

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-10-28

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:

Moreira, O., Rodrigues, M.C., Sousa, M. & Filho, A. (2025). Governance mechanisms for mitigating AI challenges to human intelligence: A conceptual framework analysis. European Academy of Management Conference, EURAM 2025. EURAM.

Further information on publisher's website:
<https://conferences.euram.academy/2025conference/>

Publisher's copyright statement:

This is the peer reviewed version of the following article: Moreira, O., Rodrigues, M.C., Sousa, M. & Filho, A. (2025). Governance mechanisms for mitigating AI challenges to human intelligence: A conceptual framework analysis. European Academy of Management Conference, EURAM 2025. EURAM.. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.



GOVERNANCE MECHANISMS FOR MITIGATING AI CHALLENGES TO HUMAN INTELLIGENCE: A
CONCEPTUAL FRAMEWORK ANALYSIS

OTACILIO MOREIRA

Universidade Federal Fluminense, Brazil

Maria Carolina Rodrigues

Cinturs universidade do Algarve, Portugal

Maria Sousa

Instituto Universitário de Lisboa, Portugal

Américo Filho

Universidade Federal Fluminense, Brazil

Access to this paper is restricted to registered delegates of the EURAM 2025 (European Academy of Management) Conference.

ISSN 2466-7498 and ISBN 978-2-9602195-7-9

GOVERNANCE MECHANISMS FOR MITIGATING AI CHALLENGES TO HUMAN INTELLIGENCE: A CONCEPTUAL FRAMEWORK ANALYSIS

ABSTRACT: This study proposes a dynamic governance model to guide the interaction between artificial intelligence (AI) and human intelligence. It focuses on mechanisms that mitigate AI's impact on human cognition while promoting innovation, ethical responsibility, and alignment with the Sustainable Development Goals (SDGs) of the 2030 Agenda. Using a mixed-methods approach, the study combines a narrative literature review (2019–2024) from Web of Science, ScienceDirect, and Research Rabbit with a structured development process.

The model comprises three interconnected blocks: **Governance**, **Human Intelligence**, and **Social Intelligence**, each linked to specific SDGs. The Governance block emphasizes the responsible use of AI through regulatory policies that promote algorithmic fairness and data governance, aligning with SDGs 9 and 16. It also advocates for AI literacy and ethics education to foster inclusion and equality (SDGs 4 and 10) and implements safeguards such as transparency and accountability to ensure the ethical use of AI (SDGs 9 and 16). The Human Intelligence block examines AI's cognitive impacts, balancing cognitive augmentation in problem-solving (SDGs 4 and 9) with caution against the risks of over-reliance, such as cognitive erosion (SDG 4). The Social block highlights AI system assessments to enhance adaptation, inclusion, and equitable access (SDGs 10 and 16) while encouraging policy adjustments for sustainable innovation (SDGs 9 and 17).

The study concludes that integrating these blocks is essential for achieving fair, inclusive, and sustainable technological development. Structured governance, incorporating ethical principles, social reform, and continuous monitoring, ensures the responsible adoption of AI, supporting the harmonious interaction between AI and human intelligence in achieving the Sustainable Development Goals (SDGs).

KEYWORDS: Governance mechanisms; AI challenges; Human intelligence; Conceptual framework; Human-AI interaction SDGs; AI governance SDGs

1. INTRODUCTION

The rapid proliferation of artificial intelligence requires robust governance mechanisms to address its ethical implications and social impact. AI technologies have become integral to various industries and services, including healthcare, finance, security, agriculture, and transportation, raising critical questions about their responsible use and application. While AI offers transformative potential, it also carries significant risks, including privacy violations, bias, misinformation, and erosion of human cognitive autonomy, requiring a careful balance between innovation and regulation.

The epistemological positioning of technological ethics in the broader context of moral values research presents a significant methodological challenge. While academic discourse often situates technological ethics within the framework of applied ethics, this categorization requires further critical analysis and theoretical deconstruction. The absence of a comprehensive regulatory framework has elevated ethical considerations to a dominant position in AI governance, championed by diverse stakeholders, including governments, NGOs, industry leaders, and academic institutions.

Ethical guidelines, often framed as voluntary principles, address issues such as fairness, accountability, transparency, and respect for human dignity (Müller, 2020; Coeckelbergh, 2020; Nyholm, 2020, 2022a, 2020b; Nyholm & Rüther, 2023; UNESCO, 2021). However, critics argue that this reliance on voluntary principles can strategically undermine the implementation of stricter regulatory oversight, prioritizing the perception of digital freedom over effective public scrutiny. The tension arises from the interaction between ethical guidelines and regulatory mechanisms, as well as the application of ethical principles. Often drafted by experts and private actors, these

principles can lack democratic legitimacy, raising questions about the specific interests they represent. Despite their nature, these principles are not legally binding and can be applied selectively or disregarded when they conflict with commercial interests (Jasanoff, 2011). The dynamic and context-dependent nature of social values further complicates the application of ethical principles in diverse cultural and socioeconomic contexts.

The sheer volume of data processed by AI systems amplifies ethical complexities while providing valuable insights for decision-making and efficiency. According to Vasconcelos et al. (2023), AI also poses significant risks, including mass surveillance, algorithmic bias, and *the proliferation* of misinformation. This inherent duality highlights the necessity for governance mechanisms that address AI's ethical concerns and ensure its alignment with broader societal objectives. Therefore, this research aims to develop a comprehensive governance model that transcends a one-size-fits-all approach, considering the diverse applications and implications of AI across various industries.

For the Academy, a comprehensive governance model for AI is *essential* and must transcend generalist approaches, considering the diverse applications and implications of AI across various sectors. This model must integrate ethical principles and applicable regulations, promoting accountability, transparency, and inclusion. As AI becomes an integral part of daily life, advancing medicine, optimizing transportation, and transforming industries, its use and the ethical and social implications it raises demand immediate and sustained attention. Only through informed policy, collaborative governance, and continuous evaluation can we reap the benefits of AI while mitigating its potential dangers.

This research examines the complex interplay between artificial intelligence (AI) and human intelligence, developing a comprehensive governance model that integrates institutional technology, ethics, and dynamics. The research question is: How can governance mechanisms

effectively mediate interactions between AI and humans?

The primary objective is to propose a dynamic framework that governs interactions between artificial and human intelligence. The focus is on identifying and structuring the mechanisms that moderate.

AI's impacts on human cognition and formulating guidelines to balance innovation and responsibility, as well as ethical and social.

The proposed model examines how ethical guidelines, regulatory policies, and participatory approaches can mitigate the impact of AI on human cognition and society. It aims to provide a roadmap to balance innovation with social responsibility.

With this broader objective, the study seeks to answer the research question, aiming to deepen the analysis of the interaction between artificial intelligence and human intelligence and to investigate the governance mechanisms necessary to moderate this relationship and its broader implications for society.

Including the Sustainable Development Goals (SDGs) in studies and research is highly relevant, as it aligns global strategies and challenges with the goals outlined in the SDGs. Regulatory policies play a crucial role in ensuring that artificial intelligence (AI) is utilized to make a positive contribution to society and the economy, aligning with Sustainable Development Goal 9 (Industry, Innovation, and Infrastructure). In addition, the social and educational reform component prepares individuals to interact critically with AI, ensuring that society benefits equitably from technological innovations, in line with Sustainable Development Goal 10 (Reducing Inequalities). Technological safeguards, in turn, are essential for protecting the social environment and promoting an ethical and sustainable advancement of technology, thereby contributing to the construction of fairer and more innovative societies, which aligns with SDG 16 (Peace, Justice, and Strong Institutions).

Regarding human intelligence, applying ethical principles is crucial for the responsible development of AI and for constructing an equitable society in line with Sustainable Development Goal 10. The cognitive augmentation indicates that AI can act as a transformative agent in education and innovation, supporting SDG 4 (Quality Education) and SDG 9. On the other hand, cognitive erosion raises social concerns that demand a critical approach to interaction with AI, emphasizing the need to ensure educational quality and equity in access to resources in alignment with SDG 4. Additionally, cognitive discharge, by impacting individual development and institutional effectiveness, highlights the importance of promoting conscious interaction with technology, once again reinforcing the relevance of SDG 4.

