Article

Trustworthy AI for Whom? GenAI Detection Techniques of Trust Through Decentralized Web3 Ecosystems

Igor Calzada [*] , Géza Németh , Mohammed Salah Al-Radhi

Posted Date: 27 January 2025

doi: 10.20944/preprints202501.2018.v1

Keywords: Generative AI; Decentralization; Web3; Trustworthy AI; Blockchain; DAOs; Data Cooperatives; Big Data; Detection Techniques; Democracy

Article

Trustworthy AI for Whom? GenAI Detection Techniques of Trust Through Decentralized Web3 Ecosystems

Igor Calzada [1,2,3,4,5,6,7,*], Géza Németh [7] and Mohammed Salah Al-Radhi [7]

[1] Public Policy & Economic History Department, Faculty of Economy and Business, University of the Basque Country, UPV-EHU, Oñati Square 1, 20018 Donostia-San Sebastián, Spain
[2] Basque Foundation for Science, Ikerbasque, Plaza Euskadi 5, 48009 Bilbao, Spain
[3] School of Social Sciences, Social Science Research Park (Sbarc/Spark), Wales Institute of Social and Economic Research and Data (WISERD), Cardiff University, Maindy Road, Cathays, Cardiff CF24 4HQ, UK
[4] Decentralization Research Centre, 545 King St. W, Toronto, ON W5V 1M1, Canada
[5] Fulbright Scholar-In-Residence (S-I-R), US-UK Fulbright Commission, Unit 302, 3rd Floor Camelford House, 89 Albert Embankment, London SE1 7TP, UK
[6] Astera Institute, 2625 Alcatraz Ave #201, Berkeley, CA 94705, USA
[7] Department of Telecommunications and Artificial Intelligence, Budapest University of Technology and Economics, ENFIELD Horizon, BEM, 1117 Budapest, Hungary
* Correspondence: igor.calzada@ehu.eus; Tel.: (+34 630 752876; IC)

Abstract: As generative AI (GenAI) technologies proliferate, the need for trust and transparency in digital ecosystems intensifies, especially within democratic frameworks. This article investigates decentralized Web3 mechanisms, specifically those based on blockchain, decentralized autonomous organizations (DAOs), and data cooperatives, to establish robust detection techniques fostering trust in GenAI. These mechanisms are explored against the backdrop of the EU-funded Horizon Europe lighthouse project on Trustworthy AI entitled ENFIELD as foundational elements that support content authenticity, community-driven verification, and data sovereignty, aligning with the EU's AI Act and Draghi Report policy framework. After a state-of-the-art deep analysis, this article presents a multi-layered framework to address the risks associated with AI-generated misinformation encompassing seven detection techniques of trust, including (i) federated learning for decentralized AI detection, (ii) blockchain-based provenance tracking, (iii) Zero-Knowledge Proofs for content authentication, (iv) DAOs for crowdsourced verification, (v) AI-powered digital watermarking, (vi) explainable AI (XAI) for content detection, and (vii) Privacy-Preserving Machine Learning (PPML). This approach not only strengthens AI governance through P2P frameworks but also mitigates the socio-political impacts of AI on public trust, offering a pathway through these seven techniques to allow resilient democratic systems in an era of increasing technopolitical polarization.

Keywords: generative AI; decentralization; Web3; trustworthy AI; blockchain; DAOs; data cooperatives; big data; detection techniques; democracy

1. Introduction: Trustworthy AI for Whom?

The rise of generative artificial intelligence (GenAI) has introduced transformative tools capable of generating complex, human-like content in text, imagery, and sound [1,2]. While these technologies hold vast potential for innovation across industries, they also pose significant risks related to trust, authenticity, and accountability. As the European Commission has globally advanced the AI Act framework to regulate and assure "trustworthy AI," the question of clarifying trustworthy for whom becomes increasingly urgent [3,4]. This inquiry is foundational, especially as GenAI tools permeate sensitive sectors such as healthcare, law enforcement, and governance, where misuse could erode democratic principles, spread misinformation and disinformation, or reinforce biases [5–20].

Generally speaking, the evolution of artificial intelligence (AI), particularly in its generative forms (GenAI), has sparked both admiration for its potential and concerns over its societal impact [21–23]. GenAI's ability to autonomously create text, images, and other forms of content challenges not only the boundaries of creativity but also the very foundations of truth and trust in the digital era [24,25]. GenAI has often been portrayed as a technological marvel, capable of revolutionizing industries and improving efficiencies [11,12]. However, the "elephant in the room," to borrow a metaphor, is the potential for AI to erode democratic systems by flooding information channels with highly persuasive, fabricated content [15,24]. As Amoore et al. highlight [7], the political logic of GenAI transcends mere technicality, embedding itself into the political fabric of societies by altering how information is produced, disseminated, and consumed. This shift challenges democratic institutions that rely on transparency, accountability, and trust in information, prompting urgent questions about the governance of GenAI in decentralized systems [26–41].

Against this backdrop, the European Commission's AI Act, alongside the Draghi Report [3,4], provides a comprehensive policy framework aimed at balancing innovation with ethical oversight in AI systems. The AI Act introduces a risk-based classification model, categorizing AI applications into unacceptable, high, limited, and minimal risk levels, with tailored regulatory measures for each category. This ensures that high-stakes sectors such as healthcare, law enforcement, and governance adhere to stringent requirements for transparency, accountability, and human oversight. Complementing this, the Draghi Report emphasizes AI as a strategic enabler of economic resilience, competitiveness, and sustainability within the European Union, framing AI technologies as infrastructure essential for diverse sectors. It underscores the importance of innovation sandboxes, fostering experimentation while ensuring compliance with ethical standards. Together, these policy frameworks advocate for the development of trustworthy AI systems that align technical standards with societal values, addressing challenges such as data sovereignty, democratic resilience, and public trust in AI-driven systems. The convergence of these policies represents Europe's commitment to navigating the complexities of digital transformation while safeguarding democratic integrity and equity [42–50].

Building on these policy foundations, the research question central to this article is: *Trustworthy AI for whom* (and for what)*?* This inquiry challenges conventional narratives of technological neutrality, emphasizing the need to scrutinize the social, political, and economic implications of trust in AI systems [51]. GenAI models, while transformative, introduce complex challenges, particularly in ensuring transparency, accountability, and equity in their outputs and processes [21]. The notion of "trustworthy AI" must move beyond technical compliance to consider whose trust is prioritized, what ethical frameworks are employed, and how diverse stakeholders—including minority communities—are included in decision-making processes [52–60].

In this context, the role of decentralized Web3 technologies—blockchain [61–66], decentralized autonomous organizations (DAOs) [67], and data cooperatives [68,69]—emerges as a critical countermeasure to the risks associated with centralized AI models [23]. Web3 structures prioritize transparency, data sovereignty, and community participation, aligning with democratic ideals by enabling users to directly influence AI development and governance. By distributing control across peer-to-peer networks rather than within isolated data monopolies, these frameworks offer a pathway to more resilient, socially accountable AI. This article explores these questions through the lens of decentralized Web3 ecosystems, focusing on how technologies like blockchain, decentralized autonomous organizations (DAOs), and data cooperatives can redefine the governance and detection of trust in GenAI systems. By integrating these decentralized mechanisms, the research examines seven key techniques for fostering trust: (i) federated learning for decentralized AI creation and detection, (ii) blockchain-based provenance tracking, (iii) zero-knowledge proofs for content authentication, (iv) DAOs for crowdsourced verification, (v) AI-powered digital watermarking, (vi) explainable AI (XAI) for content detection, and (vii) privacy-preserving machine learning (PPML) for secure content verification. These techniques collectively present a multi-layered framework for detecting and governing GenAI outputs, emphasizing transparency, participatory governance, and data sovereignty. The article positions these approaches as critical to addressing the socio-political risks of AI, including misinformation, disinformation, and democratic erosion [30], while aligning with the broader aspirations of the European Union's AI Act and the Draghi Report. Through this exploration, the title—*Trustworthy AI for Whom? GenAI Detection Techniques of Trust Through Decentralized Web3 Ecosystems*—underscores the urgency of rethinking trust in AI as a shared responsibility that transcends traditional regulatory paradigms.

Web3 refers to a decentralized, blockchain-based ecosystem that enables peer-to-peer networks without reliance on central authorities. The integration of AI into decentralized Web3 ecosystems introduces further complexities [25], as these networks operate without central authority, making traditional forms of governance and control inadequate [26,27]. The proliferation of AI-generated content poses profound implications for democratic integrity [28–30], as the line between real and synthetic content blurs, creating fertile ground for misinformation and disinformation. Furthermore, these challenges are compounded by the unsustainability of current data ecosystems that underlie GenAI [16], requiring innovative strategies for navigating the contradictions between digitalization and sustainability [55].

This introduction sets the stage for an exploration of *trust detection techniques* within decentralized Web3 ecosystems that can enhance GenAI's democratic accountability. Through this lens, we examine not only the technical approaches required to verify AI-generated content but also the socio-political imperatives of establishing a multi-layered trust infrastructure. This analysis is rooted in the collaborative efforts under the umbrella of the Horizon Europe-funded project ENFIELD and within the AI4SI research programme supported by the Basque Foundation for Science, which focuses on leveraging Web3 detection techniques to ensure that AI applications serve public trust, transparency, and resilience, especially within urban and governance contexts, while being committed to leveraging the social impact of the European regulation, including the AI Act and the Draghi Report, by contextualizing them in each specific regional uniqueness but being committed to fundamental rights and to equitable European digital futures.

The development of GenAI technologies has redefined trust, democratizing access to content creation but also amplifying concerns around authenticity and misuse. As these technologies become more pervasive, the challenge lies in determining "trustworthiness" in an environment where AI can impersonate humans and autonomously generate realistic content. This issue is intensified by the varying standards and perceptions of digital trust across cultural, political, and technological contexts, leading to the pressing research question of this article. To address these complexities, decentralized Web3 technologies—blockchain, DAOs, and data cooperatives—have gained attention as tools for fostering transparency and safeguarding democratic integrity in digital spaces [30]. These technologies present an alternative to centralized AI governance by embedding verification mechanisms within peer-to-peer networks, potentially enhancing the reliability of AI-driven content in democratic contexts.

But how to frame the research question *Trustworthy AI for whom*? There are four preliminary considerations and caveats that should be acknowledged before developing the structure of this article:

(i) Recent advances in digital watermarking present a scalable solution for distinguishing AI-generated content from human-authored material. SynthID-Text, a watermarking algorithm discussed by Dathathri et al. [70], provides an effective way to mark AI-generated text, ensuring that content remains identifiable without compromising its quality. This watermarking framework offers a pathway for managing AI's outputs on a massive scale, potentially curbing the spread of misinformation. However, questions of accessibility and scalability remain, particularly in jurisdictions where trust infrastructures are underdeveloped. SynthID-Text's deployment exemplifies how watermarking can help maintain trust in AI content, yet its application primarily serves contexts where technological infrastructure supports high computational demands, leaving out communities with limited resources.

