Towards an Ethical Governance of Artificial Intelligence: Issues of Transparency, Accountability and Data Protection

[1] Daanoune Meryem, [2] Khadir Mohammed, [3] Taoufik Wafaa, [4] Rajaa Hadri

[1, 3,4] PHD Student, LRBIGOFE Laboratory, Faculty of Legal, Economic and Social Sciences, Ain Chock, Casablanca.

[2] PHD in Management Sciences - Professor at the Faculty of Legal, Economic and Social Sciences, Ain Chock, Casablanca.

Paper Number: 240204

Abstract :

This article explores the conditions for implementing ethical and transparent digital governance in the era of artificial intelligence. Drawing on a literature review and the analysis of concrete case studies (ChatGPT, Clearview AI, CNSS, etc.), it identifies major challenges in terms of data protection, algorithmic fairness, and legal responsibility.Thearticleproposesa typology of emerging regulatory frameworks (AI Act, UNESCO principles, ISO standards) and formulates concrete recommendations for public and private decision-makers.

Key Words : *Digital governance, artificial intelligence, ethics, data protection, responsibility, algorithmic bias*

1. Introduction

The digital revolution and the rapid advancement of Artificial Intelligence (AI) are reshaping every sector of society (economy, healthcare, education, security, etc.). This transformation simultaneously raises unprecedented issues of trust and accountability. International organizations (UNESCO, OECD, European Union, etc.) emphasize the necessity of designing "trustworthy AI" that respects fundamental human rights. Recent data breaches—such as the massive cyberattack on the Moroccan National Social Security Fund in April 2025, which compromised millions of personal records—illustrate the risks associated with deficient governance.

In this context, digital governance must be redesigned to integrate robust ethical principles. It encompasses the entirety of rules, processes, and stakeholders (States, regulators, corporations, civil society) that organize the use of digital technologies and algorithms. The objective is to maximize the benefits of innovation while mitigating risks (discrimination, privacy

violations, cyberattacks, algorithmic opacity, etc.) and protecting the public interest. This article examines the core principles of ethical AI governance, drawing on academic literature and concrete case studies, to formulate strategic policy orientations.

2. Research Question

How can ethical and transparent digital governance be established to protect fundamental rights against the risks associated with artificial intelligence, while simultaneously fostering technological innovation?

3. Methodology

This article is based on a narrative literature review, drawing on recent scientific papers from indexed academic journals, as well as landmark institutional reports (UNESCO, OECD, CNIL) and regulatory frameworks such as the GDPR and the AI Act. The objective is to cross-reference theoretical contributions with empirical realities, relying on the analysis of emblematic case studies (ChatGPT, Clearview AI, CNSS Morocco) that illustrate the concrete risks linked to a lack of ethical governance. This approach makes it possible to identify current trends in digital governance, evaluate the relevance of existing mechanisms, and propose operational recommendations adapted to the contemporary challenges of artificial intelligence.

4. Data Protection and Privacy in an AI-Driven World

Modern AI thrives on massive datasets, a significant portion of which consists of personal data (information regarding identities, behaviors, communications, etc.). In the race for increasingly high-performance models, this massive data collection raises a major challenge: reconciling technological innovation with respect for privacy (Veale & Edwards, 2018).
Through the GDPR, the European Union mandates principles of lawfulness, transparency, and data minimization, which also apply to AI systems. However, recent cases demonstrate that these principles are being severely tested.

- **ChatGPT and User Data:** Launched in late 2022, OpenAI's ChatGPT rapidly scaled to tens of millions of users, generating a colossal volume of conversations. However, the methods for collecting and utilizing this data raised concerns among regulators. In March 2023, Italy temporarily blocked access to ChatGPT due to alleged non-compliance with personal data legislation. The Italian authority highlighted "the absence of a legal basis justifying the massive collection and storage of personal data for the purpose of training the service's algorithms." This case illustrates the tension between AI data requirements and individual consent. Under pressure, OpenAI has since introduced

options for users to opt out of having their data used for model training. Nevertheless, the ChatGPT–Italy episode demonstrates that GDPR compliance is not a given for consumer-facing generative AI, and that increased vigilance from authorities (the French CNIL is also investigating ChatGPT's practices) remains necessary.