Finally, monitoring and evaluation are crucial for establishing accountability and trust in governance practices, as well as promoting equitable access to information directly related to Sustainable Development Goal 16. Policy adjustments should be adaptable and collaborative, ensuring equitable and sustainable strategies and reaffirming the connection to SDGs 9 and 10. Impact assessment helps mitigate risks and builds a fairer and more inclusive society in line with SDG 10. Thus, the interconnection between these three blocks and the SDGs is essential to ensure that artificial intelligence promotes sustainable and responsible development, benefiting society.

2. THEORETICAL FRAMEWORK

This study proposes a dynamic governance model that incorporates practical elements of monitoring and evaluation, fostering continuous collaboration among diverse stakeholders and acknowledging the complex interplay between artificial intelligence and human intelligence. By addressing these issues in an integrated and coordinated manner, we can ensure that AI is developed and used responsibly and ethically, promoting technological development that is fair, inclusive, and sustainable. This not only maximizes the benefits of AI but also minimizes its

potential risks, ensuring the well-being of individuals and society. Unlike static models that treat this relationship in a binary or simplistic way, it translates the dynamic and complex nature of its interaction, integrating ethical, regulatory, and participatory perspectives for a more comprehensive understanding of the challenges and opportunities of AI, structured in three interconnected blocks: governance, human intelligence, and social dimension. This is what is discussed below:

2.1. Governance Block

In this section, the essential mechanisms for governing the development and deployment of artificial intelligence (AI) are analyzed, integrating legal, social, and technological considerations (Cath, 2018). The analysis is divided into three main components: regulatory policies, social and educational reforms, and technological safeguards.

Regulatory policies: This component focuses on regulatory policies and their ability to address the challenges presented by AI. SDG 16, which promotes peace, justice, and strong institutions, emphasizes the importance of establishing legal frameworks that ensure algorithmic fairness and effective data governance. This goal emphasizes the need to build institutions that are transparent and accountable, which are crucial to ensuring that technology benefits all segments of society. Moreover, SDG 9, which focuses on industry, innovation, and infrastructure, emphasizes that implementing effective regulatory policies is crucial for creating an environment conducive to innovation and sustainable development. These policies should collaborate internationally to address emerging legal challenges in AI, ensuring that technological innovations drive economic growth while aligning with social and ethical values.

The component underscores the importance of international cooperation in creating common legal frameworks. Zhuk (2024) compared how each European Union jurisdiction addresses authorship challenges, intellectual property rights, and ownership in the context of AI creativity, highlighting

differences in regulatory approaches and their implications for innovation, cross-border collaboration, and ethical concerns. Abedrabo (2024) explores the importance of establishing legal frameworks and standards to ensure algorithmic fairness and data governance in artificial intelligence (AI)

Social and Educational Reforms: This component addresses the social and educational measures necessary to ensure the responsible development and deployment of AI. This component is aligned with SDG 4, which aims to ensure quality education and underscores the importance of promoting.

AI literacy and ethics education. These initiatives are crucial in ensuring that everyone has access to learning and development opportunities, thereby contributing to the fight against social inequalities. In addition to SDG 10, which focuses on reducing inequalities, it emphasizes the importance of access to inclusive educational programs and AI training. This represents a crucial step in ensuring that technologies do not deepen existing inequalities but promote inclusion and equity.

For Mohanty et al. (2023), Shukla (2024), and Ifenthaler et al. (2024), while limited access to AI education and training programs can widen inequalities, AI-powered language translation systems, according to Pflanzner (2023), offer a promising solution to promote inclusion and ensure everyone can access educational and skill-building opportunities. Wirtz, Weyerer, and Geyer (2018) emphasize the importance of multisectoral collaboration, encompassing governments, civil society, academia, and industry, as a crucial element in effective AI governance.

Technological safeguards: Adopting technological safeguards is essential to mitigate risks and ensure that AI is used ethically, safely, and effectively. These safeguards protect users and institutions, promoting trust in technology. Its implementation is directly related to SDG 16, which aims to achieve peace, justice, and the strengthening of institutions. This highlights that

safeguards such as transparency and accountability are crucial for establishing robust institutions that can effectively regulate the ethical use of AI, thereby promoting social justice and upholding human rights. Similarly, SDG 9, which focuses on industry, innovation, and infrastructure, highlights that technological safeguards that ensure the integrity and reliability of AI systems are key to responsible innovation and sustainable infrastructure development.

According to Felzmann *et al.* (2020), deploying technological safeguards is essential to ensure that AI is developed and used responsibly. Cheong (2024) and Balasubramaniam and Kujala (2023) emphasize the importance of creating AI systems that prioritize transparency and explainability to foster trust.

And ensure accountability. Tocchetti *et al.* (2024) examined the robustness of AI from a human-centered perspective, emphasizing the intersection of technological challenges and societal needs.

Table 1. Summary of the Governance Block

Component	Description	Authors
Regulatory Policies	Assessing the adequacy of legal frameworks to regulate AI, including privacy, bias, and rights.	Zhuk (2024); Abedrabo (2024)
Social and educational reforms	Measures to promote AI literacy, reduce inequalities, and foster multisectoral collaboration.	Shukla (2024); Wirtz, Weyerer & Geyer (2018)
Technological safeguards	Implementation of measures for transparency, privacy, robustness, and human-centered governance.	Felzmann et al. (2020); Cheong (2024); Tocchetti et al. (2024); Balasubramaniam & Kujala (2023)

Source: The authors

2.2. Human Intelligence Block

This block analyzes the multiple impacts of artificial intelligence (AI) on human cognitive processes and social structures. It explores ethical, cognitive, and social aspects, both positive and negative, and divides them into four main components: Ethical Principles, Cognitive Augmentation, Cognitive Erosion, and Cognitive Offload.

Ethical Principles are a fundamental component that examines the ethical principles that guide the development and use of AI, including fairness, transparency, accountability, and respect for human dignity. This section is based on these principles. The following SDGs are relevant to this component.

These principles form the basis of this discussion and have significant implications for several Sustainable Development Goals. SDG 16, which promotes peace, justice, and institution-building, is closely aligned with promoting fairness, transparency, and accountability in the use of AI, ensuring that systems operate in a manner that respects human rights and society's fundamental values.

Additionally, SDG 10, which focuses on reducing inequalities, highlights the importance of eliminating structural biases and ensuring accessibility and fairness in AI systems. According to OECD 2019 and UNESCO 2021, AI should contribute positively to society and the environment by ensuring the equitable distribution of its benefits. This aligns with the notion that trustworthy AI can play a pivotal role in promoting inclusive growth, sustainable development, and overall well-being.

Cognitive Augmentation is a component that looks at artificial intelligence's (AI) ability to enhance human cognitive capabilities through automation, generating insights, and assisting in solving complex problems. This aspect of AI is closely related to several of the United Nations' Sustainable Development Goals (SDGs). SDG 4, which aims to ensure quality education, highlights how AI can enhance learning processes and improve decision-making efficiency, making it crucial for educational improvement and the development of cognitive skills. SDG 9, which addresses industry, innovation, and infrastructure, is closely linked to the role of AI in promoting innovation in industries, particularly by developing technologies that enhance productivity and quality of work.

In addition, SDG 10, which focuses on reducing inequalities, underscores the importance of eliminating structural biases and ensuring that AI systems are accessible and fair, thus reducing social inequalities.

Humble and Mozelius (2024) mention how AI-based educational technologies influence students' learning processes and cognitive resources in programming and related technical domains. Chen and Yadollahpour (2024) report that AI is driving a new era in cognitive neuroscience, providing powerful tools for both research and practical applications. Horvatić and Lipić (2021) emphasize the need for human-centric AI to address biases arising from the black box nature and lack of transparency in AI systems. This transparency is essential for enabling users to control and improve AI performance, while ensuring fairness and accountability. Mao *et al.* (2024) highlight the ethical.

Moreover, technical challenges, including algorithmic bias and privacy issues, must be addressed to ensure that these advances are beneficial and sustainable in the long term.

Cognitive erosion is a component that examines the potential for adverse consequences, such as overreliance on AI systems, which can lead to a decline in critical thinking, problem-solving, and creativity; the risk of skill obsolescence as AI automates tasks previously performed by humans; and the potential for AI to manipulate user attention and impact decision-making processes, as is often observed on social media platforms.

This component is closely related to SDG 4, which promotes quality education and emphasizes that the unexamined acceptance of AI suggestions can compromise critical thinking, a key element for effective and reflective education. For its part, SDG 10, which aims to reduce inequalities, emphasizes that the trend toward addiction to automated solutions can perpetuate existing disparities, as vulnerable populations may lack access to the necessary tools to develop their capacities independently.