(ii) The concept of “personhood credentials” (PHCs) provides another lens for exploring trust. According to Adler et al. [71], PHCs allow users to authenticate as real individuals rather than AI agents, introducing a novel method for countering AI-powered deception. This system, based on zero-knowledge proofs, ensures privacy by verifying individuals’ authenticity without exposing personal details. While promising, PHCs may inadvertently centralize trust among issuing authorities, which could undermine local, decentralized trust systems. Additionally, the adoption of PHCs presents ethical challenges, particularly in regions where digital access is limited, raising further questions about inclusivity in digital spaces purportedly designed to be “trustworthy.”

(iii) In the context of decentralized governance, Poblet et al. [61] highlight the role of blockchain-based oracles as tools for digital democracy, providing external information to support decision-making within blockchain networks. Oracles serve as intermediaries between real-world events and digital contracts, enabling secure, decentralized information transfer in applications like voting and community governance. Their use in digital democracy platforms has demonstrated potential for enhancing transparency and collective decision-making. Yet, this approach is not without challenges; the integration of oracles requires robust governance mechanisms to address biases and inaccuracies, especially when scaling across diverse socio-political landscapes. Thus, oracles provide valuable insights into building trustworthy systems, but their implementation remains context-dependent, raising critical questions about the universality of digital trust.

(iv) Lastly, the discourse on digital sovereignty, as discussed by Fratini et al. [72], is integral to understanding the layers of trust in decentralized Web3 ecosystems. Their research outlines various digital sovereignty models, illustrating how governance frameworks vary from state-based to rights-based approaches [73–79]. The rights-based model emphasizes protecting user autonomy and data privacy, resonating with democratic ideals but facing practical challenges in globalized digital economies. In contrast, state-based models prioritize national security and centralized control, often clashing with decentralized ethos. These sovereignty models underscore the need for adaptable governance structures that consider the diversity of trust needs across regions, reflecting the complexities of fostering “trustworthy” AI in decentralized contexts.

After presenting the research question in this introduction, a European policy analysis will be carried out in the next section around Trustworthy AI through the AI Act and the Draghi report. Stemming from this European policy analysis, the third section will present the seven techniques for detecting trust as part of the ongoing research project within the framework of the Enfield EU lighthouse project. The final section of the article will present several discussions and conclusions, limitations, and future research avenues.

2. Method: The State-of-the-Art of the European Trustworthy AI Policy Analysis Through AI Act and Draghi Report

This section explores the European Union’s policy response to the challenges and opportunities presented by AI through the lens of two critical and timely documents: the AI Act [4,80] and the Draghi Report [3]. As such, this methodological section aims to frame the research question by conducting a state-of-the-art analysis of the European Trustworthy AI Policy through the AI Act and the Draghi Report. This state-of-the-art analysis underscores the need to position AI as both an enabler of economic growth and a guarantor of ethical, trustworthy, and socially beneficial outcomes. The Draghi Report situates AI as a pivotal driver of economic growth, competitiveness, and resilience across the EU. It articulates a vision where AI extends beyond being a collection of tools to becoming an integral infrastructure underpinning diverse sectors, such as healthcare, urban development, and governance. This transformative potential, however, comes with the shared challenge of ensuring that AI contributes to resilient and sustainable societies while mitigating the risks of democratic erosion and inequality. This is related to the research question of this article when asking, "Trustworthy AI for whom?" [81].

Complementing the Draghi Report, the AI Act introduces a comprehensive regulatory framework aimed at balancing innovation with ethical standards. It emphasizes establishing trust in high-stakes applications such as healthcare, law enforcement, and autonomous systems, where societal impact is profound. The Act outlines a risk-based approach, categorizing AI systems by potential harm and implementing oversight mechanisms accordingly.

2.1. AI Act at the Crossroads of Innovation and Responsibility

The European Union's AI Act represents a landmark effort to balance innovation with societal protection. Its risk-based framework establishes a uniform classification of AI risks across Member States, ensuring consistent governance while allowing flexibility to accommodate national priorities. For technologists, the Draghi Report resonates with its call to align technical advancements with societal priorities, urging stakeholders to build trust into the AI lifecycle. The report challenges policymakers and practitioners alike to leverage AI in ways that bolster economic and social resilience, aligning innovation with responsibility. This regulatory blueprint aims to harmonize technical standards with societal needs, ensuring that innovation does not come at the expense of ethical integrity. By promoting transparency, accountability, and explainability, the AI Act aspires to prevent misuse and discrimination while fostering trust in AI systems. This article explores key aspects of the Act, emphasizing the interplay between uniform standards and localized implementation strategies as follows:

2.1.1. Risk Classification [43,95]: A Unified Framework with Tailored Enforcement

At the heart of the AI Act lies a risk classification system that categorizes AI applications based on their potential harm. This uniform framework ensures that all Member States adhere to a shared baseline for assessing and managing AI risks. However, the enforcement of these classifications may differ based on national priorities [82]. For instance, while all countries are required to address high-risk AI applications, specific sectors may receive heightened attention depending on their relevance to the state's economic or strategic interests, which requires a deep understanding of digital rights regarding whose stakeholders are involved in the trustworthy AI process [96–100].

2.1.2. Human Oversight [101–107]: Enhancing Governance in Critical Sectors

Human oversight remains a cornerstone of the AI Act, ensuring accountability and ethical compliance in AI deployment. Member States have the discretion to amplify oversight measures in sectors critical to their national interests. For example, countries with robust healthcare systems may focus on ensuring transparency and explainability in AI-driven medical applications, whereas others may prioritize oversight in domains such as defence or financial technology.

2.1.3. Innovation Sandboxes: Bridging Compliance and Creativity

To foster innovation while maintaining regulatory compliance, the AI Act encourages the creation of innovation sandboxes—controlled environments where AI technologies can be tested and refined. Countries with strong AI ecosystems or ambitious technological agendas may leverage these sandboxes to accelerate AI adoption while adhering to ethical standards. By providing a space for experimentation, innovation sandboxes allow Member States to remain competitive in the global AI landscape without compromising on safety and trustworthiness.

2.1.4. Sector-Specific Priorities: Aligning AI with Regional Significance

The flexibility of the AI Act extends to addressing sector-specific priorities. Member States are encouraged to focus on sectors of regional significance or those deemed high-risk. For example, Germany, known for its manufacturing prowess, might emphasize AI compliance in industrial automation, while Sweden could prioritize energy sector applications, aligning regulatory efforts with national strengths and challenges. This approach ensures that AI regulations not only address universal concerns but also support localized economic development.

2.1.5. A Unified Vision with Localized Flexibility

The overarching goal of the AI Act is to foster a high level of protection across the EU while enabling Member States to tailor their approaches to AI governance. This dual objective reflects the EU's commitment to harmonizing innovation and responsibility. By accommodating each state's unique priorities and industries, the AI Act provides a framework that promotes both competitiveness and ethical integrity.

2.1.6. Toward a Balanced Future?

As AI continues to reshape societies and economies, the EU's AI Act serves as a model for navigating the complexities of regulation in a rapidly evolving landscape. By blending uniform risk classification with localized flexibility, the Act ensures that Member States can address their unique challenges while contributing to a collective vision of trustworthy and innovative AI.

In Table 1, several key aspects of AI Act are examined including (i) risk classification, (ii) high-risk AI requirements, (iii) transparency obligations, (iii) data governance, (iv) human oversight, (v) compliance and penalties, (vi) innovation sandboxes, (vii) national AI strategies, and (viii) public sector AI applications.

Table 1. European Trustworthy AI Policy Analysis Through AI Act [4].

Aspect	**EU-Wide Application Under AI Act**	**Country-Specific Focus** [3,4]
Risk Classification	AI systems are classified as unacceptable, high, limited, or minimal risk.	Individual states may prioritize specific sectors (e.g., healthcare in Germany, transportation in the Netherlands) where high-risk AI applications are more prevalent.
High-Risk AI Requirements	Mandatory requirements for data quality, transparency, robustness, and oversight.	Enforcement and oversight approaches may vary, with some countries opting for stricter testing and certification processes.
Transparency Obligations	Users must be informed when interacting with AI (e.g., chatbots, deepfakes).	Implementation might vary, with some countries adding requirements for specific sectors like finance (France) or public services (Sweden).
Data Governance	Data used by AI systems must be free from bias and respect privacy.	States with stronger data protection laws, like Germany, may adopt stricter data governance and audit practices.
Human Oversight	High-risk AI requires mechanisms for human intervention and control.	Emphasis may vary, with some states prioritizing human oversight in sectors

		like education (Spain) or labor (Italy).
Compliance and Penalties	Non-compliance can result in fines up to 6% of global turnover.	While fines are harmonized, enforcement strategies may differ based on each country's regulatory framework.
Innovation Sandboxes	Creation of sandboxes to promote safe innovation in AI.	Some countries, like Denmark and Finland, have existing sandbox initiatives and may expand them to further support AI development.
National AI Strategies	Member States align their AI strategies with the AI Act's principles.	Countries may adapt strategies to their economic strengths (e.g., robotics in Czechia, AI-driven fintech in Luxembourg).
Public Sector AI Applications	Public services using AI must comply with the Act's requirements.	Some countries prioritize transparency and ethics in government AI applications, with additional guidelines (e.g., Estonia and digital services).

2.2. Draghi Report

Both the Draghi Report and the AI Act converge on a critical question: How can AI technologies be developed and deployed in ways that support resilient, sustainable economies and societies [83–86]? Central to this inquiry is the recognition that trustworthiness and technical excellence must go hand in hand. From creating inclusive datasets to enabling user-friendly and context-aware applications, these policies highlight the role of AI in shaping not only markets but also democratic norms and citizen engagement. The Draghi Report serves as a critical policy document framing AI as both an enabler of economic growth and a tool for addressing societal challenges. By examining the beneficiaries and potential disparities embedded in the report's vision, we can better understand the socio-political dynamics of AI governance and the pathways toward equitable innovation [108–115].

2.2.1. Trustworthiness Beyond Technological Robustness

The Draghi Report underscores trustworthiness as a fundamental pillar of AI, encompassing transparency, accountability, and ethical integrity [116]. Yet, these values are often interpreted through the lens of technocratic frameworks, emphasizing regulatory compliance and algorithmic reliability. While these aspects are crucial, the focus on technical standards risks sidelining the broader social contexts in which AI systems operate. For whom, and by whom, are these systems deemed trustworthy? This perspective shifts the debate from technological robustness to societal inclusivity, interrogating whether current frameworks adequately address the needs of marginalized or underrepresented groups. The report's call for trustworthy AI resonates with high-stakes domains such as healthcare, education, and public services, but these applications often disproportionately impact vulnerable populations. Trust, in this sense, should be co-constructed through participatory governance mechanisms that empower affected communities to shape the design, deployment, and oversight of AI technologies.

