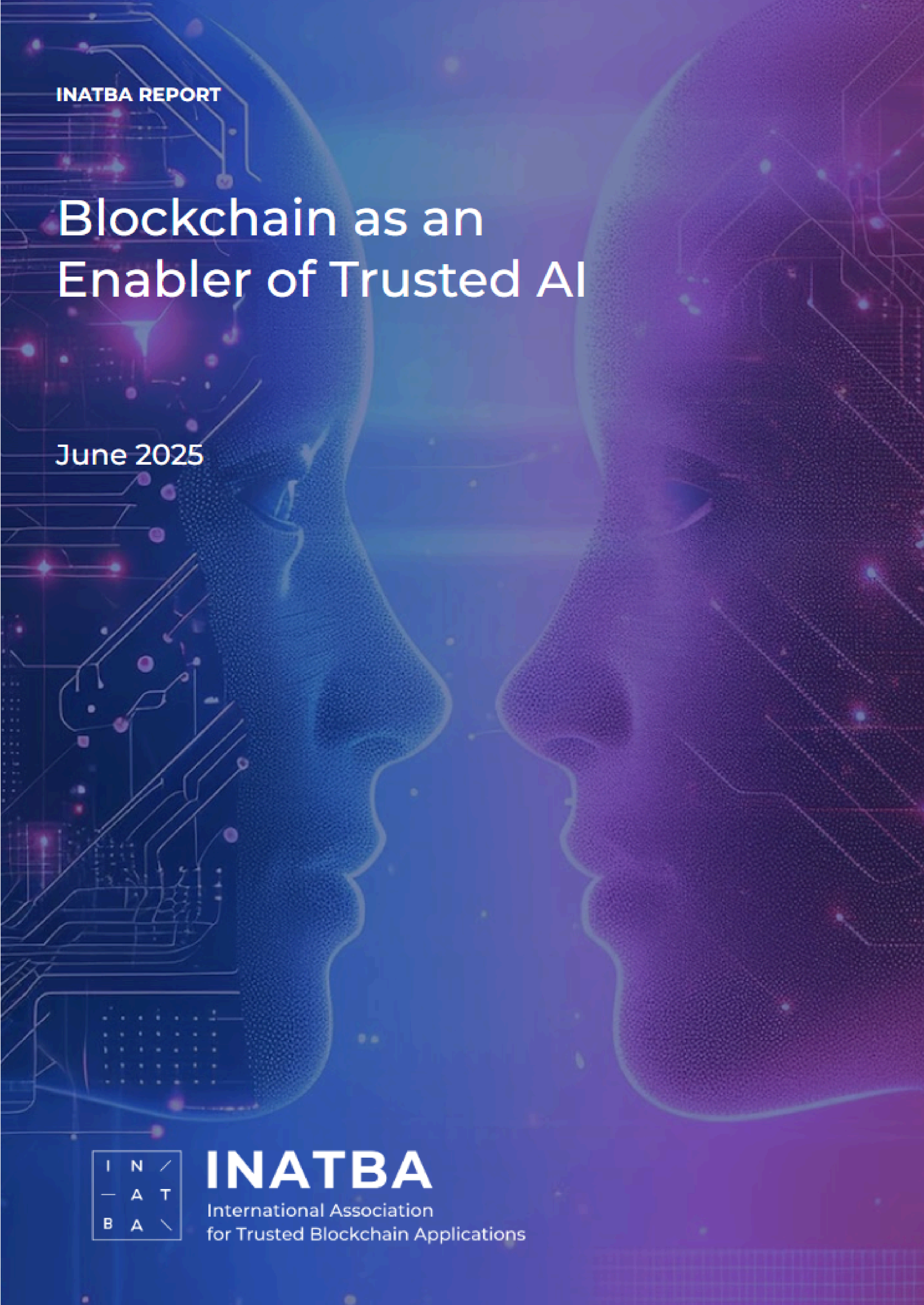

INATBA REPORT

Blockchain as an Enabler of Trusted AI

June 2025

INATBA
International Association
for Trusted Blockchain Applications

Authors

Mariana de la Roche W., INATBA AI & Blockchain Convergences Task Force Co-Chair, BlackVogel