- **Facial Recognition and "Scraping" (Wild Data Collection):** Another emblematic example is Clearview AI, known for its facial recognition tool. Clearview scraped over 20 billion photos available online (notably from social media) to build its facial database without the consent of the individuals concerned. This process is equivalent to creating a massive global biometric database for surveillance purposes. In Europe, regulators have taken action: the French CNIL fined Clearview €20 million (late 2022) and an additional €5.2 million in periodic penalty payments in 2023 for failure to comply (Eubanks, 2018). The authority criticized the company for having no legal basis for this massive collection of sensitive data without consent. Since Clearview has no headquarters in the EU, enforcing these sanctions remains difficult, raising a question of global governance.
- **Other Privacy Issues:** Beyond these cases, the ubiquity of AI poses further challenges to privacy. Voice assistants and connected objects continuously collect information within homes. Predictive models can infer sensitive data (health, orientation, etc.) from digital behaviors. Techniques exist to mitigate risks, such as anonymization and differential privacy (injecting statistical noise to protect individuals within training databases). However, these technical solutions have limits, and the risk of re-identification persists whenever a model handles large amounts of cross-referenced data. Hence the importance of promoting "privacy by design" AI and mandating regular audits of data management practices.
- **Technical Analysis:** From a technical standpoint, data protection in AI questions how models are trained and deployed. Large Language Models (LLMs) like ChatGPT memorize vast text corpora; it has been shown that without safeguards, they can output sensitive information extracted verbatim from the web or training conversations. The opacity of these models ("black box") makes identifying such leaks difficult. Measures such as data filtering before training or the implementation of "kill switches" to prevent a model from responding to certain queries (e.g., requesting personal data) are being explored. Furthermore, homomorphic encryption or federated learning techniques (training AI on local servers without centralizing raw data) could reconcile performance and privacy, though they remain in experimental stages for large-scale deployment (O'Neil, 2016).

- **Societal Analysis:** Societally, privacy is a fundamental right intimately linked to dignity and individual freedoms. In a world of omnipresent AI, the fear is one of generalized surveillance or the abusive exploitation of data. For instance, if every face in public spaces can be identified and tracked by AI, what margin remains for anonymity or simple peace of mind? Recent scandals (e.g., unauthorized use of facial recognition during protests or for dubious commercial purposes) have eroded public trust. According to Eurobarometer surveys, a majority of Europeans express concern about the use of their personal data by companies and public authorities. This mistrust can hinder the adoption of innovative services if they are perceived as intrusive. Conversely, ethical governance that guarantees transparency regarding who collects what, why, and with what safeguards, could strengthen trust and social acceptance of AI (Floridi et al., 2018). We are thus witnessing an increasing demand for algorithmic transparency, through mechanisms like labeling systems that use personal data or the right of access to data even after algorithmic processing (a right enshrined in the GDPR, but often difficult to exercise in practice).

Finally, privacy protection in AI is also a matter of digital inclusion: the most vulnerable populations (those less aware of their digital rights or lacking the means to protect their online privacy) risk being the primary victims of intrusive practices. Ethical governance must ensure that the digital divide in data protection does not widen; for example, by raising public awareness (not just for experts) regarding AI and privacy issues, and by providing accessible remedies in case of abuse (complaints to authorities, class action lawsuits, etc.).

5. Legal and Ethical Accountability of AI Decisions

When AI makes a decision with significant consequences—denying a bank loan, diagnosing a disease, operating an autonomous vehicle, or even moderating online content—who bears the responsibility? This question of accountability regarding algorithmic decisions lies at the heart of digital governance (Mittelstadt, 2021), as it pertains to the capacity of society to oversee and control the tools it creates. Two dimensions are intertwined: legal liability (who is legally responsible in the event of harm or error) and ethical responsibility (the moral obligations of AI designers and users toward society).