2.2.2. Economic Competitiveness vs. Ethical Equity

A central tension in the Draghi Report is its dual emphasis on fostering economic competitiveness and maintaining ethical AI standards. This tension reflects broader policy challenges within the EU: balancing the need to lead in the global AI race while safeguarding fundamental rights. However, questions arise regarding whose economic interests are prioritized. Large technology firms and well-resourced industries are better positioned to align with regulatory frameworks, whereas smaller enterprises or non-profit initiatives may struggle to compete [29,30]. This uneven playing field raises concerns about the inclusivity of AI-driven economic growth. From a *trustworthy AI for whom* perspective, policies must address these disparities by ensuring that AI innovation benefits a broad spectrum of stakeholders, including SMEs, grassroots organizations, and historically marginalized communities. Initiatives such as innovation sandboxes, proposed in the Draghi Report, offer a promising avenue for bridging this gap. These controlled environments can democratize access to cutting-edge technologies, enabling smaller actors to experiment with AI solutions while adhering to regulatory standards.

2.2.3. Trustworthiness in High-Stakes Sectors

The Draghi Report emphasizes trustworthiness in high-stakes sectors such as healthcare, law enforcement, and energy. While these applications promise transformative benefits, they also entail significant risks of misuse and bias, particularly for minority and marginalized populations. For example, AI systems deployed in healthcare must navigate complex ethical dilemmas, such as balancing personalized treatments with equitable access. Similarly, predictive policing algorithms, often cited as a high-risk application, have been criticized for perpetuating systemic biases [117–119]. The report's focus on risk-based classification is a step toward mitigating these harms, but the implementation of such frameworks requires careful consideration of social dynamics. Trustworthiness in these contexts cannot be reduced to compliance checklists; it demands continuous monitoring, stakeholder engagement, and mechanisms for redress. By involving affected communities in governance processes, policymakers can ensure that AI systems serve as tools for empowerment rather than oppression [87–93]. In several socio-political contexts, as discussed in the Enfield project, presidential elections could also face trustworthiness challenges, particularly when it comes to the manipulation of social media, fake news and post-truth strategies.

2.2.4. Toward a Participatory and Inclusive Vision

The Draghi Report's framing of trustworthy AI implicitly raises the question of inclusion in governance processes (Table 2). Who gets to define what is trustworthy, and whose voices are excluded in these deliberations? The report highlights the importance of public trust in AI systems, but this trust must be earned through meaningful engagement with diverse stakeholders [32,48]. Participatory approaches, such as citizen assemblies, living labs, or co-design workshops, can bridge the gap between policymakers, technologists, and end-users, fostering a shared understanding of AI's societal impact. In conclusion, the Draghi Report provides a robust foundation for advancing trustworthy AI, but its effectiveness will ultimately depend on how inclusively these principles are implemented. By centering the *trustworthy AI for whom* perspective, policymakers can ensure that AI technologies contribute to a fairer, more equitable society. The EU's commitment to ethical AI must go beyond regulatory compliance, embedding principles of inclusivity, equity, and justice into the fabric of AI governance [120–124].

Table 2. European Trustworthy AI Policy Analysis Through Draghi Report [3].

Dimension	Key Insights	Implications

Trustworthiness Definition	Encompasses transparency, accountability, ethical integrity.	Calls for participatory governance to ensure inclusivity and co-construction of trust.
Economic Competitiveness	Tension between fostering innovation and maintaining ethical standards.	Uneven playing fields for SMEs and grassroots initiatives; innovation sandboxes as a potential equalizer.
High-Stakes Sectors	Focus on healthcare, law enforcement, energy; risks of bias and misuse.	Continuous monitoring and inclusive frameworks to ensure systems empower rather than oppress vulnerable populations.
Participatory Governance	Advocates for inclusion via citizen assemblies, living labs, and co-design workshops.	Encourages diverse stakeholder engagement to align technological advancements with democratic values.
Regulatory Frameworks	Balances economic growth with societal equity.	Promotes innovation while safeguarding against tech concentration and ethical oversights.
Challenges in Decentralization	Risks of bias, misinformation, and reduced accountability in decentralized ecosystems.	Emphasizes blockchain and other tech as solutions to enhance accountability without compromising user privacy.
Equitable Innovation	Highlights disparities in economic benefits across industries and societal groups.	Need for policies that ensure AI benefits reach marginalized communities and foster equity.
Technological vs. Societal Context	Debate over prioritizing technological robustness vs. societal inclusivity in trustworthiness.	Shift required towards frameworks addressing underrepresented groups.

2.3. Trustworthy AI for Whom: Approaching from Decentralized Web3 Ecosystem Perspective

This state-of-the-art in this section underscores the importance of creating AI frameworks that are trustworthy, explainable, and aligned with societal needs. As GenAI becomes increasingly influential, there is an urgent need to ensure it serves as a force for social innovation rather than a tool for democratic erosion [30]. By leveraging cutting-edge technologies and adopting a transdisciplinary perspective, the EU's approach—through the Draghi Report and AI Act—offers a model for navigating the challenges posed by GenAI. These frameworks encourage the integration of innovation with ethical safeguards, bolstering the integrity of democratic systems in an increasingly digital world [94].

2.3.1. The Challenges of Detection Techniques for Trust Through Decentralized Web3 Ecosystems

One of the fundamental challenges that GenAI presents to democratic integrity is its capacity to produce content that is indistinguishable from human-generated material [1,2,101]. This is exacerbated in decentralized Web3 ecosystems [125], where information flows without centralized oversight [41]. As Tsai et al. [12] explain in their exploration of GenAI across various domains, the sophistication of AI models has reached a point where they can mimic human creativity, making detection increasingly difficult [126,127]. In a decentralized environment, the lack of a central authority to verify and validate content further amplifies the problem, necessitating the development of robust detection methodologies. Although it remains to be seen, whether decentralization actually implies distributing power or, by contrast, is concentrated in a few tech-savvy elites [128].

In decentralized Web3 ecosystems, such as blockchain-based social media platforms [31,49,64,65,87,88] or DAOs [38], information authenticity and trust are key to maintaining democratic integrity. However, the very nature of these ecosystems—characterized by their peer-to-peer architecture and absence of central control—poses a challenge for ensuring the credibility of content. The traditional mechanisms of verification, such as third-party fact-checkers or centralized content moderation, are ineffective in these spaces. As Magro suggests [97], emerging digital technologies in the public sector, particularly within decentralized networks, require new frameworks for governance, transparency, and trust [71].

2.3.2. GenAI and Disinformation/Misinformation [11]: A Perfect Storm?

The capacity of GenAI to produce convincing, yet false, content creates a perfect storm for disinformation and misinformation. Shin et al. [11] explore how disinformation and misinformation from GenAI influences user information processing, revealing that people often struggle to discern AI-generated misinformation from real content [144]. This cognitive challenge becomes even more problematic in decentralized ecosystems, where the volume of information and the lack of centralized curation make it difficult for users to verify the authenticity of the content they encounter. In the context of democratic systems, this creates a fertile ground for disinformation and misinformation campaigns designed to manipulate public opinion and undermine trust in democratic institutions [129–131].

The ability of GenAI to produce highly realistic yet false content heightens the threat of disinformation campaigns, particularly in decentralized Web3 environments. Such campaigns have already been used by state and non-state actors to influence democratic processes, as seen in the disinformation tactics employed to sway public opinion during the U.S. election [76], probably due to the recent surge of the social media platform BlueSky (a decentralized social media with a friendly user experience). With AI amplifying these manipulative tactics, the scale and reach of disinformation are dramatically increased, making it harder for citizens to differentiate between truthful and misleading information. Furthermore, as the lines between media and tech platforms blur, power is shifting from traditional media moguls to influential figures like Elon Musk and Mark Zuckerberg, whose control over major platforms directly impacts the spread of information [77], avoiding a pluralistic democratic debate beyond polarization [88]. This shift raises concerns over national sovereignty, freedom of speech, and the rule of law, especially as tech companies exert increasing influence over democratic discourse [74]. Moreover, initiatives like legislative efforts to restrict children's access to social media platforms highlight the growing recognition of the dangers posed by unchecked digital environments. The convergence of GenAI and decentralized ecosystems thus necessitates new regulatory frameworks to safeguard the integrity of democratic systems from AI-driven disinformation [132].

As noted by Farina et al. [17], the economic and societal impacts of GenAI tools are profound, with misinformation being one of the most pressing issues. The spread of AI-generated misinformation can erode public trust in democratic processes, as citizens are bombarded with conflicting and false information. This not only skews public perception but also challenges the very foundations of democracy, which rely on informed citizenry and transparent, trustworthy information channels [129,130].

2.3.3. Ethical AI and Accountability in Decentralized Systems

One of the central questions raised by the proliferation of GenAI in decentralized ecosystems is accountability [133]. In traditional, centralized systems, accountability for content lies with publishers, platforms, and regulators [108]. However, in a decentralized Web3 environment, where content creation and dissemination occur in a peer-to-peer manner, assigning responsibility becomes more complex. As Spathoulas et al. [134] discuss, privacy-preserving and verifiable AI processing are essential in maintaining the balance between innovation and accountability, especially in cyber-physical systems that interact with AI-generated content. Roio et al. [64] further highlight how privacy-preserving selective disclosure of verifiable credentials can be employed to enhance accountability without compromising user privacy, ensuring that entities can be held responsible while maintaining anonymity where necessary. Similarly, Adler et al. [71] explore the importance of personhood credentials, emphasizing the value of privacy-preserving tools in distinguishing real individuals from AI-generated entities online. These developments signal a shift towards creating tools and frameworks that address the accountability challenges posed by GenAI in decentralized ecosystems, enabling both trust and privacy.

The challenge of accountability is compounded by the opacity of AI systems themselves [135]. As Guersenzvaig and Sánchez-Monedero [116] argue, the intrinsic values guiding AI research and development are often misaligned with societal needs, creating a disconnect between the creators of AI systems and their users. This disconnect is particularly problematic in decentralized ecosystems, where the absence of oversight mechanisms means that harmful content can spread unchecked, with no clear path for recourse or accountability.

2.3.4. The Role of Blockchain in AI Content Authentication

A promising solution to the problem of accountability in decentralized ecosystems is the use of blockchain technology to trace and authenticate AI-generated content [23,27–31]. Blockchain's decentralized ledger system provides a transparent and immutable record of content creation and modification, making it possible to track the provenance of AI-generated material. As Chafetz et al. suggest [8], the integration of GenAI with open data frameworks can create a new wave of transparency in content creation, allowing users to verify the authenticity of the information they consume [124,136,137].

Digital watermarking techniques, which embed unique identifiers into AI-generated content, can be further enhanced by blockchain technology [102]. These identifiers serve as a form of digital fingerprint, allowing content to be traced back to its source and verified for authenticity. This approach not only enhances transparency but also provides a scalable solution for detecting AI-generated misinformation in decentralized systems, as noted by Estévez Almenzar et al. in their glossary of human-centric AI frameworks [98].