Mat Yarger, Demia

Prof. Dr. Ingrid Vasiliu-Feltes, Institute for Science, Entrepreneurship and Investments

Tan Gürpinar, Quinnipiac University

Reviewers

Antonio Lanotte, EU Blockchain Observatory and Forum

Fabio Budris, INATBA AI & Blockchain Convergences Task Force Co-Chair, BlackVogel

Horst Treiblmaier, INATBA Academic Advisory Body, Modul University Vienna

Jim Mason, EU Blockchain Observatory and Forum

Kalpana Tyagi, INATBA Academic Advisory Body, Maastricht University

Katarina Krüger, INATBA Academic Advisory Body, HTW Berlin

Nina Siedler, thinkBLOCKtank, EU Blockchain Observatory and Forum, siedler legal

Paolo Giudici, INATBA Academic Advisory Body, Università di Pavia

Sharmin Chougule, INATBA Academic Advisory Body, University of Camerino & UNIDROIT

Xiaochen Zhang, AI 2030

Abstract

Artificial Intelligence (AI) is revolutionizing modern life, yet its rapid expansion presents profound ethical challenges to achieving trustworthy AI systems. This report delves into core ethical principles crucial for building trust, ranging from fairness and transparency to privacy and accountability, and highlights both the risks and transformative potential of AI adoption. Drawing on insights from the INATBA AI-BC Task Force and emerging global regulations (such as the EU AI Act), we explore how tools such as blockchain can directly contribute to establishing trusted AI by bolstering its integrity through decentralized governance, immutable audit trails, and enhanced data privacy. Emphasizing environmental, social, and governance (ESG) considerations, the discussion highlights the urgency of building interdisciplinary expertise, fostering transparent regulatory frameworks, and ensuring that AI's development remains firmly rooted in human-centric values, all essential components for fostering trust. This paper offers recommendations for harmonizing innovation with ethical responsibility, steering AI toward a more equitable, sustainable, and trustworthy digital future.

Table of Contents

1. Introduction

Artificial Intelligence (AI) technologies are rapidly reshaping the way we interact with technology, driving innovation across various sectors, including finance and healthcare. Despite its transformative power, AI also creates complex questions about transparency, accountability, and social responsibility. These questions are critical in the context of public awareness around data privacy, confidentiality, security, digital rights, and algorithmic bias. The pursuit of ethical and trustworthy AI underscores the indispensable role of high-quality data. The integrity and fidelity of data inputs are foundational, as AI systems, regardless of their sophistication, can inadvertently amplify biases or inaccuracies if their foundational data is flawed. AI ethics, a multidisciplinary field, aims to ensure that AI technologies respect human values, avoid undue harm, and act as a beneficial force for humanity and the planet. The urgency for robust ethical frameworks intensifies as AI systems influence critical decisions with potentially life-altering consequences in areas such as healthcare, finance, and criminal justice.

In this opening section of this industry report, we discuss the driving forces behind AI's explosive growth, the ethical dilemmas it introduces, and how specialized task forces like the AI-BC Task Force are working to ensure that AI's benefits do not come at the expense of human dignity and societal well-being.

2. Foundational Ethical Principles in AI

At the heart of AI ethics lie several interconnected principles intended to guide the development and deployment of technologies in a manner that respects human rights and fosters public trust. These principles, often articulated as beneficence (doing good), respect for autonomy, fairness, and transparency, offer a moral compass for navigating AI's many applications. While these concepts are crucial to framing ethical AI, translating them into measurable operational metrics remains a complex task. Human rights interpretations can vary across jurisdictions and may at times conflict with national laws. Public trust, though essential, is inherently subjective and shaped by user perception, which may not always align with the actual integrity or capabilities of AI systems. Operationalizing these principles in practice will require standardized metrics, digital guardrails, and policy frameworks that can be implemented at scale.

A widely held view in the tech ethics community is that technology is a tool, and its impact depends on how humanity chooses to apply it responsibly. At INATBA, there is a firm belief that leaders in the blockchain and emerging technology space must proactively ensure that innovation serves a clear and meaningful social purpose. This ethos translates into a commitment to upholding international legal standards and human rights in every endeavor.