Challenges of Accountability: Modern AI systems, particularly those based on machine learning, are unique in their ability to learn from experience and execute tasks in ways that are sometimes unpredictable, even to their creators. These are referred to as "black-box" algorithms, where neither the

internal functioning nor the exact logic behind a decision can be easily explained. This opacity complicates accountability. For instance, if an AI used for CV screening commits discrimination, is the fault with the programmer (who may have unintentionally embedded a bias), the employing company, or the training database (which might reflect historical discrimination)? (CNIL, 2017). Often, there is a gap between the design chain (software providers, data scientists, integrators) and the operational chain (the client company, the agents applying the decision).

To address this "grey area," current legal trends tend to view AI as a tool where the responsible party is the one who utilizes or profits from it (the principle of strict liability or vicarious liability). For example, if an autonomous vehicle causes an accident, the manufacturer or service operator can be held liable, similar to a product defect. Likewise, a company using an algorithm to screen applicants cannot simply blame "an AI error": in France, the Labor Code specifies that automated recruitment decisions must be explainable and justifiable to candidates.

- ✓ **Case Studies of Problematic AI Decisions:**

Algorithmic Bias and Discrimination: Numerous studies have revealed that AI decisions can inherit discriminatory biases (O'Neil, 2016) present in their training data or induced by their design. A prominent example is COMPAS, a predictive algorithm used in the United States to estimate the risk of recidivism among defendants. In 2016, a ProPublica investigation revealed that COMPAS tended to overrate the risk of African American defendants compared to white defendants, demonstrating a systematic racial bias (Binns, 2018).

Specifically, the ProPublica analysis found that black defendants were about twice as likely (44.9% vs. 23.5%) to be mistakenly flagged as higher risk, while white defendants were more likely to be mistakenly labeled as low risk (47.7% vs. 28.0%).

Predictive Policing and Institutional Scandals: Similarly, predictive policing algorithms have disproportionately targeted disadvantaged neighborhoods, creating a vicious cycle of over-surveillance of minorities. In Europe, one of the most striking cases is the Dutch childcare benefits scandal (Eubanks, 2018): an algorithmic fraud detection system unfairly accused thousands of families—primarily from immigrant backgrounds—of fraud based on ethnic profiling criteria. This resulted in a massive scandal and the resignation of the Dutch government in 2021. Amnesty International labeled these algorithms "xenophobic machines" that destroyed thousands of lives by targeting marginalized groups.

These examples highlight the potentially devastating impact of biased AI decisions on fundamental rights (the principle of equality, presumption of innocence, etc.). They also illustrate the critical issue of transparency: often, these biases are only brought to light long after the fact by journalists or researchers auditing the systems, because the models were initially opaque and protected as proprietary information or trade secrets.

Example – Facial Recognition Bias and Discrimination

Figure 2: Visualization of biometric face identification by a facial recognition algorithm.

Studies have shown that these systems misidentify the faces of certain minorities, with error rates potentially 10 to 100 times higher than for white male faces.

Facial Recognition and Technological Bias: Facial recognition algorithms are used for various decisions (identity authentication, police surveillance, access control, etc.). However, their accuracy depends on the data on which they were trained. Tests conducted by NIST (National Institute of Standards and Technology) highlighted alarming performance disparities: for example, certain commercial algorithms misidentified African American female faces up to 100 times more often than white male faces (Floridi et al., 2018), with the lowest error rates generally observed for middle-aged white men. In other words, AI reliability is not uniform, posing a serious problem of equity and indirect discrimination. Specifically, an individual belonging to a poorly recognized group is more likely to fall victim to a false positive (e.g., being wrongly suspected because the AI mistook them for someone else). Several cases have been documented where African Americans in the United States were wrongfully arrested following erroneous facial recognition identifications. These injustices cause a loss of public trust in technologies deployed by police or administrations and have led to moratoriums (San Francisco, Boston, etc.) or strict regulations on facial recognition in certain cities and countries, based on the precautionary principle.