2.3.5. Transdisciplinary Approaches to AI Governance

The complexity of GenAI's impact on democratic integrity necessitates a transdisciplinary approach, integrating insights from fields such as cybersecurity, data ethics, and digital humanities [138]. As highlighted by the United Nations High-level Advisory Body on Artificial Intelligence [103], governing AI for humanity requires a multifaceted perspective that considers not only the technical aspects of AI systems but also their societal implications. This includes addressing issues of bias, fairness, and transparency, all of which are critical to maintaining democratic integrity in decentralized ecosystems [42].

Karatzogianni et al. emphasize the importance of digital citizenship in navigating the ethical and political challenges posed by emerging technologies [50]. In the context of GenAI, this involves creating systems that are not only technically robust but also aligned with democratic values of transparency, accountability, and trust [14,109]. By incorporating these ethical considerations into the design and governance of AI systems, it is possible to mitigate the risks of disinformation and democratic erosion [139].

2.3.6. Addressing the Elephant in the Room

GenAI represents a double-edged sword in the context of decentralized Web3 ecosystems [104]. On the one hand, it offers unprecedented opportunities for innovation, creativity, and efficiency [140]. On the other hand, it poses significant risks to democratic integrity by enabling the widespread dissemination of misinformation and disinformation. Addressing this "elephant in the room" requires a comprehensive, transdisciplinary approach that integrates advanced detection methodologies, blockchain-based content authentication, and ethical AI frameworks [10,19,95,117].

The integration of AI in decentralized data ecosystems [99], particularly through GenAI, has raised significant questions about the future of democratic integrity. Trust in AI systems is crucial for maintaining democratic integrity. In recent years, discussions surrounding AI governance have shifted towards ensuring that AI technologies operate within ethical frameworks that protect human rights, privacy, and public trust. As stated by the partners working in the EU-funded project AI4Gov [141], which focuses on implementing AI in governance structures, the trustworthiness of AI is foundational for democratic processes. In decentralized Web3 ecosystems, trust is even more critical due to the lack of centralized oversight. AI4Gov emphasizes the importance of developing AI systems that are transparent, explainable, and aligned with democratic principles, thereby ensuring that AI serves the public interest rather than undermining it.

Decentralized systems, such as those powered by blockchain technology, pose unique challenges for AI governance. Traditional governance models, which rely on centralized authorities to ensure accountability and transparency, are not applicable in decentralized ecosystems. In line with what Belanche et al. argue [79], this article states that the "dark side" of AI in services can manifest when there is no clear mechanism for accountability, leading to potential abuses of power and erosion of public trust. This issue becomes particularly relevant in decentralized ecosystems, where peer-to-peer networks operate without central authorities to regulate AI-generated content.

Ethical considerations are at the forefront of discussions about GenAI and its impact on democratic systems. Buolamwini highlights the ethical dilemmas posed by AI systems, particularly in terms of bias, discrimination, and the dehumanization of individuals [105]. These ethical concerns are magnified in decentralized ecosystems, where GenAI can produce content without oversight. The absence of centralized control raises questions about how to ensure that AI systems respect ethical boundaries, particularly in terms of fairness, transparency, and accountability.

One of the key ethical challenges of GenAI is its potential to exacerbate existing inequalities. For example, AI-generated content can perpetuate harmful stereotypes, influence public opinion, and manipulate democratic processes. The lack of diversity in AI development teams often leads to biased algorithms that disproportionately harm marginalized communities [110]. In decentralized systems, where there is little oversight, these biases can go unchecked, further undermining democratic integrity [142].

To address these ethical concerns, AI governance frameworks must incorporate principles of fairness and accountability. In line with ongoing advancements by the ENFIELD EU-funded project [143], which explores AI's role in shaping future governance models, decentralized systems must be designed with ethical considerations at their core. ENFIELD project advancements advocate for the development of AI systems that prioritize human rights, transparency, and accountability, ensuring that AI serves as a force for social good rather than a tool for manipulation. The ability of AI to generate content that is indistinguishable from human-produced material raises serious concerns about the spread of false information.

In conclusion, the challenges posed by GenAI in decentralized Web3 ecosystems demand a robust and transdisciplinary approach to maintain democratic integrity and public trust. Building on the EU's regulatory framework through the AI Act and the Draghi Report, this article identifies the necessity of advanced detection methodologies as foundational pillars of trustworthy AI governance (Figure 1). The following seven techniques—federated learning for decentralized AI detection, blockchain-based provenance tracking, Zero-Knowledge Proofs for content authentication, DAOs for crowdsourced verification, AI-powered digital watermarking, Explainable AI (XAI) for content detection, and Privacy-Preserving Machine Learning (PPML)—are presented as a comprehensive framework to counter the risks of misinformation, disinformation, and erosion of accountability. These techniques not only address the technical complexities but also align with ethical principles, offering a pathway for fostering innovation while safeguarding democratic resilience in an increasingly decentralized digital landscape.

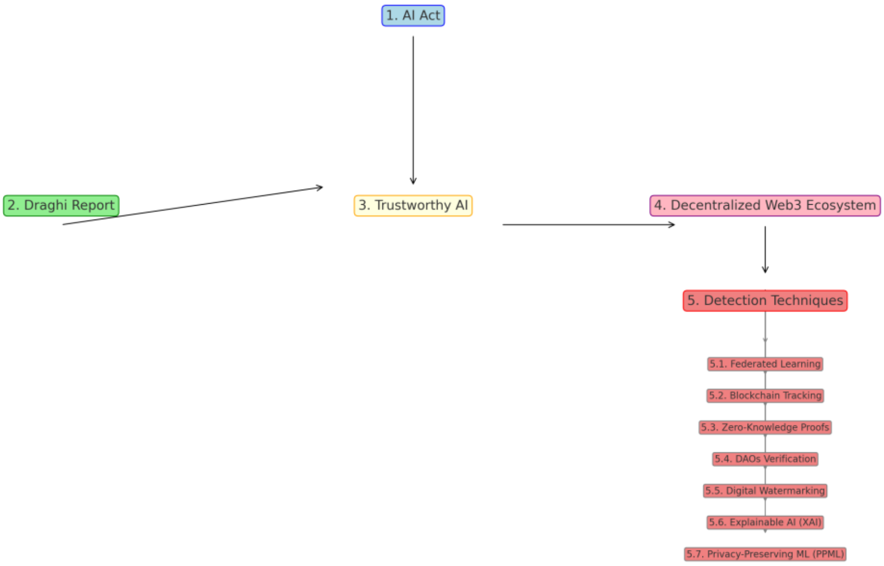
1. AI Act
2. Draghi Report
3. Trustworthy AI
4. Decentralized Web3 Ecosystem
5. Detection Techniques
5.1. Federated Learning
5.2. Blockchain Tracking
5.3. Zero-Knowledge Proofs
5.4. DAOs Verification
5.5. Digital Watermarking
5.6. Explainable AI (XAI)
5.7. Privacy-Preserving ML (PPML)

Figure 1. Bridging Methods (AI Act and Draghi Report) and Results (7 Techniques of Detection of Trust).

2.4. *Justification for the Relevance and Rigor of the Methodology*

The choice to examine the European Trustworthy AI Policy landscape through the dual lenses of the AI Act and the Draghi Report is both highly relevant and methodologically rigorous, given the current challenges posed by GenAI in decentralized ecosystems. For the readership of the journal *Big Data and Cognitive Computing*, which often emphasizes computational and technical advancements, this approach highlights the critical, timely, and novel importance of integrating perspectives beyond pure computer scientific methodologies as the action research process has shown working in a hybrid team of computer scientist and non-computer scientists (i.e., social and political scientists). Addressing societal, ethical, and policy dimensions is not a peripheral concern but a fundamental requirement for ensuring that AI-driven systems are trustworthy, inclusive, and aligned with democratic values. By grounding this research in policy analysis, the methodology enriches the scope of *Big Data and Cognitive Computing*, illustrating how interdisciplinary approaches can bridge the gap between innovation and accountability in AI systems.

2.4.1. Bridging Policy and Practice for Technological Communities

Computer scientists and engineers often focus narrowly on the technological aspects of AI development, potentially overlooking the broader societal implications. By analyzing the AI Act and Draghi Report, this methodology contextualizes the governance mechanisms that underpin AI systems. These policies are not just abstract regulatory frameworks, but actionable blueprints designed to ensure that innovations align with ethical standards and democratic principles.

Highlighting this connection makes the case for why computer scientists should engage with policy, as it directly impacts how their (technical and social) innovations are deployed and regulated in real-world contexts.

2.4.2. The AI Act as a Framework for Risk Classification and Ethical Safeguards

The AI Act introduces a risk-based classification model that is integral to aligning technical innovation with societal protection. By categorizing AI systems based on risk, from minimal to high, the Act enforces stringent measures on high-stakes applications such as healthcare, law enforcement, and governance. For researchers in Big Data and AI, understanding this framework is critical for designing systems that meet these regulatory benchmarks while fostering public trust. This policy-driven alignment bridges the technical and ethical aspects of AI, making the methodology both relevant and indispensable for advancing trustworthiness in AI systems.

2.4.3. The Draghi Report as a Vision for Strategic Resilience

Complementing the AI Act, the Draghi Report positions AI as a strategic enabler of economic resilience and sustainability. This focus on innovation sandboxes and sector-specific priorities offers a roadmap for experimental and regionally adaptive AI systems. For computer scientists, these insights provide a structured way to think about scalable yet responsible innovation. The Draghi Report's emphasis on public trust and ethical equity also aligns with the foundational principles of decentralized ecosystems, making it a valuable reference for designing AI systems that are not only technologically advanced but also socially accountable.

2.4.4. Policy Relevance in Decentralized Web3 Ecosystems

The integration of decentralized Web3 ecosystems into this methodological framework adds another layer of rigor. Web3 technologies such as blockchain, DAOs, and data cooperatives represent cutting-edge solutions for fostering transparency and data sovereignty. By situating these technologies within the policy frameworks of the AI Act and Draghi Report, this approach offers a robust mechanism for addressing the unique challenges posed by decentralized environments, such as misinformation, data misuse, and democratic erosion. This analysis is particularly relevant for computer scientists working on Web3 technologies, as it provides a roadmap for embedding trust and accountability into their systems.

2.4.5. Advancing Detection Techniques of Trust

The methodology extends beyond policy analysis to operationalize its findings through seven advanced detection techniques. These techniques are rooted in state-of-the-art technological solutions. They offer a comprehensive toolkit for countering the risks associated with GenAI, particularly in decentralized contexts. For computer scientists, these techniques are not just theoretical constructs but practical solutions that can be directly implemented in their systems.

2.4.6. A Transdisciplinary Perspective for a Complex Problem

Finally, this methodology is inherently transdisciplinary, combining insights from policy analysis, computer science, and social sciences. This holistic perspective ensures that the research not only advances technical innovation but also addresses the broader societal, ethical, and democratic implications of AI technologies. For reviewers in *Big Data and Cognitive Computing*, this approach underscores the necessity of integrating policy analysis into technological research to create systems that are not only innovative but also equitable and trustworthy.