"Our mission is to harness the power of blockchain and cutting-edge technologies to create more equitable and sustainable societies everywhere. As such, our commitment to international law and human rights is fundamental to all our endeavors." (INATBA, Commitment to International Law and Human Rights (INATBA, 2023) https://inatba.org/commitment accessed 30 May 2025.))

All technological applications should uphold ethical standards, comply with international law, and protect fundamental rights and freedoms, promoting equality, non-discrimination, and data privacy. Promoting human rights and ensuring digital trust encompasses ensuring that human dignity, equality, and justice guide the design and use of blockchain and Distributed Ledger Technologies (DLT). By fostering an ecosystem rooted in trust, stakeholders can be confident that technological advancement does not infringe upon fundamental human liberties.

A growing international consensus, including initiatives such as INATBA and the OECD AI principles, is emerging around the following core principles:

1. Human, Social, and Environmental Wellbeing (Beneficence/Non-Maleficence): AI systems should benefit individuals, society, and the environment throughout their lifecycle. This includes promoting public well-being, advancing environmental sustainability, and actively avoiding harm, for instance, through early disease detection or optimizing renewable energy distribution.

2. Human-Centered Values (Autonomy): AI systems should respect human rights, diversity, and individual autonomy. Humans should retain meaningful control, especially in high-stakes decisions such as medical treatments or legal outcomes.

3. Fairness: AI systems should be inclusive, accessible, and designed to minimize discrimination by addressing structural and algorithmic bias.

4. Justice: The development and deployment of AI should support equitable access and fair treatment, particularly across regions and demographic groups, ensuring that AI does not exacerbate existing inequalities.

5. Transparency and Explainability: There should be openness about how AI systems operate and make decisions, enabling affected individuals to understand outcomes and challenge automated decisions when necessary.

6. Accountability: Those responsible for AI systems must be identifiable and accountable for outcomes. This includes having control mechanisms in place and external oversight.

7. Privacy and Security: AI systems must protect personal data, adhere to privacy regulations, and ensure robust cybersecurity throughout their lifecycle.

These principles are reinforced through transparent operations, accountability mechanisms, and inclusive collaboration among stakeholders. Maintaining ongoing dialogue with public authorities across jurisdictions is essential for harmonizing regulatory approaches to blockchain and AI, enabling innovation while safeguarding ethical and legal standards.

Sustainability and energy efficiency are also central to ethical technology development. Blockchain applications, when designed with clear guidelines and concrete use cases, can promote not only economic growth but also measurable social and environmental benefits.

In the era of growing cyber threats, aligning AI and blockchain ethics with advanced cybersecurity frameworks such as zero-trust architectures can further strengthen digital resilience. Furthermore, the emergence of quantum computing introduces new security challenges. Integrating quantum-resistant algorithms, such as those recommended by NIST, into blockchain and AI systems is vital for preserving digital trust, privacy, and long-term cyber resilience.

These ethical principles are not abstract ideals but actionable guidelines. While full operationalization remains an evolving process, some initiatives, such as the EU's AI Act regulatory sandbox programs and Estonia's ethical-by-design digital government infrastructure, offer early examples of how principles can be implemented in real-world systems. These cases illustrate how shared moral frameworks can be translated into AI quality monitoring and risk assessment protocols, serving as practical models for responsible innovation in blockchain and AI ecosystems.

As explored further in Section 5, blockchain's transparency, immutability, and decentralized governance offer concrete mechanisms for embedding these ethical principles in practice, contributing to a digital infrastructure that is both trusted and resilient.

3. Key Ethical Challenges

While ethical principles offer a conceptual foundation, real-world AI implementation frequently encounters obstacles such as biased datasets, opaque decision-making processes, and potential privacy breaches. Corporations and governments alike grapple with questions of accountability: Who is responsible when AI systems cause harm? This section reviews the most pressing challenges facing AI developers and regulators, underscoring the importance of robust data governance, multidisciplinary oversight, and consistent compliance with legal standards such as the EU's General Data Protection Regulation (GDPR) or the EU AI Act,

- **Robustness:**

  AI systems must function reliably throughout their lifecycle, including under unexpected or adverse conditions. Ensuring algorithmic robustness helps reduce the risk of failures that could cause harm or misinformed decisions.