Automated Decisions in the Private Sector: Corporations also face accountability for algorithmic decisions. A frequently cited case is that of Amazon, which developed an internal AI system around 2014 to screen job applicants' resumes. It was soon discovered that the algorithm systematically penalized female candidates: any mention suggesting the candidate was a woman (e.g., "women's chess club" or attending a "women's college") resulted in a lower score. The cause: the model was trained on the company's past resumes, which were predominantly male for technical positions—it had "learned" the gender bias of the tech industry. Amazon attempted to correct the bias without success and ultimately abandoned the project. This episode, along with others (recommendation algorithms showing high-paying job ads preferentially to men, or credit scoring disadvantaging certain zip codes linked to minorities), has highlighted the need for ethical algorithmic impact assessments before deployment. We are increasingly seeing the emergence of independent algorithmic audits and guidelines encouraging models to be tested for bias ("fairness" benchmarks). This is both an ethical responsibility (not deploying an unjust system) and a legal one, as equal treatment legislation applies: a discriminatory AI decision can be challenged in court just as a human decision would be.

6. Analysis and Discussion

6.1 AI Regulation: Key International Actors and Frameworks

At the international level, institutions such as UNESCO and the European Union have proposed ethical frameworks to guide AI development. The goal is to build Trustworthy AI that respects fundamental rights and places humans at the center of every decision.

- **UNESCO**

In November 2021, UNESCO developed the Recommendation on the Ethics of AI, adopted by all 194 Member States. This text defines universal principles (dignity, non-discrimination, transparency, sustainability, accountability, etc.) and recommends governance measures. To implement this, UNESCO launched the Global Observatory on AI Ethics and Governance and an ethics laboratory. It also developed two practical tools: the Readiness Assessment Methodology (RAM) (a macro-tool to evaluate a country's preparedness for ethical AI) and the Ethical Impact Assessment (EIA) (a tool to evaluate the risks and impacts of specific AI systems). Published in 2023, these instruments aim to guide governments in creating regulatory frameworks and action plans. Furthermore, UNESCO regularly organizes international forums (such as the 2024 Global Forum) to share best practices.

❖ **OECD**

Since 2019, the OECD has established AI Principles aimed at promoting innovative, trustworthy AI that respects human rights and democratic values. These principles, revised in May 2024 to incorporate recent advancements, serve as a global benchmark, adopted or adapted by the EU, the Council of Europe, and the United States. To support their implementation, the OECD launched OECD.AI, a global observatory tracking over 1,000 AI policies.

Among its concrete actions, the OECD initiated the HAIP Reporting Framework (within the G7) to enhance the transparency of advanced AI systems, particularly large-scale models. In parallel, it developed an AI Incidents Monitor to track major AI-related failures and is preparing a Trustworthy AI Index. Finally, the OECD facilitates international interoperability—ensuring that public policies can function together coherently and harmoniously on a global scale, even within different regulatory contexts.

❖ **European Union – The AI Act (2024)**

In 2024, the European Union adopted the AI Act, the world's first comprehensive law dedicated to AI. This legislation classifies AI systems into four levels of risk, with stricter rules applied as the level of danger increases.

Minimal Risk	**Transparency Risk**	**High Risk**	**Unacceptable Risk**
No specific controls (anti-spam filters, video games).	Obligation to inform users that they are interacting with an AI (chatbots, image generators).	Strict obligations (AI in healthcare, justice, HR, etc.).	Total Prohibition (social scoring, discriminatory AI).

Implementation began in 2025, but the full set of obligations will become mandatory by 2026. To ensure oversight, the EU has established a European AI Office, and each member state must appoint a supervisory authority.

❖ **United States**

In the United States, there is currently no single federal law governing AI. In 2023, the Biden administration published core principles (security, privacy, human oversight, etc.) to frame AI development. In response, major tech

companies (Google, OpenAI, etc.) have voluntarily committed to securing their systems.
Agencies such as the NIST and the FTC are working on risk assessment and regulation. Furthermore, several states (such as California) have begun adopting their own laws to regulate AI.

- **China**

China aims to become the world leader in AI by 2030. It applies a strict national strategy, combining innovation with reinforced legal oversight.
Several laws regarding cybersecurity and data protection impose heavy obligations on companies: the Cybersecurity Law (2017), the Data Security Law (Sept. 2021), and the Personal Information Protection Law (PIPL). Consequently, specific rules govern generative AI and algorithms, emphasizing national values. Additionally, the state exercises rigorous control through audits, security reviews, and surveillance of sensitive sectors.