In conclusion, the methodological framework centered on the AI Act and Draghi Report provides a robust foundation for addressing the complex interplay between technological innovation and societal impact. By bridging policy analysis with practical detection techniques, it offers a comprehensive approach to advancing trustworthy AI in decentralized ecosystems. This methodology is highly relevant for computer scientists, as it equips them with the insights and tools needed to navigate the regulatory and ethical landscape of AI development.

3. Results: Seven Detection Techniques of Trust ThSrough Decentralized Web3 Ecosystems

Building upon the comprehensive analysis of the AI Act [4,80] and Draghi Report [3], which collectively establish the foundational frameworks for European Trustworthy AI governance, the transition from policy to practice becomes imperative. These policies underscore the critical need for AI systems that balance innovation with ethical and societal responsibilities, particularly in decentralized Web3 ecosystems where traditional oversight mechanisms are challenged. The frameworks discussed in the Methods section highlight the EU's commitment to transparency, accountability, and inclusivity. However, addressing the operational challenges of trust in decentralized environments requires actionable methodologies.

To address the trust deficit in GenAI, decentralized Web3 mechanisms offer innovative solutions by leveraging their inherent features of transparency, immutability, and peer-to-peer governance. Blockchain technology provides a robust foundation for establishing content provenance, ensuring that information can be traced to its origin with a transparent, tamper-proof ledger. DAOs facilitate community-driven verification processes, enabling collective oversight that aligns with democratic values and reduces reliance on centralized authorities. Additionally, data cooperatives empower individuals and communities by granting them control over their data, fostering trust through participatory governance and ethical stewardship [23]. Together, these decentralized mechanisms challenge traditional approaches to trust and accountability, offering scalable, resilient frameworks to detect and mitigate AI-generated misinformation and disinformation while maintaining alignment with the ethical imperatives outlined in the AI Act and Draghi Report.

This section introduces seven advanced detection techniques as a practical bridge from the theoretical underpinnings of Trustworthy AI to the operational realities of combating disinformation, ensuring content authenticity, and fostering democratic resilience. These techniques—federated learning, blockchain-based provenance tracking, Zero-Knowledge Proofs, DAOs for crowdsourced verification, digital watermarking, Explainable AI (XAI), and Privacy-Preserving Machine Learning (PPML) [144]—serve as a toolkit to uphold trust, transparency, and accountability in AI applications, aligning with the principles set forth in the AI Act and Draghi Report. By operationalizing these techniques, this article navigates the pathway from policy analysis to tangible solutions that safeguard democratic systems in the face of GenAI's transformative potential [145–148].

These seven detection techniques outlined in this study were systematically identified and developed under the framework of the ENFIELD Horizon Europe project, which seeks to establish trustworthy AI governance through innovative, decentralized methodologies. ENFIELD, bringing together computer scientists and political and social scientists, provides a transdisciplinary platform that integrates insights from policy, technology, and societal impact to tackle the challenges of AI in decentralized Web3 ecosystems. These techniques—federated learning, blockchain-based provenance tracking, zero-knowledge proofs, DAOs, digital watermarking, explainable AI, and privacy-preserving machine learning—were chosen for their alignment with the project's mission to foster transparency, accountability, and participatory governance. Each technique was rigorously evaluated in terms of its applicability to the EU's AI Act and Draghi Report, ensuring they are both operationally feasible and ethically sound. This integration underscores ENFIELD's commitment to bridging theoretical frameworks with real-world applications, enabling scalable solutions that address misinformation, democratic erosion, and public trust deficits in AI systems. By embedding these detection mechanisms into the broader policy landscape, the ENFIELD project not only operationalizes the principles of the AI Act and Draghi Report but also positions Europe as a global leader in trustworthy AI governance, while in North America, there is a strong appetite for such technical and social experimentation [11,29,33,128]. This *Big Data and Cognitive Computing* article aims to open up a new entrepreneurial research avenue by exploring the robust Trustworthy AI European regulatory framework as well as incorporating a proactive entrepreneurial approach for socio-technical initiatives taking place in North America.

Against this backdrop, the rise of decentralized Web3 ecosystems presents unique challenges to the detection of AI-generated content and the establishment of trust in such environments while fostering social innovation [149] as was the case with the previous buzz around smart cities [150–153]. Unlike traditional centralized systems, where oversight and governance are clearly defined, decentralized systems rely on peer-to-peer networks, leaving the authenticity and trustworthiness of information to be validated by the users themselves. As GenAI continues to evolve, its capacity to produce convincing yet fabricated content makes it increasingly difficult to detect disinformation, posing risks to democratic integrity, particularly spread from highly concentrated groups of people giving rise to the relevance of AI urbanism in post-smart cities momentum [154–158].

Detection of AI-generated content is crucial for preserving trust in decentralized Web3 ecosystems. As Eubanks notes [159], the automation of high-tech tools, including AI, has historically been employed to profile, police, and punish marginalized groups. This power dynamic becomes even more problematic when applied to decentralized networks, where there is no central authority to govern the flow of information. Without reliable detection tools, AI-generated disinformation can quickly undermine the credibility of decentralized platforms, exacerbating social inequalities and eroding trust in the system [160–164].

Web3 ecosystems rely on distributed nodes and smart contracts, which complicates the development of reliable detection frameworks [145]. However, detecting AI-generated disinformation in a decentralized environment remains an unresolved issue, requiring innovative approaches that balance privacy, security, and verification [13,66,90,146].

Trust is the backbone of any democratic process, and it becomes even more critical in decentralized ecosystems where traditional forms of oversight are absent. Gohdes [163], in *Repression in the Digital Age*, highlights the ways in which states have historically employed digital tools for surveillance and censorship, which are increasingly integrated into decentralized systems [160]. The diffusion of power in decentralized networks makes it easier for bad actors to spread disinformation without accountability. This poses a significant threat to public trust, as users struggle to discern authentic content from AI-generated misinformation. As the HAI notes [13,147], trust in AI systems is contingent on their transparency and explainability, both of which are challenging to implement in decentralized networks. The absence of centralized control complicates efforts to establish verification protocols, making it essential to develop new methods for detecting and authenticating content in decentralized Web3 ecosystems.

Here there is the list of seven techniques of trust through Decentralized Web3 ecosystems studied in light of the ENFIELD EU project [143]:

Table 3. Seven Detection Techniques of Trust through Decentralized Web3 Ecosystems.

Techniques	Description
T1. Federated Learning for Decentralized AI Detection	Collaborative AI model training across decentralized platforms, preserving privacy without sharing raw data.
T2. Blockchain-Based Provenance Tracking	Blockchain technology records content creation and dissemination, enabling transparent tracking of content authenticity.
T3. Zero-Knowledge Proofs for Content Authentication	Cryptographic method to verify content authenticity without revealing underlying private data.
T4. Decentralized Autonomous Organizations (DAOs) for Crowdsourced Verification	Crowdsourced content verification through DAOs, allowing communities to collectively vote and verify content authenticity.
T5. AI-Powered Digital Watermarking	Embedding unique identifiers into AI-generated content to trace and authenticate its origin.

T6. Explainable AI (XAI) for Content Detection	Provides transparency in AI model decision-making [164], explaining why content was flagged as AI-generated.
T7. Privacy-Preserving Machine Learning (PPML) for Secure Content Verification	Enables secure detection and verification of content while preserving user privacy, leveraging homomorphic encryption and other techniques.

Each technique aligns with the ENFIELD project's goals of fostering transparency, accountability, and privacy in AI detection across urban decentralized systems, helping bolster public trust [143].

The seven detection techniques presented in this article are not mutually exclusive; rather, they represent a cohesive and complementary framework for fostering trust in decentralized Web3 ecosystems. Each technique addresses a unique aspect of trustworthiness—ranging from privacy preservation to transparency, traceability, and participatory governance—and their integration amplifies their collective effectiveness. For example, T1 can be enhanced with T7 techniques to ensure secure and decentralized model training. Similarly, T2 can work in tandem with T3 to validate content authenticity while maintaining user privacy. T5 benefits from T2 to ensure traceability, while T6 provides transparency for T4. These synergies exemplify the ENFIELD Horizon Europe project's focus on leveraging interdisciplinary approaches to operationalize the principles of the AI Act and Draghi Report. By combining these techniques, decentralized AI governance can address the multifaceted challenges of misinformation, disinformation, and democratic erosion, delivering scalable and ethically aligned solutions to safeguard public trust.

3.1. Federated Learning for Decentralized AI Detection (T1)

Federated learning represents a transformative methodology for decentralized AI detection, aligning with the AI Act's focus on safeguarding user privacy while promoting innovation [4,80]. By enabling multiple decentralized nodes to collaboratively train AI models without sharing raw data, federated learning ensures that sensitive information remains local, addressing privacy concerns emphasized in the Draghi Report [3] and supporting the privacy-preserving goals of the ENFIELD Horizon Europe project [143]. This technique addresses the operational challenge of balancing decentralized governance with global model accuracy. For instance, as Burton et al. [165] emphasize, collective intelligence frameworks benefit from federated learning's ability to refine detection capabilities without the need for centralized control.

A practical European example of federated learning can be seen in the *GAIA-X* (www.gaia-x.eu) initiative, which promotes secure and decentralized data ecosystems for industries across Europe. GAIA-X leverages federated approaches to enable cross-border data sharing while maintaining strict data protection standards, aligning with the EU's General Data Protection Regulation (GDPR) and the AI Act's principles. By pooling decentralized resources, federated learning enhances disinformation detection while fostering autonomy within Web3 ecosystems. This scalability enables trust-building across decentralized networks, ensuring compliance with the EU's emphasis on transparency and user-centric AI.

3.2. Blockchain-Based Provenance Tracking (T2)

Blockchain provides a transparent, immutable ledger that enables robust provenance tracking for AI-generated content, aligning with the AI Act's requirement for traceability in high-risk AI applications and the Draghi Report's emphasis on transparency [3,4,80]. By recording every instance of content creation, modification, and dissemination, blockchain ensures the authenticity and accountability of digital information. This approach directly addresses the ENFIELD Horizon Europe project's objective of fostering public trust in decentralized ecosystems. As Lalka [166] and Li [167] note, blockchain's application in tracking content provenance is pivotal in combating misinformation. By embedding digital signatures or hash functions, blockchain provides a verifiable trail of content origin, ensuring stakeholders can distinguish authentic from manipulated materials, which is critical for maintaining trust in decentralized AI governance.

A practical European example is the *OriginTrail* project (www.originaltrail.io), which employs blockchain technology to ensure the traceability of data and products in supply chains across Europe. OriginTrail's decentralized knowledge graph leverages blockchain to authenticate the provenance of goods, ranging from food to pharmaceuticals, ensuring compliance with EU regulations.