The dual perspectives on AI's impact on human intelligence, as analyzed by Passi and Vorvoreanu (2022) and Nyholm (2024), suggest that heavy reliance on AI technologies can lead to cognitive biases and a decline in critical thinking and problem-solving skills. Zhai, Wibowo, and Li (2024) investigate the potential impacts of AI dialogue systems on students' cognitive abilities, highlighting the risks of reduced critical thinking and creativity, and offer suggestions for mitigating these challenges in educational settings. Finlay and Takeda (2021) and Williams and Lim (2024) broadly address the impact of AI on decision-making and its influence on human feelings, thoughts, and behaviors.

Cognitive Offload examines the social and cultural implications of utilizing AI. Often, people accept AI suggestions without sufficient scrutiny in various contexts, especially when the AI fails to offer simple and understandable explanations. The following SDGs are related: SDG 4, which aims to ensure quality education, emphasizes that this uncritical acceptance can undermine critical learning, which is essential for practical and formative education. SDG 16, which promotes peace, justice, and strong institutions, addresses the need for more transparent and understandable decision-making, reflecting the demand for accountability and fairness in institutions that implement AI in society.

Meyerhoff *et al.* (2021) emphasize the need for task-specific strategies rather than a one-size-fits-all approach. Wahn *et al.* (2023) underscore the importance of understanding how task structure and perceived algorithm effectiveness affect decision-making in cognitive discharge.

The study contributes to the broader understanding of human-algorithm collaboration and its potential to improve efficiency in demanding environments.

Table 2. Summary of the Human Intelligence Block

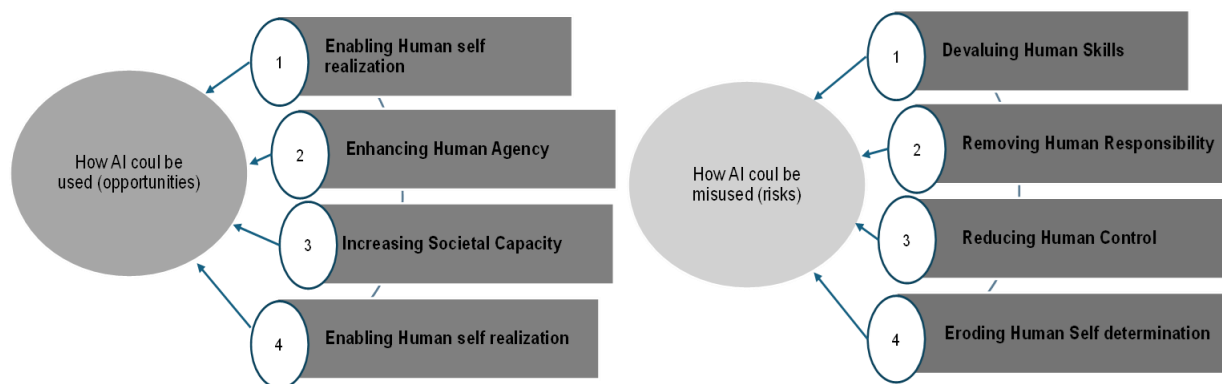
Component	Description	Authors
-----------	-------------	---------

Ethical Principles	Analysis of ethical principles such as fairness, transparency, and accountability, with a focus on equity and mitigation of algorithmic biases	OECD (2019); UNESCO (2021); Horvatić & Lipić (2021)
Cognitive Augmentation	Assessing the potential of AI to enhance human capabilities, such as problem-solving and creativity, while addressing the risks of bias and authenticity.	Humble & Mozelius (2024); Chen & Yadollahpour (2024); Mao et al. (2024); Nyholm (2024)
Cognitive Erosion	Exploring the risks of over-reliance on AI, such as reduced critical thinking, loss of skills, and attention manipulation.	Zhai, Wibowo & Li (2024); Passi & Vorvoreanu (2022); Finlay & Takeda (2021); Williams & Lim (2024)
Cognitive Offload	Study of the social and cultural implications of accepting AI suggestions without critical evaluation, with recommendations to improve human-algorithm collaboration	Meyerhoff et al. (2021); Wahn et al. (2023).

Source: The authors

At the same time, civil society organizations offer crucial perspectives on how AI affects different communities, particularly those that may be vulnerable to algorithmic biases or other unintended consequences. Academia, with its focus on rigorous independent research, is equally vital in helping society understand the broader implications of rapid advances in AI. (Figure 1)

Figure 1. Overview of the top four opportunities offered by AI, four corresponding risks, and the opportunity cost of AI underutilization



Source: Adapted from Cows & Floridi (2018)

2.3. Social Block

This block examines adaptive and inclusive governance strategies to ensure AI systems align with broader societal objectives. Its components function as pillars that ensure that policies and practices reflect society's values and priorities. They are divided into three main components: Impact Assessment, Policy Adjustments, and Monitoring and Evaluation.

Impact Assessment is a component that highlights the relevance of conducting systematic assessments to audit AI systems and mitigate associated risks. This component is key to ensuring that technologies are implemented responsibly and safely. It is mainly related to SDG 10, which aims to reduce inequalities and underlines that impact assessment should consider how technologies affect different social groups. This is crucial to ensure that the benefits of AI are distributed equitably and that any potential harms are identified and mitigated before implementation. In addition, SDG 16, which focuses on peace, justice, and building strong institutions, emphasizes accountability and fairness in impact assessments, reflecting a commitment to developing systems that operate transparently and ethically.

According to Watkins et al. (2021), Stahl et al. (2023), Mantelero and Esposito (2021), and Qian, Sian, and No (2024), the process involves adopting transparent and accessible methodologies that consider both the long-term consequences and the inclusion of diverse stakeholders. Buçinca *et al.* (2023) highlight the challenge of predicting the long-term effects of AI systems by introducing AHA!. (Anticipating Harms of AI), which identifies potential harms before development or deployment. Jiang et al. (2023) and Kazim, Thame, and Koshiyama (2021) examined situational awareness, or the perception of environmental elements, highlighting how applying this concept, as per Siemon and Strohmman (2020), can enhance collaboration between human intelligence and AI, as well as decision-making processes.

Policy adjustments examine the need for adaptable and responsive policies and regulations that

can evolve in response to technological advances and societal changes. This component is crucial in ensuring that public policies respond effectively to the ever-evolving dynamics of society and technology. It is directly related to SDG 9, which promotes industry, innovation, and infrastructure and emphasizes the importance of adaptive policies that encourage innovation and sustainable development. This is essential to keep up with the rapid technological and social transformations that characterize today's world. On the other hand, SDG 17, which focuses on partnerships and means of implementation, emphasizes the importance of collaboration among stakeholders, including governments, academia, and industry. This encouragement of cooperation is crucial for formulating effective and responsive policies that address emerging social needs.

Yar *et al.* (2024) propose a research agenda in their study that applies machine learning (ML) models throughout the entire public policy formulation cycle to measure the performance of these models.

Models in terms of accuracy and precision for specific data. It encourages the adoption of innovative AI and ML approaches to address socioeconomic challenges and increase the effectiveness of public policies. Benneer and Wiener (2019) note that static regulation leads to increasing maladjustments and social losses in the face of rapid technological, scientific, and societal changes. Therefore, adaptive regulation offers a promising alternative.

Monitoring and Evaluation is a fundamental component concerned with continuous monitoring and evaluation to verify the effectiveness of AI systems and governance mechanisms. This includes establishing feedback mechanisms and iterative improvements to ensure transparency, accountability, and responsiveness to emerging challenges.

The relationship of this component with the SDGs is significant. SDG 16, which promotes peace, justice, and strong institution-building, highlights that effective monitoring and evaluation ensure that institutions can be held accountable and that governance practices operate transparently,

promoting an environment of trust and justice in society. Similarly, SDG 4, dedicated to quality education, emphasizes that the continuous feedback and adaptability of AI systems can improve education and skills development by aligning technologies with users' needs. This ensures that everyone has access to informed and ethical decision-making, reinforcing the role of education as a vital pillar for sustainable development.

Boldt and Gilbert (2019) and Emery-Xu, Jordan, and Trager (2024) emphasize the importance of evaluating AI systems based on usability, efficiency, reliability, and accountability for automated decisions, as well as the protection of personal data and intellectual property rights. Amershi et al. (2019), Crawford (2021), and Abedin et al. (2022) discuss ethical considerations in formulating metrics to monitor AI systems, ensuring they align with human values. Palenski (2024), Zaidan and Ibrahim (2024), Chesterman et al. (2023), and De Almeida, dos Santos, and Farias (2021) highlight the complexity.