- **Morocco**

Morocco is beginning to integrate AI into its digital strategy, with a strong emphasis on ethics and data protection. Law 09-08 (the law relating to the protection of personal data, in effect since 2009) applies to AI systems that process personal information. In March 2025, the CNDP (National Commission for the Protection of Personal Data) issued a statement clarifying that any AI processing involving personal data must adhere to the principles of the law (integrity, transparency, fairness, etc.). Furthermore, the CNDP has initiated consultations with national and international experts to prepare a specific deliberation on AI.

- **International Standards, Labels, and Certifications**

International organizations such as ISO/IEC have developed technical standards (e.g., ISO/IEC 42001) to assist organizations in managing AI responsibly. These standards provide a framework for establishing controlled and verifiable systems, allowing entities to obtain ethical certifications.
To date, according to OECD reports, very few countries have developed official ethical labels for AI systems (with only a few pilot projects in the United Kingdom, Germany, etc.). Consequently, the most reliable benchmarks remain ISO standards, the GDPR, and the voluntary principles established by UNESCO or the OECD.
Major technology firms (Google, IBM, Microsoft, etc.) are already implementing these rules to demonstrate their commitment to developing safer and more ethical AI.

6.2 Exemplary Case Studies

Several recent examples illustrate the challenges and best practices in digital governance:

• **The CNSS Case (Morocco, 2025):** The cyberattack on the National Social Security Fund led to the exfiltration of 54,000 PDF files (approximately 2 million insured individuals' records). The unprecedented scale of this breach sent political and legal shockwaves through Morocco. Analysis reveals several lessons: public officials had not allocated necessary budgets to cybersecurity, despite warnings from the DGSSI (General Directorate of Information Systems Security). Today, the CNSS is being sued for failing to uphold its "duty of vigilance." Institutionally, this incident led to a critical review of national standards (Laws 09-08 and 05-20 and their compatibility with the GDPR) and disciplinary actions against certain officials. The case demonstrates that without proactive governance, even a strategic public body can become a target for state-sponsored or ideological attacks, as Moroccan authorities have asserted.

• **Healthcare Infrastructure (France, 2021):** In September 2021, the IT systems of the AP-HP (Assistance Publique – Hôpitaux de Paris) were hacked, exposing the personal and biological data of 1.4 million people tested for Covid-19. Beyond medical concerns, this incident highlighted the vulnerability of healthcare structures to ransomware. In response, hospitals strengthened their security (business continuity plans, new procedures, staff training), and France accelerated its cybersecurity policy in critical sectors. Ethically, the leak of patient data reignited the debate on health data sharing, arguing for a balance between research (Medical AI) and the informed consent of citizens.

• **Cambridge Analytica (UK/USA, 2018):** Although more political than industrial, this case remains emblematic: a private company transparently collected Facebook data from tens of millions of users for political advertising. The scandal's revelation resulted in a "hard reset" of personal data governance (in-depth audits, strengthening of the GDPR, and Mark Zuckerberg's testimony before Congress). This case shows that a lack of clear ethical and regulatory rules on data can cause a massive loss of trust in digital platforms.

• **Corporate AI (Technology Examples):** Several major corporations have established ethical AI governance bodies (Microsoft, IBM, Google), including internal committees of engineers, lawyers, and philosophers, AI codes of conduct, and internal auditing tools. For instance, IBM publishes the "AI Fairness 360" toolkit to help developers test their models. These voluntary

initiatives illustrate proactive corporate governance, either in anticipation of or as a complement to legal frameworks (Floridi et al., 2020).
These case studies reflect both the urgency of governance (the damage caused by crises) and the path forward (preventive measures, transparency, accountability). They support the idea that ethical digital governance requires close coordination across all sectors.

6.3 Ethical Challenges and Major Risks

The following table summarizes the primary ethical issues posed by AI, their associated challenges, and the solutions implemented to address them:

Issue	Ethical Challenges	Proposed Responses
Data Protection	Personal data leaks, infringement on privacy.	Legislations (GDPR, Law 09-08), supervisory authorities (CNDP), regular audits, security standards (ISO/IEC 27701).
Cybersecurity	Cyberattacks, obsolete infrastructures.	National cybersecurity strategies, adoption of ISO/IEC standards, investment in system security.
Algorithmic Fairness	Discriminatory biases (gender, origin, etc.).	Ethical charters, algorithmic audits, diversification of training data, developer awareness.
Transparency Explainabillity	Unexplainable automated decisions, "black-box" algorithms.	Right to explanation for users, AI codes of conduct, reliability labels, development of explainability tools
Responsibility	Ambiguity regarding liability in case of AI-induced error or harm.	Establishment of ethical committees, clarification of responsibilities within organizations, legal sanctions for damages.
International Governance	Disparate standards, global technological competition.	Adoption of global principles (UNESCO, OECD), multilateral agreements, harmonization of international standards.

Table 1: Ethical AI Challenges and Proposed Responses

➢ **Analysis of Ethical Pillars :**

This table outlines the core ethical issues raised by AI, their corresponding challenges, and the solutions being deployed to mitigate them:

Data Protection: Ensuring the confidentiality and security of personal information.

- ✓ **Responses:** Specific legislation (GDPR, Law 09-08), compliance audits, ISO/IEC 27701 standards.

Cybersecurity: Protecting AI systems against cyberattacks and vulnerabilities.

- ✓ **Responses:** National cybersecurity strategies, adoption of international standards, and security investments.

Algorithmic Fairness: Avoiding bias and discrimination in automated decision-making.

- ✓ **Responses:** Algorithmic audits, diversification of datasets, and ethical **charters.**

Transparency / Explainability: Making the decisions of AI systems understandable for users.

- ✓ **Responses:** The right to an explanation, development of explainability tools, and reliability labels.

Accountability: Clearly defining who is responsible in the event of a malfunction or damage caused by AI.

- ✓ **Responses:** Ethical committees, clarification of liability frameworks, and legal sanctions.

International Governance: Harmonizing standards and regulations globally for ethical AI.

- ✓ **Responses:** Global principles (UNESCO, OECD), multilateral agreements, and international standards.

6.4 Recommendations for Responsible Governance

Based on the analyses and best practices studied, several levers can be activated:

Multi-stakeholder Approach: Governance cannot be unilateral; it must involve governments, regulatory authorities, businesses, NGOs, technical experts, and citizens. Dialogue forums (public/private observatories such as OECD.AI) facilitate the sharing of lessons learned and the construction of common standards. At the national level, multidisciplinary bodies can oversee the implementation of ethical principles by adapting laws and supervising critical systems.

Regulatory Innovation: Beyond existing data protection and cybersecurity laws, AI-specific regulations are necessary. These include mandatory "trust score" labeling for high-risk applications, prior algorithmic review in sensitive sectors (justice, healthcare, social benefits), and the creation of

ethical specifications for public AI tenders. The European Union is advancing these topics through the AI Act and the encouragement of external AI audits. In Morocco, this could inspire a strengthening of the prerogatives of regulators (CNDP, DGSSI) and judicial authorities.

Training and Ethical Culture: Governance relies on the individuals who design and use AI. It is crucial to integrate ethical and security training throughout the project lifecycle (DevSecOps culture, training for data scientists). Organizations must allocate dedicated budgets for cybersecurity and "ethics by design". As noted by the CEO of Deloitte Morocco regarding the CNSS crisis, organizations must move from a reactive to a proactive stance in cybersecurity.

International Standards and Labels: Promoting adherence to global standards (UNESCO, OECD) and participating in international partnerships (Global Partnership on AI) creates a shared framework of trust. Regulatory interoperability is facilitated by the convergence of definitions, such as the OECD's AI lifecycle. Compliance labels, such as ISO/IEC 42001, can certify that companies respect fundamental ethical principles. Furthermore, encouraging responsible innovation through R&D grants ensures technology remains aligned with the public interest.