3.3. Zero-Knowledge Proofs (ZKPs) for Content Authentication (T3)

ZKPs exemplify the EU's dual commitment to innovation and data protection as outlined in the AI Act and Draghi Report [3,4,80]. ZKPs enable the verification of AI-generated content's authenticity without disclosing sensitive details, ensuring compliance with the privacy-first approach championed by the ENFIELD Horizon Europe project [143]. This technique is particularly relevant for decentralized ecosystems, where users demand confidentiality and transparency. As Medrado and Verdegem argue [168], cryptographic tools like ZKPs are vital for addressing the ethical challenges of decentralized governance. By allowing platforms to confirm content authenticity while protecting proprietary information, ZKPs provide a scalable solution that fosters trust and aligns with the EU's focus on inclusive and secure AI systems.

A European example of ZKP application can be found in the ***European Blockchain Services Infrastructure (EBSI)*** **(**https://digital-strategy.ec.europa.eu/en/policies/european-blockchain-services-infrastructure**)**, an initiative led by the European Commission and the European Blockchain Partnership (EBP). EBSI integrates ZKPs to enhance data security and privacy across multiple use cases, including verifying digital credentials for education and cross-border administrative processes. By enabling institutions to confirm the authenticity of diplomas or professional certifications without exposing personal data, EBSI demonstrates how ZKPs can address privacy concerns while ensuring trust in decentralized systems.

3.4. DAOs for Crowdsourced Verification (T4)

DAOs democratize the verification of AI-generated content, reflecting the Draghi Report's call for participatory governance and the AI Act's emphasis on inclusivity [3,4,80]. By integrating peer review mechanisms, voting systems, and reputation scores, DAOs empower communities to collectively assess content authenticity, fostering trust in decentralized networks. This community-driven approach resonates with the ENFIELD Horizon Europe project's objective to embed trust within local ecosystems [143]. As Mejias & Couldry [169] highlight, DAOs counteract the concentration of power in digital platforms by decentralizing decision-making. This framework democratizes AI governance, creating a collaborative and transparent system for content verification that directly aligns with EU regulatory goals.

A European example of DAOs in practice is ***Aragon*** (https://www.aragon.org/), an open-source platform that provides tools for creating and managing decentralized organizations. Founded in Spain and widely adopted across Europe, Aragon enables communities to set up DAOs for collaborative decision-making and governance. For instance, it has been used in environmental projects where stakeholders collectively verify the authenticity of sustainability claims and vote on funding allocations.

3.5. AI-Powered Digital Watermarking (T5)

AI-powered digital watermarking embeds unique identifiers into AI-generated content, ensuring traceability throughout its lifecycle. This technique directly supports the AI Act's transparency obligations and the Draghi Report's emphasis on accountability in high-risk applications [3,4,80]. By providing a digital fingerprint, watermarking enables real-time detection and verification of content authenticity.

This approach advances the ENFIELD Horizon Europe project's goals by ensuring that all AI-generated materials within decentralized systems remain identifiable and verifiable. As Murgia notes

[170], digital watermarking enhances the ethical deployment of AI by making alterations traceable, thus addressing concerns over content manipulation in decentralized networks.

A European example is the ***C2PA (Coalition for Content Provenance and Authenticity)*** initiative (https://c2pa.org/), which includes European stakeholders and collaborates on open standards for embedding metadata and watermarks in digital media. For instance, Adobe, a key participant in C2PA, has partnered with European media organizations to pilot digital watermarking solutions that help verify the origin and integrity of visual content. These efforts align with the EU's regulatory focus on combating misinformation and ensuring content authenticity, particularly in journalism and digital communications.

3.6. Explainable AI (XAI) for Content Detection (T6)

XAI enhances transparency by clarifying AI decision-making processes, a core principle of the AI Act and Draghi Report [3,4,80]. By providing insights into why specific content was flagged as AI-generated or misinformative, XAI fosters public trust in decentralized systems.

This technique aligns with the ENFIELD Horizon Europe project's focus on explainability as a cornerstone of ethical AI. As Johnson & Acemoglu argue [171], transparent AI systems are essential for sustaining public trust and democratic resilience. XAI bridges the gap between technical robustness and societal understanding, ensuring accountability and adherence to EU principles in decentralized AI ecosystems.

A European example is the ***Horizon 2020 Trust-AI project*** (www.trustai.eu), which focuses on developing explainable and trustworthy AI models across various sectors, including healthcare, finance, and public administration. For instance, in the healthcare domain, Trust-AI collaborates with European institutions to implement XAI systems that explain diagnostic decisions made by AI-powered tools, enabling medical professionals to validate and trust the outputs. This work aligns with EU principles by ensuring that AI systems remain transparent, interpretable, and accountable.

3.7. Privacy-Preserving Machine Learning (PPML) for Secure Content Verification (T7)

PPML enables AI models to verify content authenticity without compromising user privacy, reflecting the AI Act's data protection requirements and the Draghi Report's focus on equitable innovation [3,4,80]]. Techniques such as homomorphic encryption and secure multi-party computation allow sensitive data to remain secure while enabling robust analysis.

PPML supports the ENFIELD Horizon Europe project's vision of decentralized and privacy-focused AI systems. As Rella et al. emphasize [172], integrating PPML into federated learning ensures that detection processes are both secure and ethical. This approach fosters user trust and addresses operational challenges in decentralized ecosystems, aligning with EU mandates for trustworthy and inclusive AI.

A European example of PPML in practice is the ***MUSKETEER project*** **(**www.musketeer.eu**)**, funded under the EU Horizon 2020 program, which focuses on developing privacy-preserving machine learning frameworks for industrial and societal applications. MUSKETEER integrates homomorphic encryption and secure multi-party computation to enable collaborative model training across organizations without exposing sensitive data. For instance, it has been piloted in the healthcare sector to allow hospitals across Europe to train AI models on patient data while complying with GDPR and safeguarding privacy.

The seven techniques collectively operationalize the EU's regulatory principles as outlined in the AI Act and Draghi Report, bridging policy frameworks with actionable methodologies [3,4,80]. They align with the Enfield Horizon Europe project's mission to advance decentralized governance, privacy, and public trust in AI systems [143]. By integrating these techniques, decentralized Web3 ecosystems can ensure transparency, accountability, and resilience against AI-driven challenges while adhering to the EU's commitment to fostering ethical and innovative AI environments.

The seven European examples underscore that Trustworthy AI is designed not just for governments and regulatory bodies but for a diverse set of stakeholders. This inclusivity is central to the EU's approach, as reflected in the AI Act and Draghi Report. The examples reveal that Trustworthy AI benefits multiple audiences as we can observe in Table 4 that respond to the Research Question of this article: *Trustworthy AI for Whom?*

Table 4. *Trustworthy AI for Whom?* Responding to the Research Question per Each of the Seven Detection Techniques of Trust.

Technique	European Example	Response to the Research Question	*Trustworthy AI for Whom?*
T1. Federated Learning for Decentralized AI Detection	*GAIA-X* initiative promoting secure and decentralized data ecosystems www.gaia-x.eu	Supports user-centric data sharing and privacy compliance across Europe	**End Users and Citizens**: Projects like ***GAIA-X*** (federated learning) focus on user-centric designs that prioritize transparency and data privacy.
T2. Blockchain-Based Provenance Tracking	***OriginTrail*** project ensuring data and product traceability www.originaltrail.io	Enhances product authenticity and trust in supply chains for consumers and industries	**Communities and Organizations**: Tools like ***OriginTrail*** (blockchain-based provenance tracking) ensures that organizations and consumers can trust the authenticity of data and products.
T3. Zero-Knowledge Proofs for Content Authentication	***European Blockchain Services Infrastructure (EBSI)*** for credential verification https://digital-strategy.ec.europa.eu/en/policies/european-blockchain-services-infrastructure	Ensures privacy and security for credential verification in education and public services	**Regulators and Policymakers**: By embedding EU principles into operational frameworks, initiatives like the ***European Blockchain Services Infrastructure (EBSI)*** demonstrate that Trustworthy AI aids regulators in enforcing compliance while maintaining transparency and inclusivity across borders.
T4. DAOs for Crowdsourced Verification	*Aragon* platform enabling collaborative decentralized governance https://www.aragon.org/	Empowers communities with participatory governance and collaborative decision-making	**Communities and Organizations**: Tools like *Aragon* (DAOs) empower decentralized decision-making, fostering collaborative governance among community members.
T5. AI-Powered Digital Watermarking	*C2PA* initiative embedding metadata and watermarks in digital media https://c2pa.org/	Improves traceability and content authenticity for media and journalism	**Industry and Innovation Ecosystems**: Projects like *C2PA* (digital watermarking) support industrial and media ecosystems by providing robust frameworks. These initiatives promote innovation while adhering to ethical guidelines.

T6. Explainable AI (XAI) for Content Detection	***Horizon 2020 Trust-AI*** project developing explainable AI models www.trustai.eu	Enhances transparency and trust in AI decision-making for users and professionals	**End Users and Citizens**: Projects like ***Trust-AI*** (XAI) focus on user-centric designs that prioritize transparency and data privacy. Citizens gain trust in AI systems when these systems explain their decisions, safeguard personal data, and remain accountable.
T7. Privacy-Preserving Machine Learning (PPML) for Secure Content Verification	***MUSKETEER*** project creating privacy-preserving machine learning frameworks www.musketeer.eu	Ensures secure AI training and compliance with privacy laws for industry stakeholders	**Industry and Innovation Ecosystems**: Projects like ***MUSKETEER*** (PPML) support industrial ecosystems by providing robust frameworks for privacy-preserving analysis and content authentication. These initiatives promote innovation while adhering to ethical guidelines.

4. Discussion and Conclusion

4.1. Discussions and Conclusions

The emergence of GenAI and its integration into decentralized Web3 ecosystems presents both transformative opportunities and profound challenges. This article has explored the tension between fostering innovation and ensuring democratic accountability through the lens of Trustworthy AI. Framed by the research question, *"Trustworthy AI for whom?"*, this inquiry has been situated within the context of the AI Act, the Draghi Report, and the ENFIELD Horizon Europe project. It argues that trust in AI systems must transcend compliance frameworks and technical excellence. Instead, it must prioritize inclusivity, societal equity, and participatory governance.

The seven detection techniques of trust—federated learning, blockchain-based provenance tracking, Zero-Knowledge Proofs (ZKPs), DAOs, AI-powered digital watermarking, Explainable AI (XAI), and Privacy-Preserving Machine Learning (PPML)—demonstrate the potential of decentralized mechanisms to enhance transparency, accountability, and public trust. These methodologies align closely with the regulatory aspirations of the AI Act and the strategic imperatives outlined in the Draghi Report, offering actionable pathways to operationalize trust in AI ecosystems.

Critically, these detection methodologies address a central challenge identified in both policy frameworks: balancing innovation with ethical and societal responsibilities. Tools such as DAOs and federated learning emphasize the importance of participatory governance, challenging the issue of "technological paternalism," as discussed by Merchant [173], where the beneficiaries of AI are often determined without sufficient input from marginalized groups. Integrating end-user perspectives into the development of decentralized Web3 tools could foster greater public trust, ensuring that these systems genuinely serve the communities they aim to empower.