Interconnectedness between ethical, legal, and socio-political considerations in establishing effective governance mechanisms for AI technologies.

This framework highlights how the various dimensions of the Social Block interact with the SDGs, emphasizing the importance of an integrated approach in AI governance.

Table 3. Summary of Social Block

Components	Description	Authors
Impact Assessment	Conducting audits to mitigate risks and anticipate potential damage with transparent, accessible, and inclusive methodologies.	Stahl et al. (2023), Mantelero & Esposito (2021), Buçinca et al. (2023), Jiang et al. (2023)
Policy Adjustments	Creation of adaptable policies that evolve according to technological advances, promoting collaborative environments and continuous learning.	Yar et al. (2024), Bennear & Wiener (2019)

Monitoring and Evaluation	Establishment of continuous feedback mechanisms and iterative improvements to ensure transparency and ethical alignment of automated decisions.	Boldt & Gilbert (2019). Emery-Xu, Jordan & Trager (2024). Amershi et al. (2019), Crawford (2021), Abedin et al. (2022)
----------------------------------	---	--

Source: The authors

Adaptive AI governance requires integrating technical and ethical approaches to ensure alignment with societal values and promote effective and accountable governance structures.

3. METHODOLOGY

This study employed a mixed-methods approach, combining a narrative literature review with a framework development process, as Grant and Booth (2009) and Booth, Papaioannou, and Sutton (2010) recommended. The literature review was limited to the period from 2019 to 2024, utilizing the Web of Science, ScienceDirect, and Research Rabbit databases. The following terms were used in the research: "Governance mechanisms," "AI challenges," "Human intelligence," and "Conceptual framework."

Including the Boolean operators "and" and "or" allows for refinement to broaden the scope of research and encompass the most significant possible number of results. The review aimed to synthesize existing knowledge, identify gaps in the literature, and inform the development of a new framework. This process was particularly crucial due to the rapid evolution of AI technologies and the associated ethical, social, and cognitive challenges.

The lack of a comprehensive model to address AI's transformative power has been identified as a significant gap in existing research, particularly given the lag in existing legislation regarding new AI capabilities (Calo, 2017; Pasquale, 2015; Gasser & Mayer-Schönberger, 2024). Developing the model involved a three-stage approach (Table 4).

Table 4. Stages of the Model Development Process

Stages	Description
--------	-------------

1. Analysis of the literature	The selected articles were analyzed to identify important themes, concepts, and perspectives on AI governance. The abstracts were reviewed to select the one most closely adhering to the proposal. re
2. Structure Design	This internship systematically organized the central themes and concepts into three blocks: Governance, Human, and Social Intelligence.
3. Refinement of the structure	The proposed framework underwent peer reviews, which provided critical feedback on its comprehensiveness, relevance, and applicability, suggesting improvements and necessary adjustments.

Source: The authors

This methodology enabled a comprehensive analysis of the state of the art in AI Governance while offering a structured format for developing a new model to fill the identified gaps. The resulting model presents a holistic perspective, considering the complexities of AI governance by emphasizing its dynamic and iterative nature.

3.1. Framework Proposal

The analysis of the selected articles made it possible to categorize them into three main Blocks: Governance, Human and Social Intelligence, and their respective components,

In the Governance Block, selected articles were considered that deal with mechanisms, strategies, and policies aimed at regulating the influence of Artificial Intelligence (AI) on human cognition and social structures (Table 5).

Table 5. Governance block

Components	Description
Regulatory Policies	Develop and enforce laws that regulate the use of AI, particularly in sensitive areas such as health, finance, and public safety. Establish industry standards for AI development and implementation through a collaborative process involving industry, academia, and civil society. Ensure that organizations comply with regulatory requirements through audits and reporting.
Social and Educational Reforms	Engage the public and stakeholders in discussions about AI ethics and governance. Promote AI literacy and education to help the public understand the benefits and risks of AI, ensuring that its benefits are distributed equitably across society.

Technological safeguards	Implement security protocols to prevent AI from causing harm and ensure robust cybersecurity measures to safeguard AI systems against malicious attacks. Promote interoperability standards to facilitate the seamless integration of AI systems.
--------------------------	---

Source: The authors

In the Human Intelligence Block, the selected articles that deal with the impact of Artificial Intelligence (AI) on human intelligence were analyzed, with a focus on cognitive processes, as presented in Table 6.

Table 6. Human Intelligence Block

Components	Description
Ethical Principles	Establish clear ethical guidelines for AI development and deployment through independent ethics committees that review AI projects, ensuring transparency in AI algorithms and decision-making processes. Establish clear lines of accountability and put effective redress mechanisms in place when ethical violations occur.
Cognitive Enhancement (Extension of skills)	Enhance human cognitive capabilities by automating tasks, providing valuable insights, and assisting in the solution of complex problems. AI-powered tools such as machine learning algorithms and generative AI can process and analyze large amounts of data far
Components	Description
	Beyond human capability, allowing for faster and more informed decision-making.
Cognitive erosion (obsolescence of skills)	As AI systems take over repetitive or complex activities, humans may lose proficiency in skills previously essential for work, decision-making, or learning.
Cognitive offload (transfer of cognitive tasks)	Ensuring cognitive autonomy and adaptability in the workforce and society while preserving human skills when AI systems perform tasks faster, more accurately, and with less effort. Examining how transferring cognitive tasks from humans to AI systems can free up human cognitive resources for other tasks. However, it can also lead to a reliance on AI for specific functions.

Source: The authors.

In the Social Block, articles on iterative improvement through monitoring, feedback, evaluation, and stakeholder contributions were considered, as shown in Table 7.

Table 7. Social Block

Components	Description
Impact assessment	Address the impact of AI on the labor market through retraining programs and initiatives that promote job creation. Develop economic policies that support innovation while mitigating the economic impacts of AI through funding research and development in AI ethics and governance.
Policy adjustments	Establish specialized government agencies to oversee the development and deployment of artificial intelligence. Promote international collaboration to develop global standards and best practices for AI governance. Promote the involvement of various stakeholders, including industry, academia, and civil society, in AI governance.
Monitoring and Evaluation	Develop metrics to assess the effectiveness of AI governance mechanisms. Implement continuous monitoring systems to track the impact of AI on society. Promote feedback loops to incorporate lessons learned from necessary adjustments to governance mechanisms.

Source: The authors.

3.2. Authors used to build the framework.

Next, for each block, the primary authors who contributed to scientific studies related to each component are highlighted.

Table 8. Distribution of authors by the components of the Governance Block

Components	Authors
Block structure	Huang, <i>et al.</i> (2024); Gasser & Almeida (2017); Taeihagh (2021); Wirtz, Weyerer Sturm (2020); Yeung (2019)
Regulatory Policies	Camilleri (2023); Hogan & Lasek-Markey (2024); Francés-Gómez (2023)
Social and Educational Reforms	Bayly-Castañeda, Ramírez-Montoya and Morita-Alexander (2024). ; Zhai, Wibowo & Li (2024); Tang Kuok Ho, D. (2024).
Collaborative Governance	Eschenbrenner et al (2022); Zaidan & Ibrahim (2024); De Almeida, dos Santos & Farias (2021)
Technological Safeguards	Coeckelbergh (2024); De Almeida, Santos & Farias (2021); Ahmed, Wahed & Thompson (2023)

Source: The authors

The authors in Table 8 ensure that when components work together, AI is developed and used responsibly, ethically, and in alignment with the Sustainable Development Goals. Collaboration

among different sectors, including governments, civil society, academia, and industry, is essential for effective AI governance.

Table 9. Distribution of authors by the components of the Human Intelligence Block

Human Intelligence	Authors
Ethical Principles	Jobin, Ienca & Vayena (2019). Cows & Floridi (2018). Al-Zahrani and Alasmari (2024).; Sifat (2023).
Cognitive Augmentation	Swanepoel (2021b); Zhai, Wibowo & Li (2024); Humble & Mozeliuss (2024)
Cognitive Erosion	Mogi (2024); Bhatt <i>et al.</i> (2023); Gerlich (2024)
Cognitive Discharge	Meyerhoff <i>et al.</i> (2021); Wahn <i>et al.</i> (2023); Boldt & Gilbert (2019)

Source: The authors

The authors in Table 9 ensure that these four components interact to outline a comprehensive understanding of how AI impacts human cognitive capabilities and well-being. It is crucial to recognize both the benefits and risks associated with the interaction between human intelligence and AI, promoting responsible and ethical use of technology.