- **Safety and Security:**

  Beyond reliability, AI systems must be safe for users and resilient against external threats. This includes building safeguards to prevent adversarial attacks, ensuring secure data handling, and implementing kill-switch mechanisms to override or decommission systems when necessary.

- **Autonomy:**

  While AI can significantly augment human capabilities, it must not undermine human autonomy. Systems should not manipulate users or engage in unjustified surveillance. Preserving meaningful human control, particularly in critical applications such as military, healthcare, or law enforcement, is a non-negotiable ethical boundary.

- **Bias:**

  Algorithmic bias remains a significant and well-documented challenge. AI systems can perpetuate or exacerbate societal inequalities due to skewed training data or flawed design processes. Examples include:

  - Healthcare algorithms that disadvantaged Black patients by using healthcare costs as a proxy for need (AMA, 2022).
  - AI hiring tools that penalized resumes containing the word "women's," effectively discriminating against female applicants (Dastin, 2018).
  - Facial recognition systems that misidentified Black individuals, leading to wrongful arrests (Wessler, 2024)

- **Justice (Global Equity and Access):**

  The deployment of AI technologies raises critical questions about global justice and access. In particular, the structure of the AI value chain, dominated by a few powerful actors with access to large-scale data and computing infrastructure, risks reinforcing systemic inequalities. High-income countries and large corporations often benefit the most, while low-income nations and marginalized communities struggle with limited access to resources, expertise, and reliable infrastructure. This imbalance fosters a digital divide that could result in forms of "AI colonialism" or "data colonialism," where technologies are introduced without local participation or fair benefit-sharing.

- **Global Governance Challenges:**

  Applying ethical standards consistently across jurisdictions is a substantial challenge. Countries vary in their regulatory maturity, commitment to data privacy, algorithmic transparency, and human rights protections. While the EU's GDPR offers a robust model, other regions may lack equivalent safeguards, creating a fragmented and uneven global landscape. This governance gap becomes even more critical when AI is deployed in high-stakes international contexts such as trade, diplomacy, and military operations.

  In the absence of global harmonization, interim governance strategies must reinforce compliance with existing local and regional regulations. While international cooperation is essential to create shared ethical frameworks, AI systems should, in the meantime, be guided by the most rigorous jurisdictional standards applicable in their context.

- **Respect for Human Rights:**

  All AI applications should align with international human rights law and uphold the dignity of individuals. Systems should never be used for discriminatory social scoring or invasive mass surveillance, as these practices undermine fundamental freedoms and erode trust in technology.

The digital divide is a central ethical concern, particularly in the context of digital assets and AI-enhanced financial services. It manifests not only as unequal access to technology or infrastructure, but also as disparities in the ability to navigate complex regulatory environments. Financial institutions play a key role in bridging this gap by offering digital tools, educational resources, and support to underserved communities. Although automated compliance is improving, smaller entities and underbanked populations often lack the technical or financial resources needed to meet evolving obligations. By equipping these actors with accessible solutions and capacity-building initiatives, institutions can help reduce systemic exclusion and promote broader participation in the digital economy.

4. Regulatory Landscape & Governance Models

Global policymakers have begun formalizing AI governance through legislation, ethical guidelines, and strategic frameworks, for instance, the proposed EU AI Act takes a risk-based approach, banning certain ‘unacceptable risk’ practices (like social scoring) and imposing strict requirements for transparency, data quality, human oversight, and risk management on high-risk AI systems. Similar efforts include the UN’s AI-focused resolutions, which align with these ethical imperatives. This article builds on the findings presented in our previous report, *AI Regulation & Blockchain: Bridging Ethics and Governance*, which provides a detailed review of existing and emerging regulations (de la Roche Wills et al., 2024). Here, we focus on how these regulations align with, or sometimes conflict with, ethical imperatives such as privacy, fairness, and accountability. We also evaluate the interplay between self-regulation (industry best practices) and external oversight (governmental or international bodies), highlighting the need for coherent, harmonized frameworks that promote both technological progress and societal values. For more information on the regulatory landscape of AI check our report on AI, Regulation and Blockchain.