The examples presented in this study highlight the broad applicability of Trustworthy AI to diverse stakeholders. *End Users and Citizens* benefit from initiatives like *GAIA-X* (federated learning) and *Trust-AI* (XAI), which prioritize transparency and privacy, empowering individuals to understand and trust AI systems. *Communities and Organizations* gain from decentralized governance mechanisms such as *Aragon* (DAOs) and *OriginTrail* (blockchain-based provenance tracking), fostering collaborative decision-making and trust in data authenticity. *Industry and Innovation Ecosystems* are supported by tools like *C2PA* (digital watermarking) and *MUSKETEER* (PPML), which provide robust frameworks for privacy-preserving analysis and content authentication while adhering to ethical standards. Finally, *Regulators and Policymakers* are aided by frameworks such as *EBSI*, which ensure privacy and compliance while maintaining transparency and inclusivity.

Equally significant is the need to shift from theoretical frameworks to practical implementation. As Sieber et al. emphasize [174], the success of AI governance relies on the public actively engaging with these technologies. Enhancing user experience (UX) is key to this engagement. For instance, sophisticated but intuitive tools that communicate the functionality of blockchain-based provenance tracking or AI-powered watermarking could bridge the gap between technical innovation and societal adoption. Similarly, improving the explainability of AI decision-making through XAI could demystify complex processes, fostering trust among diverse stakeholder groups.

Ultimately, the success of decentralized Web3 ecosystems depends on how effectively these tools are adapted to regional, cultural, and societal contexts. As Tunç observes [175], the future trajectory of AI governance will be shaped by its capacity to reconcile universal principles with localized needs. By fostering multistakeholder collaboration, the ENFIELD Horizon Europe project provides a valuable framework for integrating decentralized governance tools with public values, ensuring that AI remains a democratic enabler rather than a disruptor.

In conclusion, Trustworthy AI, as conceptualized and operationalized in this article, serves as a framework for inclusivity, equity, and transparency. The seven detection techniques outlined in this research demonstrate how AI systems can align with societal values while addressing the complexities of decentralized environments. By combining the regulatory guidance of the AI Act and Draghi Report with innovative, practical tools, this article outlines a pathway to ensure that AI becomes a tool for societal empowerment rather than disruption. Trustworthy AI, ultimately, is AI for everyone—serving diverse stakeholders and reinforcing the democratic principles that underpin its development.

4.2. Limitations

Despite its contributions, this study acknowledges several limitations in the development and deployment of trustworthy AI in decentralized Web3 ecosystems.

(i) Technical and Operational Challenges: Many of the techniques discussed, such as federated learning and PPML, require advanced computational infrastructure (Quantum Computing) and significant technical expertise. Their deployment in resource-constrained environments may be limited, perpetuating global inequalities in digital access and trust frameworks.

(ii) Ethical and Governance Gaps: While tools like DAOs and blockchain foster transparency and decentralization, they raise ethical concerns regarding power concentration among technologically savvy elites [128]. As recently noted by Calzada [128] and supported by AI hype approach by Floridi [176], decentralization does not inherently equate to democratization; instead, it risks replicating hierarchical structures in digital contexts.

(iii) Regulatory Alignment and Enforcement: The AI Act and the Draghi Report provide robust policy frameworks, but their enforcement mechanisms remain uneven across EU member states. This regulatory fragmentation may hinder the uniform implementation of the detection techniques proposed.

(iv) Public Awareness and Engagement: A significant barrier to adoption lies in the public's limited understanding of decentralized technologies. As Medrado and Verdegem highlight [168], there is a need for more inclusive educational initiatives to bridge the knowledge gap and promote trust in AI governance systems.

(v) Emergent Risks of AI: GenAI evolves rapidly, outpacing regulatory and technological safeguards. This dynamism introduces uncertainties about the long-term effectiveness of the proposed detection techniques.

4.3. Future Research Avenues

To address these limitations and advance the discourse on trustworthy AI, future research should explore the following avenues:

(i) Context-Specific Adaptations: Further research is needed to tailor decentralized Web3 tools to diverse regional and cultural contexts. This involves integrating local governance norms and socio-political dynamics into the design and implementation of detection frameworks.

(ii) Inclusive Governance Models: Building on the principles of participatory governance discussed by Mejias and Couldry [169], future studies should examine how multistakeholder frameworks can be institutionalized within decentralized ecosystems. Citizen assemblies, living labs, and co-design workshops offer promising methods for inclusive decision-making.

(iii) User-Centric Design: Enhancing UX for detection tools such as digital watermarking and blockchain provenance tracking is crucial. Future research should focus on creating user-friendly interfaces that simplify complex functionalities, fostering greater public engagement and trust.

(iv) Ethical and Legal Frameworks: Addressing the ethical and legal challenges posed by decentralized systems requires interdisciplinary collaboration. Scholars in law, ethics, and social sciences should work alongside technologists to develop governance models that balance innovation with accountability.

(v) AI Literacy Initiatives: Expanding on Sieber et al. [174], there is a need for targeted educational programs to improve public understanding of AI technologies. These initiatives could focus on empowering marginalized communities, ensuring equitable access to the benefits of AI.

(vi) Monitoring and Evaluation Mechanisms: Future studies should investigate robust metrics for assessing the efficacy of detection techniques in real-world scenarios. This includes longitudinal studies to monitor their impact on trust, transparency, and accountability in decentralized systems.

(vii) Emergent Technologies and Risks: Finally, research should anticipate the future trajectories of AI and Web3 ecosystems, exploring how emerging technologies such as quantum computing or advanced neural networks may impact trust frameworks.

(viii) Learning from Urban AI: A potentially prominent field is emerging around the concept of Urban AI, which warrants further exploration. The question *"Trustworthy AI for whom?"* echoes the earlier query *"Smart City for whom?"*, suggesting parallels between the challenges of integrating AI into urban environments and the broader quest for trustworthy AI [177–182]. Investigating the evolution of Urban AI as a distinct domain could provide valuable insights into the socio-technical dynamics of trust, governance, and inclusivity within AI-driven urban systems [183–185].

This article underscores the critical importance of trustworthy AI in navigating the complexities of GenAI and decentralized Web3 ecosystems [186]. By aligning with the AI Act, Draghi Report, and the objectives of the ENFIELD Horizon Europe project, the proposed detection techniques provide actionable pathways to strengthen democratic resilience and societal trust. However, achieving this vision requires a continued commitment to multistakeholder collaboration, inclusive governance, and user-centric innovation. As the field evolves, integrating diverse perspectives and addressing emerging challenges will be pivotal in ensuring that AI serves as a force for equitable and sustainable societal transformation [187–196].

Author Contributions: Conceptualization, I.C.; methodology, I.C.; validation, I.C. and G.N.; formal analysis, I.C.; investigation, I.C.; resources, I.C. and G.N.; data curation, I.C.; writing—original draft preparation, I.C.; writing—review and editing, I.C., and G.N.; visualization, I.C.; supervision, I.C.; project administration, I.C. and M.A-R; funding acquisition, I.C. and G.N.

Funding: This research was funded by (i) European Commission, Horizon 2020, H2020-MSCA-COFUND-2020-101034228-WOLFRAM2: Ikerbasque Start Up Fund, 3021.23.EMAJ; (ii) UPV-EHU, Research Groups, IT 1441-22 and IT 1541-22; (iii) Ayuda en Acción NGO, Innovation & Impact Unit, Research Contract: Scientific Direction and Strategic Advisory, *Social Innovation Platforms in the Age of Artificial Intelligence (AI)* (www.socialsolver.org accessed on 1 July 2024) and *AI for Social Innovation. Beyond the Noise of Algorithms and Datafication* Summer School Scientific Direction, 2-3 September 2024, Donostia-St. Sebastian, Spain (https://www.uik.eus/en/activity/artificial-intelligence-social-innovation-ai4si accessed on 1 July 2024), PT10863;

(iv) Presidency of the Basque Government, External Affairs General Secretary, Basque Communities Abroad Direction, Scientific Direction and Strategic Advisory *e-Diaspora Platform HanHemen* (www.hanhemen.eus/en accessed on 1 July 2024), PT10859; (v) European Commission, Horizon Europe, ENFIELD-European Lighthouse to Manifest Trustworthy and Green AI, HORIZON-CL4-2022-HUMAN-02-02-101120657; SGA oc1-2024-TES-01-01, https://www.enfield-project.eu/about. Invited Professor at BME, Budapest University of Technology and Economics (Hungary) (https://www.tmit.bme.hu/speechlab?language=en); (vi) Gipuzkoa Province Council, Etorkizuna Eraikiz 2024: *AI's Social Impact in the Historical Province of Gipuzkoa* (AI4SI). 2024-LAB2-007-01. www.etorkizunaeraikiz.eus/en/ and https://www.uik.eus/eu/jarduera/adimen-artifiziala-gizarte-berrikuntzarako-ai4si; (vii) Warsaw School of Economics SGH (Poland) by RID LEAD, Regional Excellence Initiative Programme (https://rid.sgh.waw.pl/en/grants-0 and https://www.sgh.waw.pl/knop/en/conferences-and-seminars-organized-by-the-institute-of-enterprise_and https://www.sgh.waw.pl/knop/en/conferences-and-seminars-organized-by-the-institute-of-enterprise; (viii) SOAM Residence Programme: Network Sovereignties (Germany) via BlockchainGov (www.soam.earth); (ix) Decentralization Research Centre (Canada) (www.thedrcenter.org/fellows-and-team/igor-calzada/); (x) The Learned Society of Wales (LSW) 524205; (xi) Fulbright Scholar-In-Residence (S-I-R) Award 2022-23, PS00334379 by the US–UK Fulbright Commission and IIE, US Department of State at the California State University; (xii) the Economic and Social Research Council (ESRC) ES/S012435/1 "WISERD Civil Society: Changing Perspectives on Civic Stratification/Repair"; (xiii) Gipuzkoa Province Council, Human Rights & Democratic Culture: *Gipuzkoa Algorithmic Territory: Socially Cohesive Digitally Sustainable?* And *Digital Inclusion, Data Divide, and Migrants* Summer School Scientific Direction, September 2025, Donostia-St. Sebastian, Spain, PT10937; and (xiv) Astera Institute, Cosmik Data Cooperatives for Open Science. *Views and opinions expressed however those of the author only and do not necessarily reflect those of these institutions. None of them can be held responsible for them.*

Data Availability Statement: No data was used for the research described in the article.