7. Conclusion

Building ethical and transparent digital governance requires a balance between protecting fundamental rights and encouraging innovation. Recent scandals, such as Italy's temporary block of ChatGPT and Clearview AI's unauthorized scraping of biometric data, highlight the tension between data-hungry AI models and the right to privacy. Moreover, the massive 2023 CNSS data breach in Morocco illustrated that without safeguards, even strategic public bodies remain vulnerable, leading to legal action for breach of "duty of vigilance".

Algorithmic biases represent a major risk to equity. Cases such as the COMPAS recidivism tool, Amazon's biased HR system, and inaccuracies in facial recognition for non-white individuals demonstrate how AI can reproduce societal prejudices. These injustices erode public trust and necessitate rigorous ethical assessments, including independent algorithmic audits and fairness testing, prior to deployment.

The central challenge remains transparency and accountability. The "black-box" nature of many machine learning models complicates the determination of liability when harm occurs. As noted by the CNIL, the lack of clarity regarding the chain of responsibility is a major legal concern. Addressing this through explainability and traceability is essential for social acceptance.

Internationally, regulatory frameworks like the EU AI Act and principles from UNESCO and the OECD are emerging to align AI with human rights. Technical standards like ISO/IEC 42001 further harmonize global requirements for "trustworthy AI". Ultimately, operational ethical governance must act on four levels: organizational (multi-stakeholder bodies), procedural (transparency and audits), human (ethics by design training), and international (multilateral cooperation).

In conclusion, ethical governance does not hinder innovation; it is the prerequisite for its sustainability and inclusion. By proactively protecting fundamental rights and establishing clear accountability, governance inspires the trust necessary for AI to truly serve the public interest. Responsible innovation, backed by shared values, ensures a deployment of AI that is safe, beneficial, and enduring.

Acknowledgements : I would like to sincerely thank the National Center for Scientific and Technical Research (CNRST) for its financial and technical support. I also express my deep gratitude to my research supervisor, Professor **Mohammed Khadir**, for his guidance and continuous support throughout this work.

Bibliograpy :

1. *Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 149–159).*
2. *Batool, A., Zowghi, D., & Bano, M. (2023). Responsible AI Governance : A Systematic Literature Review. ArXiv. arxiv.org*
3. *Comité national pilote d'éthique du numérique (CNPEN). (2023). Systèmes d'intelligence artificielle générative : enjeux d'éthique. www.ccne-ethique.fr*
4. *Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., … & d'Oliveira, N. (2022). Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance. arXiv. arxiv.org*
5. *Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.*
6. *Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … & Vayena, E. (2018). AI4People—An ethical framework for a good AI society : Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.*
7. *Garawit, S., & El Gnaoui, L. (2024). Principes éthiques internationaux de l'intelligence artificielle (IA) : Un essai d'analyse dans l'entreprise au*

Maroc. International Journal of Science, Management and Engineering Sciences (IJSMeS). www.researchgate.net

8. *Khan, A. A., Akbar, M. A., Fahmideh, M., Liang, P., Waseem, M., Ahmad, A., & Abrahamsson, P. (2022). AI Ethics: An Empirical Study on the Views of Practitioners and Lawmakers. arXiv. arxiv.org*
9. *Mittelstadt, B. D. (2021). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 3(10), 754–755.*
10. *O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.*
11. *OCDE. (2024). Principes de l'OCDE sur l'intelligence artificielle : Mise à jour et orientations. www.oecd.org*
12. *Poussart, D. (2021). Gouvernance algorithmique, intelligence artificielle, enjeux, éthique : Esquisse d'une analyse critique. Éthique publique, 23(2). journals.openedition.org*
13. *Rapport CNIL. (2017). Comment permettre à l'homme de garder la main – Les enjeux éthiques des algorithmes et de l'intelligence artificielle. www.cnil.fr*
14. *UNESCO. (2021). Recommandation sur l'éthique de l'intelligence artificielle. unesdoc.unesco.org*
15. *Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party's guidelines on automated decision-making and profiling. Computer Law & Security Review, 34(2), 398–404.*
16. *Wells, G.-P. (2025). Vers une gouvernance éthique de l'intelligence artificielle : enjeux, défis et perspectives. Institut d'études internationales de Montréal. ieim.uqam.ca*