Table 10. Distribution of authors by the components of the Social Block:

Components	Authors
AI Impact Assessments	Stahl <i>et al.</i> (2023), Mantelero & Esposito (2021), Buçinca <i>et al.</i> (2023), Jiang <i>et al.</i> (2023)
Policy Adjustments	Agrawal, Gans & Goldfarb (2019); Wang & Siau (2018); Androniceanu (2023).
Monitoring and Evaluation	Crawford, 2021; Amershi <i>et al.</i> , 2019; Abedin <i>et al.</i> , 2022.

Source: The authors

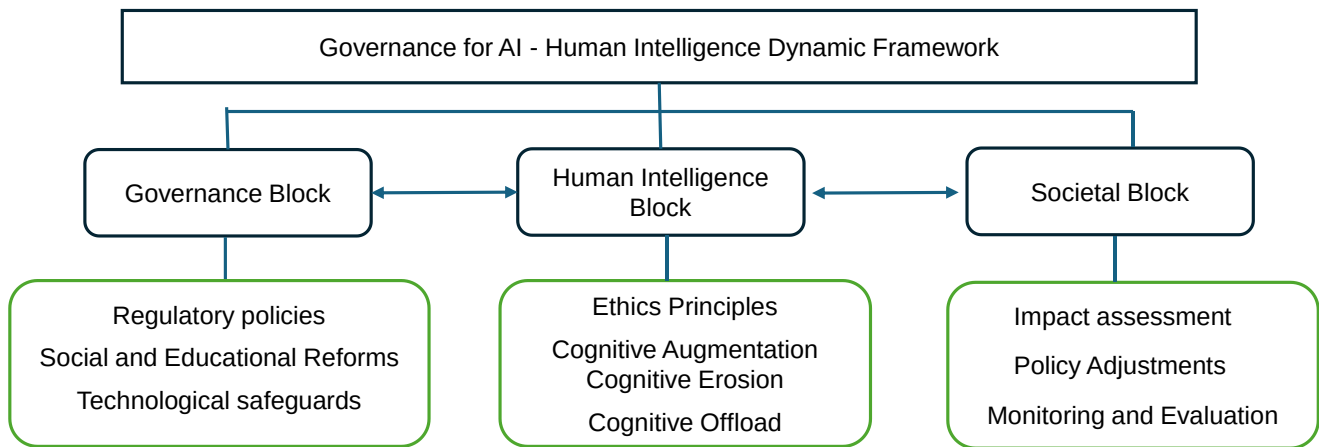
The authors in Table 10 ensure that these three components act as pillars to ensure that AI policies and practices reflect society's values and priorities, promoting fair, inclusive, and sustainable technological development. Integrating these elements enables adaptive governance that considers both societal needs and the evolution of technology, ensuring that AI is used in an ethical and responsible manner.

The need for governments, businesses, and academic institutions to collaborate on creating transparent and comprehensive policies that prioritize human well-being and justice is increasingly evident. This collaboration is essential to address the complex ethical, legal, and social challenges AI poses. By aligning AI regulation with international human rights standards and creating robust and inclusive policies that engage stakeholders from diverse backgrounds, we can ensure that AI technologies are developed and deployed responsibly.

AI governance should be dynamic and adaptable, involving ongoing monitoring and regular adjustments to keep up with technological advancements. Promoting a broad understanding of the technological implications is crucial, as it creates an informed citizenry capable of meaningfully participating in discussions about the development and use of AI. This approach requires comprehensive stakeholder engagement, ensuring that diverse perspectives are considered when developing regulatory strategies.

A proactive governance model is essential, as it anticipates future technology trajectories and potential challenges before they become critical (Figure 2). By reframing technological development as a nuanced journey of creation and responsibility, we can embrace technological progress with curiosity and optimism rather than fear. This perspective suggests that our relationship with technology can be more organic, adaptable, and fundamentally human, as well as collaborative endeavors that reflect our highest aspirations and more thoughtful judgment.

Figure 2. Governance for AI – Dynamic Human Intelligence Framework



Source: The authors

The proposed framework (Fig. 2) is aligned with the Sustainable Development Goals (SDGs) of the UN 2030 Agenda, reinforcing its relevance for developing ethical and responsible AI. Including the SDGs demonstrates concern for social justice, equality, and sustainable development, which is essential in AI governance. It is dynamic and adaptable, recognizing the rapid evolution of AI technologies and the need for governance mechanisms that keep pace with these changes. Adaptability is crucial to addressing emerging challenges and ensuring the framework's relevance over time.

The proposed framework integrates ethical, regulatory, and participatory perspectives, promoting a comprehensive understanding of AI's challenges and opportunities. This multidisciplinary approach is essential for effective governance, considering the complexity of the interaction between AI and human intelligence. Collaboration among diverse stakeholders (governments, academia, industry, and civil society) is essential for the responsible development and implementation of AI. This collaboration ensures that different perspectives are considered and that AI policies reflect society's values and priorities.

The framework addresses the risks of overreliance on AI, such as cognitive erosion and cognitive discharge, by proposing strategies to preserve human capabilities and critical thinking. This focus

on AI's cognitive impacts demonstrates the model's concern with ensuring that technology does not compromise human skills and autonomy. It is built on a comprehensive literature review, including studies from 2019 to 2024. This solid theoretical foundation lends credibility to the model, demonstrating that it is grounded in existing research and knowledge in the field.

Therefore, we highlight some limitations for future studies: The model lacks empirical validation across different contexts, and the absence of practical tests limits the assessment of its effectiveness and applicability in real-world scenarios. Future studies should focus on empirically validating the framework through case studies, experiments, and other research methodologies. The qualitative analysis used to develop the framework may introduce subjectivity and limit the generalizability of the results. Future studies could complement the qualitative approach with quantitative methods to increase the objectivity and generalizability of the results. Although the model aims to be globally relevant, it may need to be adapted to different cultural and social contexts. Future studies could examine the framework's adaptability in various cultural contexts, identifying factors that influence its implementation and adjusting it accordingly to meet local needs.

4. FINDINGS

This research proposes a model structured into three blocks — governance, human intelligence, and social intelligence — to provide a holistic understanding of the complex interaction between artificial intelligence (AI) and human intelligence. The blocks are interconnected, reflecting that while they can operate independently, they also have an essential interdependence. This dynamic relationship enables continuous feedback and highlights the importance of adaptability over time, particularly given the ever-evolving nature of AI. The governance block encompasses the policies, regulations, and guidelines that guide the use and development of AI. It includes establishing ethical standards, ensuring data privacy, and promoting social responsibility.

Governance also involves social and educational reforms as technology evolves, ensuring that AI practices remain aligned with the values and goals of the SDGs.

The human intelligence block focuses on humans' cognitive and emotional capacities when interacting with AI. As Dattathrani and De R. (2022); Fanni et al. (2023) and Felzmann et al. (2020); warn of AI's potential risks to cognitive autonomy and human agency and underline the need for strategies to preserve human capabilities and critical thinking, the block includes understanding human skills such as creativity, empathy, and decision-making, which complement and, in many cases, surpass AI's capabilities. The interaction between human intelligence and AI can lead to more effective and innovative solutions, leveraging the strengths of both.

The social block addresses the interaction between AI and society. Through policy adjustments, it promotes measurements to assess the understanding of how AI affects and is affected by social, cultural, and economic dynamics. Social intelligence examines how AI can address social issues, such as inequality and digital exclusion, and foster social well-being. Additionally, it considers the ethical and social implications of using AI, ensuring that the technology is developed and implemented in a manner that benefits society.

There is interdependence, interconnection, and adaptability between the three Blocks.

Interdependence ensures that human capacities and social dynamics inform governance policies, while ethical and regulatory guidelines guide Human and Social Intelligence. Interconnectedness and adaptability are crucial for ensuring that AI is developed and utilized ethically, responsibly, and beneficially in society.

By considering these blocks in an integrated manner, organizations can fully leverage the capabilities of AI and human intelligence, promoting innovation, problem-solving, and social well-being. Therefore, this framework provides a comprehensive approach to understanding the intricate interplay between AI and human intelligence.

The model's key findings answer the research question: How can governance mechanisms effectively mediate interactions between artificial intelligence (AI) and human intelligence?

Analysis of the Model and related literature revealed key challenges and opportunities, such as

Regulatory reforms

- Effective regulatory policies that guide responsible AI use and balance innovation with ethics.
- Integrated principles and guidelines into AI systems, considering their ethical implications.
- The adoption of a flexible approach, allowing iterative adjustments and regular risk assessments.
- The promotion of user protection and innovation through collaborative, hybrid solutions involving public and private organizations.
- Ensuring transparency and accountability while balancing data protection, intellectual property, and system complexity.

Social and Educational Reforms:

- Enhance public understanding of AI capabilities, limitations, risks, and benefits.
- Collaborate with academia, government, and industry for effective governance.
- Focus on specialized training, public education, and expert development of AI technologies for end users.

Technical Safeguards:

- Adopt explainable AI to promote transparency and accountability.
- Implement algorithmic auditing to increase user trust and mitigate risks.
- Ensure data quality, evaluation, model testing, and techniques like reinforcement learning and human feedback.

The importance of ethical principles was underlined: a) empowering individuals and communities, b) including vulnerable and marginalized groups, c) respecting their autonomy and dignity, d) ensuring that AI decisions do not compromise freedoms and human rights and utilize AI to mitigate social and economic inequality by fostering inclusion, equity, and sustainability through bias-minimizing algorithms. For responsible use, developers and organizations must be accountable for the results of AI decisions by deploying accountability mechanisms and robust governance practices.

Cognitive augmentation through human-artificial intelligence (AI) interaction reflects how machines can complement and enhance human cognitive capabilities. This collaboration: a) optimizes data analysis and information processing in complex and repetitive tasks for humans, b) frees up time for more creative and reflective activities, c) produces more effective and innovative solutions, and can even analyze the impact of economic and health crises in urban spaces and propose solutions to social and economic trends. d) Enhances human cognition in decision-making and productivity, benefiting society. However, there are ethical concerns surrounding the reliance on technology, including privacy and algorithmic discrimination. The key to the success of this interaction between humans and AI lies in understanding the complementary strengths of both human and AI intelligence. Therefore, human-AI collaboration enhances capabilities and redefines how we interact with the world around us, creating new opportunities for cognitive development.

Cognitive erosion can lead to a decline in mental abilities due to the increased use and overreliance on technology, particularly AI. Authors have stated that over-reliance on AI can a) reduce curiosity and the desire to explore new knowledge; b) lead to a cycle of intellectual apathy; c) negatively impact the human ability to think critically and solve problems independently; and d) affect the depth of understanding. Further observation suggests that even reliance on browsers

and calculators can lead to a decline in basic cognitive skills. Therefore, the future use of AI must consider its cognitive and pedagogical implications, encouraging activities that stimulate critical thinking and creativity and ensuring a balanced use of technology that does not compromise essential mental functions.

Cognitive Offload, or the delegation of cognitive functions to AI, involves transferring responsibilities for information processing, decision-making, and problem-solving from the human mind to technological systems. This approach generates both opportunities and dangers, including the reduction of human autonomy and cognitive abilities. It has been found that by making tasks easier to perform, AI can limit people's ability to complete them without technological support, requiring careful attention to avoid this limitation.

Therefore, regulatory bodies, developers, educators, and business leaders must collaborate to create a) a responsible ecosystem that promotes technological innovation and the preservation of fundamental cognitive skills, b) education programs, research on technological dependency, and c) solutions that help maintain cognitive skills are crucial to building a future in which humans and technologies coexist in a balanced way.

Impact Assessments allow us to understand how Artificial Intelligence (AI) affects individuals, groups, organizations, and society. These assessments utilize data analysis and feedback personalization. In education and other sectors, methods such as a) academic performance analysis, b) personalization of learning, and c) administrative efficiency quantify the impact of AI. The technology's most efficient and accurate ability to compile and analyze performance data, offering objective evaluations and personalized suggestions, has emerged. Impact assessments identify the benefits and risks associated with technology, informing decisions and policies related to AI.

Artificial intelligence (AI) governance should be an iterative process; as new information is

gathered and technology evolves, it enables policy adjustments aimed at promoting the well-being of society. This process involves continuous assessment of the impacts of AI policies by documenting, verifying compliance with initial planning, exposing learnings arising from implementation, and communicating the results to ensure that stakeholders are aware of and can manage the associated risks, as well as increasing accuracy in the evaluation of public policies. Additionally, creating evaluation committees that oversee the integrity of AI systems and manage regulatory risks is crucial to ensure that policies are adequate and minimize biased outcomes.

Monitoring and evaluation can make the technology be used following ethical principles and requires continuous collaboration among various stakeholders, ensuring that AI systems: a) are developed and used in a way that reflects a variety of perspectives and values and b) identify risks such as biases and privacy violations, promoting trust and innovation. The integration of monitoring practices enables a) the detection of problems about to arise and b) rapid interventions. Given that AI technology transcends borders and impacts diverse societies differently, AI management must be conducted globally to align regulatory practices and standards, prevent conflicts, and foster coordination.

A framework for governing the interaction between AI and human intelligence that integrates regulatory policies, social and educational reforms, technical safeguards, and ethical principles can address the growing cognitive challenges that arise when individuals rely excessively on AI tools, thereby reducing their engagement in critical thinking and analytical tasks.

After analyzing the findings, we consider the AI governance framework a valid academic proposal, as it aligns not only with the 2030 SDGs but also with the growing need to implement ethical and responsible systems for the development and use of AI. AI governance should be viewed as a comprehensive system of rules, processes, structures, and tools that ensures the responsible use of AI, aligning with organizational principles and ethical and social standards. In addition, adopting

governance frameworks is crucial for mitigating risks, such as bias, discrimination, and privacy violations, while promoting the efficient development of AI.

Several frameworks have been developed to guide the implementation of AI governance, including the European Union's ethical guidelines for trustworthy AI. These guidelines highlight the importance of transparency, accountability, and human oversight throughout the AI lifecycle. Integrating and enforcing these guidelines can lead to better outcomes in organizational goals and increase user acceptance of AI. Thus, the proposal of a governance framework not only facilitates compliance with new regulations but also contributes to a more ethical and responsible use of technology.

5. CONCLUSION

This study aimed to propose a dynamic framework for governing interactions between artificial intelligence (AI) and human intelligence, emphasizing the identification and structuring of mechanisms that moderate the impacts of AI on human cognition, as well as the formulation of guidelines to balance innovation with ethical and social responsibility. With this objective, the study seeks to answer the research question: How can governance mechanisms effectively mediate the interactions between artificial intelligence (AI) and human intelligence?

A new model, comprising three Blocks (Governance, Human, and Social), was presented to understand and manage the complex interaction between AI and human intelligence. In addition to simplistic binary approaches, it sought to focus on the ethical, social, and cognitive dimensions of AI's impact. By examining how these dimensions can moderate the impact of AI on human cognition and society, the study provides a roadmap for striking a balance between innovation and accountability.

The proposed model has some limitations, which the study recognizes. These limitations can be grouped into three main categories. **Empirical validation:** The lack of case studies or practical

applications of the model in real-world scenarios makes it difficult to assess its applicability and relevance. **Qualitative approach:** Although qualitative analysis has provided relevant information, it is subject to the researcher's subjectivity and limits generalization to other contexts. The qualitative nature of the research may limit the ability to replicate or verify results in different settings. **Literature selection:** The narrative review of the literature may have introduced bias in the selection of studies due to limitations in language, publication period, or specific focus areas. Despite the limitations presented, the study highlights the primary contributions of the model to management. The model provides a framework for informed management and decision-making in the field of AI, emphasizing the importance of considering the ethical, social, and cognitive dimensions to ensure that technology is developed and used for the benefit of society.

The study's main contributions include a holistic model aligned with the 17 Sustainable Development Goals of the 2030 Agenda, which integrates the ethical, regulatory, and participatory aspects of AI governance. This model guides the responsible development and implementation of AI technologies, emphasizing dynamic adaptability to address the rapid evolution of AI and unforeseen challenges.

The study emphasizes the significance of multi-stakeholder engagement in developing inclusive and responsible AI. It also addresses the mitigation of cognitive risks, warns of potential threats to cognitive autonomy and human agency, and emphasizes strategies to preserve human capabilities and critical thinking.

This research provides a foundation for future AI governance research and policymaking. Empirical validations and longitudinal studies are crucial for assessing the framework's effectiveness across various scenarios and cultural contexts. Policymakers should develop clear ethical guidelines and regulations to ensure the accountable development of AI, striking a balance between innovation and

the well-being of people.

Despite its contributions, the study has limitations, including the need for empirical validation and potential biases in literature selection. Future research should investigate factors influencing AI-human intelligence interaction, such as cultural diversity and interdisciplinary collaboration, and assess the long-term impacts of AI strategies.

The proposed framework offers a structured approach to address the complexities of AI-human interaction, promoting evidence-based strategies to tackle ethical, cognitive, and social challenges. Continuous refinement and validation in practical scenarios are crucial for the effective implementation of this approach.

Funding Declaration:

This paper is produced as part of a research project conducted within the StudiesDIG project (Models and Instruments for Transforming Higher Education Systems through Transnational Multi-Sector Links), which the European Union funds under the Horizon Europe program.

HORIZON-MSCA-2022-SE-01 Comissão Europeia - HORIZON TMA MSCA

REFERENCES

- Abedin, B., Meske, C., Junglas, I., et al. (2022).** Designing and managing human-AI interactions. *Information Systems Frontiers*, 24, 691–697. <https://doi.org/10.1007/s10796-022-10313-1>
- Abedrabo, W. (2024).** *Establishing Ethical AI: Legal Frameworks and Standards for Algorithmic Fairness and Data Governance*. <https://doi.org/10.13140/RG.2.2.23594.43204>
- Agrawal, A., Gans, J., & Goldfarb, A. (2019).** Economic policy for artificial intelligence. *Innovation Policy and Economics*, 19(1), 139–159. <https://doi.org/10.1086/699935>
- Ahmed, N., Wahed, M., & Thompson, N. C. (2023).** The Growing Influence of Industry on AI Research. *Science*, 379(6635), 884–886. <https://doi.org/10.1126/science.ade2420>
- Al-Zahrani, A.M., & Alasmari, T.M. (2024).** Exploring the Impact of Artificial Intelligence on

Higher Education: The Dynamics of Ethical, Social, and Educational Implications. *Humanities and Social Communications*, 11(1), 912. <https://doi.org/10.1057/s41599-024-03432-4>

Androniceanu, A. (2023). The New Trends in Digital Transformation and Artificial Intelligence in Public Administration. *Administratie si Management Public (RAMP) Journal*, 40, 147–155.

Balasubramaniam, N., & Kujala, S. (2023). Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements. *Information and Software Technology*, 159, 107197. <https://doi.org/10.1016/j.infsof.2023.107197>

Bayly-Castañeda, K., Ramírez-Montoya, M.-S., & Morita-Alexander, A. (2024). Designing personalized learning paths with AI for lifelong learning: A systematic review of the literature. *Frontiers in Education*, 9, 1424386. <https://doi.org/10.3389/educ.2024.1424386>

Benbear, L. S., & Wiener, J. B. (2019). Adaptive regulation: instrument choice for policy learning over time. Available at: <https://www.Has.Harvard.Edu>.

Bhatt, P., Sethi, A., Tasgaonkar, V., Shroff, J., Pendharkar, I., Desai, A., Sinha, P., Deshpande, A., Joshi, G., Rahate, A., Jain, P., Walambe, R., Kotecha, K., & Jain, N. K. (2023). Machine learning for cognitive-behavioral analysis: datasets, methods, paradigms, and research directions. *Brain Informatics*, 10(18). <https://doi.org/10.1186/s40708-023-00196-6>

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, 4(45). <https://doi.org/10.1186/s41235-019-0195-y>

Booth, A., Papaioannou, D., & Sutton, A. (2012). *Systematic approaches to a successful literature review*. Sage Publications.

Buçinca, Z., Pham, C. M., Jakesch, M., Ribeiro, M. T., Olteanu, A., & Amershi, S. (2023). AHA!: Facilitating AI impact assessment by generating examples of harms. *arXiv*. <https://doi.org/10.48550/arXiv.2306.03280>

Calo, R. (2017). Artificial intelligence policy: a primer and a roadmap. *UCDL Rev.*, 51, 399.

- Camilleri, M. A. (2023).** Artificial intelligence governance: ethical considerations and implications for social responsibility. *Systems Experts*, 41(7), e13406. <https://doi.org/10.1111/exsy.13406>
- Cath, C. (2018).** Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 376, 20180080. <https://doi.org/10.1098/rsta.2018.0080>
- Chen, & Yadollahpour. (2024).** A new era in cognitive neuroscience: The tidal wave of artificial intelligence (AI) [Editorial]. *BMC Neuroscience*, 25(23). <https://doi.org/10.1186/s12868-024-00869-w>
- Chesterman, S., Gao, Y., Hahn, J., & Sticher, V. (2023).** The evolution of AI governance. *Review of the Management of Technological Innovation*. <https://doi.org/10.36227/techrxiv.24681063>
- Coeckelbergh, M. (2020).** *AI Ethics*. The MIT Press Essential Knowledge Series. The MIT Press.
- Coeckelbergh, M. (2024).** The case for global AI governance: arguments, counterarguments, and challenges ahead. *AI and Society*. <https://doi.org/10.1007/s00146-024-01949-5>
- Cowls, J., & Floridi, L. (2018).** *Prolegomena to a white paper on an ethical framework for a good AI society*. <http://dx.doi.org/10.2139/ssrn.3198732>
- Crawford, K. (2021).** *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
- Dattathrani, S., & De, R. (2022).** The concept of agency in the age of artificial intelligence: Dimensions and degrees. *Frontiers of Information Systems*, 25(1), 1–26. <https://doi.org/10.1007/s10796-022-10336-8>
- De Almeida, P.G.R., dos Santos, C.D., & Farias, J.S. (2021).** Regulation of artificial intelligence: A framework for governance. *Ethics and Information Technology*, 23, 505–525. <https://doi.org/10.1007/s10676-021-09593-z>

- Emery-Xu, N., Jordan, R., & Trager, R. (2024).** International governance of the advancement of artificial intelligence. *AI and Society*. <https://doi.org/10.1007/s00146-024-02050-7>
- Eschenbrenner, B. L., No, F. F., Siau, K. L., Chen, L., & Qian, Y. (2022).** Towards an integrated framework for the governance of artificial intelligence. In *Annals AMCIS 2022* (Document 19).
- Fanni, R., Steinkogler, V. E., Zampedri, G., & Pierson, J. (2023).** Enhancing human agency through repair in artificial intelligence systems. *IA and Society*, 38(537–547). <https://doi.org/10.1007/s00146-022-01454-7>
- Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020).** Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, 26(6), 3333–3361. <https://doi.org/10.1007/s11948-020-00276-4>
- G20. (2019).** *G20 AI Principles*. G20. Retrieved <https://www.g20.org/>
- Gasser, U., & Almeida, V. A. F. (2017).** A layered model for AI governance. *Internet Computing IEEE*, 21(6), 58–62. <https://doi.org/10.1109/MIC.2017.4180835>
- Gasser, U.; Mayer-Schönberger, V. (2024).** *Protections: guiding human decisions in the age of AI*; Princeton University Press: Princeton, NJ, USA.
- Gerlich, M. (2024).** Balancing Enthusiasm and Cognitive Costs: Trust in AI and the Erosion of Critical Thinking Through Cognitive Offloading. *SSRN*. <https://doi.org/10.2139/ssrn.4994204>
- Grant, M. J., & Booth, A. (2009).** A typology of reviews: An analysis of 14 types of reviews and associated methodologies. *Journal of Health Information and Libraries*, 26(2), 91-108.
- Hogan, L., & Lasek-Markey, M. (2024).** Towards a human rights-based approach to ethical AI governance in Europe. *Philosophies*, 9(6), 181. <https://doi.org/10.3390/philosophies9060181>
- Horvatić, D., & Lipić, T. (2021).** Human-centric AI: The symbiosis of human and artificial intelligence. *Entropy*, 23(3), 332. <https://doi.org/10.3390/e23030332>
- Huang, J., Huang, K., Jackson, K., Zhang, L., & Toren, J. (2024).** Web3 and AI security. In K.

Huang, C. Parisi, L. J. Tan, W. Ma, & Z. W. Zhang (Eds.), *Web3 Application Security and New Security Landscape* (pp. pp. 153-179. Springer, Cham.

Humble, N., & Mozelius, P. (2024). The impact of artificial intelligence on cognitive load in computing education. In *Proceedings of the 19th SweCog Conference* (pp. 23-26). Swedish Society for Cognitive Sciences. ISBN: 978-91-989038-1-2. Available on Uppsala University DiVA Portal

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Rafaghelli, J., & Shimada, A. (2024). Artificial intelligence in education: implications for policymakers, researchers, and practitioners. *Technology, Knowledge and Learning*, 29 (1693–1710).