5. Blockchain Synergy: Enhancing Ethical AI

Blockchain technology adds a crucial layer of transparency and decentralization to AI ecosystems, helping to resolve trust and data immutability issues. This section draws heavily on the insights from our comprehensive Blockchain & AI Convergence report, where we examined how advanced cryptographic mechanisms (e.g., zero-knowledge proofs) can safeguard data privacy and secure consent management, and how decentralized governance models (e.g., governance within DAOs) can boost fairness in AI-driven applications. By logging critical processes, such as data provenance, model revisions, and user consent, on a tamper-proof ledger, organizations can maintain clearer audit trails. Moreover, a decentralized framework mitigates single points of failure or bias.

Here, we delve into the key findings of that convergence research, illustrating how carefully designed integration can improve accountability and support robust ethical standards, especially regarding data-sharing incentives, AI explainability, and real-time oversight. The foundational attributes of blockchain lay a strong base for ethical AI implementation, enhancing digital trust. Organizations that adopt blockchain features can develop AI systems that demonstrate greater transparency, accountability, and ethical performance [2].

To better illustrate these capabilities, the following set of mechanisms highlights how blockchain's inherent features contribute to ethical AI design and operation:

- Decentralization supports distributed AI governance, preventing any single entity from exerting disproportionate control over AI decision-making. Power distribution across multiple stakeholders ensures that ethical considerations are informed by diverse perspectives. Model updates and ethical revisions can require consensus among ethicists, technical experts, and community representatives.

- Immutability ensures that the history of AI model development, training data, and key decisions remains unaltered. This traceability strengthens accountability by enabling verification of ethical compliance over time.

- Transparent Access allows stakeholders to audit training data, algorithmic logic, and governance protocols. Early detection of potential bias or ethical breaches becomes feasible, facilitating proactive mitigation.

- Adjustable Consensus Mechanisms can ensure that AI model changes or feature deployments align with pre-defined ethical standards before going live.

- Smart Contracts can automate the enforcement of ethical requirements, halting system operations in the case of violations, and logging compliance actions in real time.

- Tokenization enables incentive mechanisms, rewarding participants who identify ethical risks or contribute to compliance efforts.

- Time-stamping provides a chronological record of ethical milestones and changes in the AI lifecycle, ensuring a clear progression of accountability.

- Programmability allows for the embedding of AI quality assessments and ongoing ethical checks directly into system architecture, leveraging existing AI risk management models (OECD, Framework for the Classification of AI Systems (OECD, 2023) https://oecd.ai/en/dashboards/classification accessed 30 May 2025.)

When adopted at scale, these mechanisms benefit from the network effect, reinforcing ethical norms across ecosystems. The application of ethical standards in dual blockchain-AI deployments enhances digital trust by ensuring verifiability, stakeholder accountability, and system resilience.

Despite these synergies, integrating blockchain into AI ethics is not without limitations:

- **Immutability vs. Correction:** Blockchain's unchangeable record may hinder efforts to amend biased or harmful data, conflicting with privacy rights like the GDPR's "right to be forgotten."
- **Privacy Concerns:** Public blockchains may compromise sensitive data used in AI systems. Achieving a balance between auditability and confidentiality remains a key design challenge.
- **Scalability and Performance:** Blockchain throughput may be insufficient for high-volume AI applications without technical optimization.
- **Sustainability:** Consensus mechanisms such as Proof-of-Work raise environmental concerns, especially when combined with energy-intensive AI training models, as both technologies may significantly amplify energy consumption depending on the design of their integration.
- **AI Risk Assessment:** Blockchain can support traceable input/output documentation for ongoing risk monitoring in AI systems.
- **DAO Governance Challenges:** DAOs face their own governance hurdles, including token concentration, low voter participation, and unclear legal frameworks.

These trade-offs emphasize the need for context-aware integration of blockchain within AI governance strategies, acknowledging both its potential and its constraints.