Acknowledgments: Due to my fieldwork action research under (i) the framework of Horizon Europe ENFIELD research project in Budapest (https://www.igorcalzada.com/invited-keynote-speaker-in-budapest-hungary-on-13th-november-2024-at-wire2024-week-of-innovative-regions-in-europe/) and (ii) RIA LEAD research programme in Warsaw: (i) particularly from 9 to 19 November 2024 and from 7 to 16 February 2025 and (ii) from 27 November to 5 December 2024 (https://www.igorcalzada.com/invited-keynote-speaker-in-warsa-poland-on-28-and-29th-november-2024-at-warsaw-school-of-economics-sgh/), respectively, I hereby acknowledge my gratitude to Budapest University of Economics and Technology BME (particularly to Dr Geza Nemeth and Dr Mohammed Al-Radhi).and Warsaw School of Economics SGH (particularly to Dr Anna Visvizi). During this fieldwork research, the corresponding author coordinated several research activities in Budapest and Warsaw, including (i) WIRE2024 Conference participation (https://wire2024.nkfih.gov.hu/program/13-14-november-2024), Budapest fieldwork, BME seminar and ENFIELD Workshop and (ii) RID-LEAD Public Conference (https://www.sgh.waw.pl/en/events/rid-seminar-data-markets-society), Smart Region Conference (https://www.sgh.waw.pl/knop/en/conferences-and-seminars-organized-by-the-institute-of-enterprise), and Lectureships in Public Policies and AI.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Alwaisi, S., Salah Al-Radhi, M. & Németh, G., (2023) Automated child voice generation: Methodology and implementation. *2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Bucharest, Romania, pp. 48-53. doi: 10.1109/SpeD59241.2023.10314889.
2. Alwaisi, S. & Németh, G., (2024) Advancements in expressive speech synthesis: A review. *Infocommunications Journal*, 16(1), pp. 35-49. doi: 10.36244/ICJ.2024.1.5.
3. European Commission. The Future of European Competitiveness: A Competitiveness Strategy for Europe. *European Commission*, September 2024. Available online: https://ec.europa.eu (accessed on 18 November 2024).
4. European Parliament and Council. Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations and Directives. *Official Journal of the European*

Union. 2024, L 1689, 1–144. Available online: http://data.europa.eu/eli/reg/2024/1689/oj (accessed on 18 November 2024).

5. Yang, F., Goldenfein, J., & Nickels, K. (2024). GenAI Concepts. Melbourne: ARC Centre of Excellence for Automated Decision-Making and Society RMIT University, and OVIC. DOI: 10.60836/psmc-rv23
6. Insight & Foresight (2024). How Generative AI Will Transform Strategic Foresight.
7. Amoore, L.; Campolo, A.; Jacobsen, B.; Rella, L. A world model: On the political logics of generative AI. *Political Geography* 2024, *113*, 103134. Available online: https://doi.org/10.1016/j.polgeo.2024.103134 (accessed on 18 November 2024).
8. Chafetz, H., Saxena, S., & Verhulst, S.G., 2024. *A Fourth Wave of Open Data? Exploring the Spectrum of Scenarios for Open Data and Generative AI*. The GovLab. Available from: https://arxiv.org/abs/2405.04333 (Accessed 1 Sept 2024).
9. Delacroix, S. (2024) Sustainable data rivers? Rebalancing the data ecosystem that underlies generative AI. *Critical AI*, 2(1), No Pagination Specified. https://doi.org/10.1215/2834703X-11205224.
10. Gabriel, I. et al. (2024) The ethics of advanced AI assistants. *arXiv preprint*. https://arxiv.org/abs/2404.16244..
11. Shin, D., Koerber, A., & Lim, J.S. (2024) Impact of misinformation from generative AI on user information processing: How people understand misinformation from generative AI. *New Media & Society*. https://doi.org/10.1177/14614448241234040.
12. Tsai, L. L., Pentland, A., Braley, A., Chen, N., Enríquez, J. R., & Reuel, A. (2024) *An MIT Exploration of Generative AI: From Novel Chemicals to Opera*. MIT Governance Lab. Available from: https://doi.org/10.21428/e4baedd9.5aaf489a (Accessed 1 September 2024).
13. Weidinger, L., et al. (2023) Sociotechnical Safety Evaluation of Generative AI Systems. *arXiv preprint*. https://arxiv.org/abs/2310.11986.
14. Allen, D. & Weyl, E.G. (2024) The Real Dangers of Generative AI. *Journal of Democracy*. 35(1): 147-162.
15. Kitchin, R. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences; Sage: London, UK, 2014.
16. Cugurullo, F.; Caprotti, F.; Cook, M.; Karvonen, A.; McGuirk, P.; Marvin, S., Eds. *Artificial Intelligence and the City: Urbanistic Perspectives on AI*; Routledge: Abingdon, UK, 2024. DOI: 10.4324/9781003365877.
17. Farina, M., Yu, X. & Lavazza, A. (2023). Ethical considerations and policy interventions concerning the impact of generative AI tools in the economy and in society. *AI and Ethics.* https://doi.org/10.1007/s43681-023-00405-2.
18. **Calzada, I.** (2021), *Smart City Citizenship*, Cambridge, Massachusetts: Elsevier Science Publishing Co Inc. [ISBN (Paperback): 978-0-12-815300-0]. doi:10.1016/c2017-0-02973-7.
19. Aguerre, C., Campbell-Verduyn, M. & Scholte, J.A. (2024) *Global Digital Data Governance: Polycentric Perspectives*. Abingdon, UK: Routledge.
20. Angelidou, M.; Sofianos, S. *The Future of AI in Optimizing Urban Planning: An In-Depth Overview of Emerging Fields of Application*. International Conference on Changing Cities VI: Spatial, Design, Landscape, Heritage & Socio-economic Dimensions. Rhodes Island, Greece, 24-28 June 2024.
21. Polanyi, K. (1944) The Great Transformation: The Political and Economic Origins of Our Time. New York: Farrar & Rinehart.
22. Solaiman, I., et al. (2019) Release Strategies and the Social Impacts of Language Models. *arXiv preprint*. https://arxiv.org/abs/1908.09203.
23. **Calzada, I.** (2024), Artificial Intelligence for Social Innovation: Beyond the Noise of Algorithms and Datafication. *Sustainability*, **16**(19), 8638. DOI: 10.3390/su16198638.
24. Fang, R., et al. (2024). LLM Agents can Autonomously Hack Websites. *arXiv preprint*. https://arxiv.org/abs/2402.06664.
25. Farina, M., Lavazza, A., Sartori, G. & Pedrycz, W. (2024). Machine learning in human creativity: status and perspectives. *AI & Society.* https://doi.org/10.1007/s00146-023-01836-5.
26. Abdi, I.I. Digital Capital and the Territorialization of Virtual Communities: An Analysis of Web3 Governance and Network Sovereignty. 2024.
27. Calzada, I. (2024) (Libertarian) Decentralized Web3 Map: In Search of a Post-Westphalian Territory. *SSRN.* DOI: 10.2139/ssrn.4937294.

28. Calzada, I. (2024) Decentralized Web3 Reshaping Internet Governance: Towards the Emergence of New Forms of Nation-Statehood? *Future Internet*, 16(10), 361. DOI: 10.3390/fi16100361.
29. Calzada, I. (2024) From data-opolies to decentralization? The AI disruption amid the Web3 Promiseland at stake in datafied democracies, in Visvizi, A., Corvello, V. and Troisi, O. (eds.) *Research and Innovation Forum*. Cham, Switzerland: Springer.
30. Calzada, I. (2024) Democratic erosion of data-opolies: Decentralized Web3 technological paradigm shift amidst AI disruption. *Big Data and Cognitive Computing*, 8(3), p. 26. doi:10.3390/bdcc8030026.
31. Calzada, I. (2023) Disruptive technologies for e-diasporas: Blockchain, DAOs, data cooperatives, metaverse, and ChatGPT, *Futures*, 154(C), p. 103258. doi:10.1016/j.futures.2023.103258.
32. Calzada, I. (2020) Democratising smart cities? Penta-helix multistakeholder social innovation framework. *Smart Cities*, 3, pp. 1145-1172.
33. Allen, D., Frankel, E., Lim, W., Siddarth, D., Simons, J. & Weyl, E.G. (2023) Ethics of Decentralized Social Technologies: Lessons from Web3, the Fediverse, and Beyond, *Harvard University Edmond & Lily Safra Center for Ethics*. Available from: https://myaidrivecom/view/file-A5rvW7aJ8emgJMG8wKH3WDTz (Accessed 1 September 2024).
34. De Filippi, P.; Cossar, S.; Mannan, M.; Nabben, K.; Merk, T.; Kamalova, J. Report on Blockchain Governance Dynamics. Project Liberty Institute and BlockchainGov, May 2024. Available online: https://www.projectliberty.io/institute (accessed on 20 November 2024).
35. Daraghmi, E.; Hamoudi, A.;Abu Helou, M. Decentralizing Democracy: Secure and Transparent E-Voting Systems with Blockchain Technology in the Context of Palestine. Future Internet 2024, 16, 388. https://doi.org/10.3390/fi16110388
36. Liu, X.; Xu, R.; Chen, Y. A. Decentralized Digital Watermarking Framework for Secure and Auditable Video Data in Smart Vehicular Networks. Future Internet 2024, 16, 390. https://doi.org/10.3390/fi16110390
37. Stefano Moroni, Revisiting subsidiarity: Not only administrative decentralization but also multidimensional polycentrism, Cities, Volume 155, 2024, 105463, https://doi.org/10.1016/j.cities.2024.105463.
38. Van Kerckhoven, S. & Chohan, U.W. (2024) Decentralized Autonomous Organizations: Innovation and Vulnerability in the Digital Economy. Oxon, UK: Routledge.
39. Singh, A., Lu, C., Gupta, G., Chopra, A., Blanc, J., Klinghoffer, T., Tiwary, K., & Raskar, R. (2024). A perspective on decentralizing AI. *MIT Media Lab*.
40. Mathew, A.J. The myth of the decentralised internet. *Internet Policy Review*, 2016, *9* (3). https://policyreview.info/articles/analysis/myth-decentralised-internet
41. Zook, M. (2023) Platforms, blockchains and the challenges of decentralization. *Cambridge Journal of Regions, Economy and Society*, 16(2), pp. 367–372.
42. Kneese, T. & Oduro, S. (2024) AI Governance Needs Sociotechnical Expertise: Why the Humanities and Social Sciences are Critical to Government Efforts. *Data & Society Policy Brief*. 1-10.
43. OECD. Assessing Potential Future Artificial Intelligence Risks, Benefits and Policy Imperatives. OECD Artificial Intelligence Papers. No. 27, November 2024. Available online: https://oecd.ai/site/ai-futures (accessed on 20 November 2024).
44. Nabben, K.; De Filippi, P. Accountability protocols? On-chain dynamics in blockchain governance. *Internet Policy Review* 2024, 13(4). DOI: 10.14763/2024.4.1807.
45. Nanni, R., Bizzaro, P. G., & Napolitano, M. (2024). The false promise of individual digital sovereignty in Europe: Comparing artificial intelligence and data regulations in China and the European Union. *Policy & Internet*, 1–16. https://doi.org/10.1002/poi3.424
46. Schroeder, R. (2024). Content moderation and the digital transformations of gatekeeping. *Policy & