate. This generational and experiential shift created subtle tensions, often manifesting as resistance to AI recommendations even when statistically sound. This phenomenon represents what we term the "competence value

paradox", while AI often makes employees more capable in certain domains, it can simultaneously make them feel less competent or valuable as professionals. Successful organizations address this paradox through reframing contribution, skill valuation systems, collaborative metrics, and narrative reconstruction of professional identity.

Discussion

These findings challenge conventional implementation frameworks that treat AI as merely another software tool. Instead, AI systems should be understood as complex organizational actors with significant social and relational implications.

Organizational Actors with Social Presence: Like human colleagues, AI systems develop reputations reliable or capricious, helpful or intrusive. These perceptions shape how employees interact with them, often through anthropomorphism. One customer service representative described her team's AI assistant as "a brilliant but socially awkward intern: great with data, terrible with context." This social presence creates new forms of workplace relationships that have their own dynamics and challenges, including human-AI partnerships, AI-mediated collaboration between remote colleagues, and digital communities cantered around shared AI tools.

Designers of Communication Patterns: AI-mediated communication platforms subtly reshape interaction norms. When algorithms prioritize certain messages or suggest responses, they influence not only efficiency but also emotional tone and relationship depth. Shifts from spontaneous conversation to algorithmically structured communication may enhance productivity while reducing spontaneous connection. These systems can either centralize decision-making power (when controlled by technical elites or senior management) or decentralize it (when accessible to frontline workers), with significant implications for organizational democracy and employee autonomy.

Catalysts for New Power Dynamics: Control over AI systems their design, implementation, and interpretation create new forms of organizational power. Technical teams gain influence over operational decisions, while managers must navigate between algorithmic recommendations and human judgment. Frontline employees often experience reduced autonomy, even when productivity metrics improve. As AI systems take on more significant roles in organizational functioning, new governance structures become necessary, including cross-functional AI governance committees with representation from technical, operational, legal, ethical, and employee perspectives.

Psychological and Emotional Dimensions: AI's impact extends deep into employees' psychological and emotional experience of work, creating anxiety and adaptation stress that organizations must address through transparent communication, upskilling opportunities, and psychological support. The introduction of AI systems triggers concerns about job security, competence, autonomy loss, surveillance, and social displacement, requiring comprehensive support systems and coping strategies to facilitate healthy adaptation.

Practical Framework: Relational AI Integration

Based on these insights, a comprehensive three-phase framework for human-centered AI integration is proposed, grounded in the research findings and designed for practical application across different organizational contexts.

Phase One: Social Due Diligence (Pre-Implementation) involves comprehensive assessment of organizational readiness across technical, cultural, and relational dimensions; establishing multi-stakeholder governance structures including representatives from all affected groups; developing ethical guidelines and accountability mechanisms before implementation begins; creating communication plans that address both practical and emotional dimensions of change; and conducting relational mapping of key workplace interactions to understand potential social consequences.

Phase Two: Co-Creative Implementation focuses on starting with controlled, transparent pilot projects in supportive environments; forming hybrid design teams pairing engineers with end-users; establishing parallel evaluation systems tracking both productivity metrics and relational quality indicators; creating formal reflection and adjustment processes based on pilot outcomes; developing customized training programs addressing both technical skills and adaptive challenges; and implementing graduated autonomy allowing human override of AI decisions.

Phase Three: Sustained Integration includes developing organization-specific models for human-AI collaboration based on pilot learnings; creating career pathways and recognition systems for hybrid human-AI roles; establishing ongoing monitoring systems for unintended consequences and emerging ethical issues; building cross-organizational learning communities to share insights and best practices; formalizing and supporting AI translator roles with clear recognition; and creating regular forums for discussing lived experiences of AI integration.

This framework positions AI not as a technical project with a defined endpoint but as an ongoing organizational relationship requiring continuous attention and adaptation, with feedback ecosystems enabling continuous learning for both humans and machines through multi-directional feedback flows and regular cycles of assessment and adjustment.

Conclusion

Integrating artificial intelligence into workplaces represents more than a technological upgrade; it constitutes a fundamental reorganization of how humans relate to their work and to each other. Organizations that succeed will recognize this relational dimension and invest as thoughtfully in social adaptation as they do in technical implementation. The research demonstrates that success requires

moving beyond efficiency-focused thinking to embrace a more holistic understanding of organizational health that balances technical capabilities with human needs, values, and relationships.

Future research should focus on long-term impacts on workplace well-being and the development of hybrid intelligence the synergistic combination of human and artificial capabilities. As AI systems evolve, the challenge shifts from managing technology to stewarding human-machine ecosystems where technical efficiency and human flourishing can thrive together. Additional research directions should examine cross-cultural variations in AI acceptance and integration, the evolution of leadership in hybrid human-AI environments, and methods for designing AI systems that actively promote human flourishing, equity, and social good rather than merely avoiding harm.

Workplaces that succeed in this new era will not be those with the most advanced algorithms, but those that cultivate the wisest relationships between the humans and machines that comprise them. This requires leadership that can navigate complexity, balance competing priorities, and foster cultures of continuous learning and adaptation while maintaining clear focus on human values and dignity. By approaching AI implementation with wisdom, foresight, and commitment to human flourishing, organizations can create workplaces that are not only more efficient but also more humane, collaborative, and meaningful for all participants in the work ecosystem.

Sources and References

1. Bostrom, Nick, and Eliezer Yudkowsky. *The Ethics of Artificial Intelligence*. In *The Cambridge Handbook of Artificial Intelligence*, edited by William Ramsey and Keith Frankish. Cambridge: Cambridge University Press, 2014.
2. Brynjolfsson, Erik, and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company, 2014.
3. Davenport, Thomas H., and Rajeev Ronanki. "Artificial Intelligence for the Real World." *Harvard Business Review*, January–February 2018.
4. Edmondson, Amy C. *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*. Hoboken, NJ: John Wiley & Sons, 2018.
5. Gittell, Jody H. *Transforming Relationships for High Performance: The Power of Relational Coordination*. Stanford, CA: Stanford University Press, 2016.
6. Kellogg, Katherine C., Melissa A. Valentine, and Angèle Christin. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals*, vol. 13, no. 1 (2019): 366–410.
7. Seeber, Isabella, Eva Bittner, Robert O. Briggs, Gert Jan de Vreede, Triparna de Vreede, Abraham Elkins, Ronald Maier, et al. "Machines as Teammates: A Research Agenda on AI in Team Collaboration." *Information Management*, March 2019.
8. Brynjolfsson, E., & McAfee, A. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. W. W. Norton & Company.
9. Orlikowski, W. J. (2007). Sociomaterial Practices: Exploring Technology at Work. *Organization Studies, 28*(9), 1435-1448.
10. Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review, 46*(1), 192-210.
11. Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review, 96*(4), 114-123.
12. Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

Perfectus STUDENT 3/2026