Jasanoff, S. (2011). *Projects in Nature: Science and Democracy in Europe and the United States*. Course book ed. Princeton University Press, *MUSE Project*,

Jiang, J., Karran, A. J., Coursaris, C. K., Légere, P.-M., & Beringer, J. (2023). A situation awareness perspective on human-AI interaction: Tensions and opportunities. *International Journal of Human–Computer Interaction*, 39(9), 1789–1806.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(1), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

Kazim, E., Thame, D., & Koshiyama, A. (2021). AI audit and impact assessment: According to the UK Information Commissioner's Office. *IA and Ethics*, 1(3), 1–10.

Mantelero, A., & Esposito, M. S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of data-intensive AI systems. *Computer Law and Security Review*, 41, 105561. <https://doi.org/10.1016/j.clsr.2021.105561>

Mao, R., Liu, Q., Li, X., Cambria, E., & Hussain, A. (2024). Cognitive analysis for humans and AI [Guest Editorial]. *Cognitive Computation*, 16 (3316–3318). <https://doi.org/10.1007/s12559-024-10352-7>

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences

in cognitive offloading: A comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications*, 6(1), Article 53.

Mogi, K. (2024). Artificial intelligence, human cognition and conscious supremacy. *Frontiers in Psychology*, 15, article 1364714. <https://doi.org/10.3389/fpsyg.2024.1364714>

Mohanty, S., Arabzadeh, N., Kiseleva, J., Zholus, A., Teruel, M., Awadallah, A., Sun, Y., Srinet, K., & Szlam, A. (2023). Transforming human-centric AI collaboration: Redefining embedded agent capabilities through interactive, reasoned language instructions. *Archiv.org*. <https://doi.org/2305.10783v1>

Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, Summer 2020; Palo Alto: CSLI, Stanford University.

Nyholm, S. (2020). *Humans and robots: ethics, agency and anthropomorphism*. London: Rowman & Littlefield International.

Nyholm, S. (2022a). Artificial intelligence, ethics of. In M. Sellers & S. Kirste (Eds.), *Encyclopedia of the Philosophy of Law and Social Philosophy*. Berlin: Springer.

Nyholm, S. (2022b). A new control problem? Humanoid robots, artificial intelligence and the value of control. *IA and Ethics*, 3, 1229–1239. <https://doi.org/10.1007/s43681-022-00231-y>

Nyholm, S. (2023). *This is the ethics of technology: an introduction*. Hoboken: Wiley-Blackwell.

Nyholm, S. (2024). Artificial intelligence and human enhancement: Can AI technologies make us more (artificially) intelligent? *Cambridge Quarterly of Healthcare Ethics*, 33(1), 76–88. <https://doi.org/10.1017/S0963180123000464>

Nyholm, S., & Rüter, M. (2023). Meaning of life in AI ethics: some trends and perspectives. *Philosophy and Technology*, 36. <https://doi.org/10.1007/s13347-023-00620-Z>

Organization for Economic Cooperation and Development (OECD). (2019). *Council Recommendation on Artificial Intelligence*. OECD

Palenski, T., Hills, L., Unnikrishnan, S., & Eynon, R. (2024). How AI works: reconfiguring lifelong learning. *Post-Digital Science and Education*. <https://doi.org/10.1007/s42438-024-00496-y>

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

Passi, S., & Vorvoreanu, M. (2022). Overreliance on AI: Literature review. *Aether*. Microsoft.

Pflanzer, M., Dubljević, V., Bauer, W. A., Orcutt, D., List, G., & Singh, M. P. (2023). Embedding AI in society: ethics, politics, governance, and impacts. *IA and Society*, 38, 1267–1271. <https://doi.org/10.1007/s00146-023-01704-2>

Qian, Y., Sian, K. L., & No, F. F. (2024). Social impacts of artificial intelligence: ethical, legal and governance issues. *Social Impacts*, 100040. <https://doi.org/10.1016/j.socimp.2024.100040>

Shukla, S. (2024). Principles governing ethical development and deployment of AI. *International Journal of Engineering, Business and Management*, 8(2), 5. <https://doi.org/10.22161/ijebm.8.2.5>

Siemon, D., & Strohmman, T. (2020). Human-AI collaboration: Introducing the virtual collaborator. In *Collaborative Convergence and Virtual Teamwork for Organizational Transformation* (pp. 101-120). IGI Global. <https://doi.org/10.4018/978-1-7998-4891-2.ch005>

Sifat, I. (2023). *Artificial intelligence (AI) and future investment in retail*. RELX Group (Netherlands). <https://doi.org/10.2139/ssrn.4539625>

Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, 11, 12799–12831. <https://doi.org/10.1007/s10462-023-10420-8>

Swanepoel, D. (2021a). Does artificial intelligence have agency? In R. Clowes, K. Gartner, and I. Hippolyte (Eds.), *The Problem of Mind Technology: Investigating 21st Century Minds, Selves, and*

Artifacts (Studies in Brain and Mind Book Series). Springer.

Swanepoel, D. (2021b). The possibility of deliberate adherence to AI standards. *Ethics and Information Technology*, 23, 157–163. <https://doi.org/10.1007/s10676-021-09578-2>

Swanepoel, D., & Corks, D. (2024). Artificial Intelligence and Agency: Tiebreaker in AI decision-making. *Ethics in Science and Engineering*, 30(11). <https://doi.org/10.1007/s11948-024-00476-2>

Taeihagh, A. (2021). Governance of artificial intelligence. *Politics and Society*, 40(2), 137–157. <https://doi.org/10.1080/14494035.2021.1928377>

Tang Kuok Ho, D. (2024). Implications of artificial intelligence for teaching and learning. *Acta Pedagogica Asiana*, 3(21), 65–79. <https://doi.org/10.53623/apga.v3i2.404>

Tocchetti, A., Corti, L., Balayn, A., Yurrita, M., Lippmann, P., Brambilla, M., & Yang, J. (2024). A.I. Robustness: A human-centered perspective on technological challenges and opportunities. *ACM Computing Surveys*. <https://doi.org/10.1145/3665926>

UNESCO (2021). *Recommendation on the ethics of artificial intelligence*. Retrieved December 1, 2024, from <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce over-reliance on AI systems during decision-making. *ACM Proceedings on Human-Computer Interaction*, 7(CSCW1), Article 129.

Wahn, B., Schmitz, L., Gerster, F. N., & Weiss, M. (2023). Offloading under cognitive load: Humans are willing to offload parts of an attentionally demanding task to an algorithm. *PLoS ONE*, 18(5), e0286102. <https://doi.org/10.1371/journal.pone.0286102>

Watkins, E. A., Singh, R., Moss, E., Metcalf, J., & Elish, M. C. (2021). Governing algorithmic systems with impact assessments: Six observations. In *Proceedings of the ACM 2021 Conference on AI, Ethics and Society (AIES '21)* (pp. [page numbers]). ACM. <https://doi.org/10.1145/3461702.346258>.

Williams, G. Y., & Lim, S. (2024). Psychology of AI: How AI impacts the way people feel, think, and behave. *Current Opinion in Psychology*, 58, 101835.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial intelligence and the public sector – Applications and challenges. *International Journal of Public Administration*.
<https://doi.org/10.1080/01900692.2018.1498103>

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration. *International Journal of Public Administration*, 43(9), 818–829. <https://doi.org/10.1080/01900692.2020.1749851>

World Economic Forum. (2024). *Top 10 emerging technologies of 2024*. World Economic Forum.
<https://www.weforum.org/publications/top-10-emerging-technologies-2024>

Yar, M. A., Hamdan, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). Governing with Intelligence: The Impact of Artificial Intelligence on Policy Development. *Information*, 15(9), 556. <https://doi.org/10.3390/info15090556>

Yeung, K. (2019). *A study of the implications of advanced digital technologies (including AI systems) for the concept of accountability within a human rights framework*. DGI (2019)05. Council of Europe.

Zaidan, E., & Ibrahim, I. A. (2024). AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective. *Humanities and Social Communications*, 11(1), Article 1121. <https://doi.org/10.1057/s41599-024-03560-x>

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of overreliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Intelligent Learning Environments*, 11(1), 28.

Zhuk, A. (2024). Navigating the AI Copyright Legal Landscape: A Comparative Analysis of EU, US, and China Approaches. *IA and Ethics*, 4, 1299–1306. <https://doi.org/10.1007/s43681-023-00299-0>