6. DAOs and Web3 for Participatory AI Governance

While blockchain's role in enhancing AI transparency and accountability is well recognized, the broader Web3 ecosystem offers additional tools to support ethical and participatory governance models.

Decentralized Autonomous Organizations (DAOs) in particular offer a promising approach to collective AI oversight. They enable decentralized decision-making on critical matters, such as ethical standards, dataset curation, and model deployment, by involving diverse actors, including developers, ethicists, regulators, and impacted communities. DAOs can fund audits, manage open-source development, and create democratic structures for revising AI practices in alignment with evolving ethical norms.

For instance, a DAO could govern the use of sensitive training datasets in healthcare by:

- Approving inclusion based on ethical criteria,
- Funding third-party bias audits,
- Controlling access via community-approved smart contracts.

Token-based governance mechanisms give stakeholders voting power based on predefined roles or reputational factors. This structure promotes inclusivity while anchoring accountability to transparent and verifiable records. However, DAO models face challenges: low participation, disproportionate influence by "whales," legal ambiguity regarding liability, and vulnerabilities in smart contracts.

To address these risks, DAOs can incorporate:

- Reputation-based or quadratic voting systems to mitigate power imbalance,
- Hybrid governance (e.g., combining expert panels with community voting),
- Transparent conflict-resolution processes supported by smart contract logic.

Beyond DAOs, Web3 identity systems such as Decentralized Identifiers (DIDs) and verifiable credentials can strengthen AI accountability by ensuring traceable provenance for contributions, decisions, and audits. These mechanisms enhance digital trust in multi-agent systems by making every participant's actions verifiable and rights-based, without exposing sensitive personal data.

Together, DAOs, DIDs, and tokenized governance frameworks provide infrastructure for building participatory, transparent, and accountable AI ecosystems, enabling broader public involvement in the shaping of AI's ethical direction.

7. ESG (Environmental, Social, and Governance) Considerations

The environmental footprint of AI, along with its broader social and governance implications, is coming under increased scrutiny. As corporations seek alignment with globally recognized ESG metrics, factors such as AI's energy consumption, carbon output, and impact on workforce dynamics have become central to evaluating corporate social responsibility

- **Tokens and ESG Accountability**

  A token is a digital asset registered within a blockchain infrastructure. These tokens, while often associated with cryptocurrencies, are not interchangeable with them. Blockchain infrastructures ensure that token exchanges occur securely and without intermediaries, enabling use cases far beyond finance. Tokens can represent sustainability claims, carbon credits, or voting rights in governance systems. When integrated into ESG frameworks, they serve as key instruments to measure, incentivize, and reward ethical performance, while facilitating decentralized participation and compliance monitoring.

In this section, we map how the intersection of AI and blockchain can either exacerbate or alleviate inequalities, depending on their design and implementation. We also explore how sustainable AI practices, including energy-efficient hardware and transparent data pipelines, can contribute to broader sustainability goals.

Harmonizing blockchain, AI, and Sustainable Development Goal (SDG) indicators is vital for embedding ethical values into emerging digital infrastructure. This technological convergence allows for platforms that promote responsibility, transparency, and verifiable compliance, while maintaining security and auditability.

- **Automated ESG Compliance via Smart Contracts**

  At the core of this integration is blockchain's smart contract capability. These programmable contracts automate ESG compliance by executing actions based on predefined sustainability metrics. Smart contracts can generate real-time performance reports, trigger alerts for threshold breaches, or enable stakeholder interventions based on environmental performance or ethical benchmarks.

- **Tokenization for Governance and Incentives**

  Tokenized ESG frameworks enable stakeholders to:

  - Measure and verify sustainability practices.
  - Reward ESG achievements with utility or reputation tokens.
  - Participate in governance through weighted or democratic voting rights.

  This approach enhances inclusivity and transparency while aligning stakeholder incentives with long-term sustainability goals.

- **Consensus Protocols for Environmental Accountability**

  Modern proof-of-stake systems reduce the energy demands traditionally associated with blockchain, allowing consensus mechanisms to validate ethical and environmental achievements while incorporating environmental impact directly into governance decisions.

- **Decentralized ESG Data Management**

  Distributed ledger architecture enables cross-organizational collaboration on ESG tracking by:

  - Hosting decentralized repositories for sustainability and ethics data.
  - Preventing unilateral manipulation of records.
  - Ensuring shared accountability across stakeholders.

- **Advanced Cryptography for ESG Verification**

  Quantum-resistant cryptography and zero-knowledge proofs (ZKPs) ensure long-term privacy and secure verification for sensitive ESG data. While powerful in blockchain systems, these technologies are particularly relevant when paired with **AI-driven compliance models** that require privacy-preserving real-time data access. For example, ZKPs allow systems to verify that AI models meet ESG thresholds without disclosing proprietary model details, preserving both trust and IP.

- **Hierarchical and Multi-Signature Governance**

  Multi-signature approval protocols and hierarchical validation systems support collective ESG decision-making across enterprises. These mechanisms offer transparency and traceability for sustainability commitments and compliance records.

- **Interoperability and Unified Reporting**

  Secure interoperability protocols facilitate data sharing across blockchains and institutions. This capability supports the development of unified ESG reporting systems across industries, reducing the burden of redundant data verification and improving traceability of ESG claims.

- **Real-Time Monitoring and Audit Trails**

  Blockchain enables immutable audit trails for real-time monitoring of ESG performance. AI systems can analyze these data streams, detect anomalies or compliance violations, and trigger automated reports or actions, offering a self-regulating layer for ESG accountability.

- **Environmental Impact Monitoring**

  The combined deployment of AI and blockchain can enable organizations to track their energy consumption, carbon footprint, and resource usage with enhanced granularity. These insights are foundational for meeting decarbonization goals and improving efficiency across supply chains.

- **Sustainable Development Alignment**

  Through their combined capabilities, AI and blockchain allow for:

  - Monitoring SDG-aligned performance.
  - Aligning organizational goals with verified ESG metrics.
  - Generating actionable insights for real-time interventions.

- **Zero-Trust Architecture for ESG Integrity**

  Implementing ethical AI and blockchain systems at scale requires infrastructure that enforces standards in real time. **Zero-trust architectures** provide secure, verifiable, and selective data sharing across stakeholders. These features are particularly valuable in complex, multi-party value chains (e.g., mining, metals, or manufacturing), where transparency must be balanced with confidentiality.

  By integrating zero-trust systems with blockchain and AI, organizations can:

  - Establish digital twins to track environmental and social impacts across production lifecycles.
  - Enable Digital Product Passports, EPDs, and compliance with EU frameworks such as the Carbon Border Adjustment Mechanism (CBAM) and Corporate Sustainability Reporting Directive (CSRD).
  - Enforce selective transparency, allowing only necessary data to be shared for audits or certification, while preserving sensitive IP or operational data.

- **Energy Infrastructure: Blockchain, AI, and IoT**

  Decentralized energy management systems combining blockchain and IoT are already being used in applications such as Virtual Power Plants (VPPs) and peer-to-peer microgrid trading [Chougule, 2022]. Building on this foundation, the integration of AI enables real-time optimization of supply–demand flows, improving efficiency and transparency in renewable energy distribution.

  Practical examples include:

  - VPPs that use blockchain to coordinate distributed generation.
  - Microgrids for localized peer-to-peer energy exchange.
  - Spot-market platforms like EPEX for secure, small-scale electricity trading.
  - Smart contracts for predictive diagnostics and encrypted access in IoT-connected smart grids.
  - Blockchain-based coordination of **EV charging and smart battery systems** for more resilient, flexible energy infrastructures.

Ultimately, embedding governance, oversight, and accountability into blockchain–AI–ESG systems mitigates core ethical risks such as data colonialism, opaque decision-making, and inequitable benefit-sharing. It also supports trusted data exchanges, enables participation in low-carbon markets, and reduces the cost of ESG verification.

Rather than viewing compliance as a constraint, this convergence should be seen as a strategic enabler for building resilient, responsible, and competitive digital economies.

8. Recommendations & Future Directions

Striking the right balance between innovation and ethics calls for a multifaceted strategy, involving developers, policymakers, educators, and end-users. Here, we propose actionable steps such as expanding interdisciplinary AI curricula, creating regulatory sandboxes for AI-blockchain experimentation, and promoting open-source libraries for ethical auditing tools. We also discuss potential areas for future research, emphasizing the need for continuous engagement with local and global communities to ensure that evolving AI capabilities adhere to evolving ethical norms.

To enhance digital and societal trust, we propose developing dual AI and blockchain ethics frameworks in conjunction with cybersecurity measures and international standards, such as ISO certifications. Integrating these technologies with customized ethics auditing tools will support transparency and accountability. As these technologies evolve, creating tailored ethics frameworks for agentic and multi-agentic AI and blockchain systems *(multi-agentic meaning systems of multiple interacting AI agents)* will be essential to maintaining autonomy, reducing bias, ensuring equity and safety in their deployment, and making them safe and ethical by design through the use of an appropriate risk management system.

As we continue to explore the ethical implications of AI and blockchain, it is essential to recognize the increasing interplay between these technologies and other emerging fields, particularly immersive technologies like XR (Extended Reality), VR (Virtual Reality), AR (Augmented Reality), and the Metaverse. These technologies are rapidly gaining traction and often leverage AI and blockchain to enhance their functionality. The virtual spaces they create introduce new ethical concerns, particularly around privacy, data security, user autonomy, and the equitable distribution of resources. For example, AI-driven algorithms and blockchain's decentralized systems are being integrated into virtual worlds, enabling users to interact, trade, and create digital identities. These developments raise critical questions about the psychological and societal impacts of immersive experiences. As these technologies evolve, it is crucial to develop new ethical frameworks that align with the principles established for AI and blockchain [9] as well as for the broader Web 3.0 and 4.0 movement (with Web 4.0 referring to the next-generation 'agentic web' of decentralised, autonomous AI-driven ecosystems).

9. Conclusion

Artificial intelligence holds immense transformative potential but simultaneously presents complex ethical challenges that demand careful navigation. Achieving trustworthy AI requires adherence to core principles like fairness, transparency, accountability, privacy, and human oversight. While blockchain technology offers potential synergies for enhancing AI ethics through features like immutability and decentralization, it also introduces significant limitations, particularly concerning bias correction, privacy, and scalability. Addressing AI's environmental footprint and ensuring global equity are critical ESG considerations. Ultimately, a human-centric approach, integrating robust governance, continuous adaptation, multi-stakeholder collaboration, and potentially binding regulations, is essential to steer AI development towards a future that is innovative but also ethical, equitable, and sustainable for all.

10. References / Further Reading

- **AMA.** (2022, March 22). *Feds warned: Algorithms can introduce bias into clinical decisions*. American Medical Association. https://www.ama-assn.org/delivering-care/health-equity/feds-warned-algorithms-can-introduce-bias-clinical-decisions
- **Dastin, J.** (2018, October 10). *Insight: Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG
- **European Commission.** (2022). *AI liability directive proposal*. https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en
- **INATBA.** (2023). *Commitment to International Law and Human Rights*. https://inatba.org/commitment
- **Lanotte, A.** (2023) From Artificial to Circular Intelligence: The Role of Generative AI. https://www.taxnotes.com/special-reports/tax-technology/artificial-circular-intelligence-role-generative-ai/2023/10/13/7hf17
- **OECD.** (2023). *Framework for the classification of AI systems*. https://oecd.ai/en/dashboards/classification
- **de la Roche Wills, M., et al.** (2024). *AI Regulation & Blockchain: Bridging Ethics and Governance*. BlackVogel. https://www.blackvogel.com/_files/ugd/4a504d_c02b574cbcac4edfab8c3e7015c64a04.pdf
- **Wessler, N. F.** (2024, April 30). *Police say a simple warning will prevent face recognition wrongful arrests. That's just not true*. ACLU. https://www.aclu.org/news/privacy-technology/police-say-a-simple-warning-will-prevent-face-recognition-wrongful-arrests-thats-just-not-true
- **European Union. (2024).** *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence and amending certain Union legislative acts.* EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Contact details

Website inatba.org

Contact contact@inatba.org

Join INATBA membership@inatba.org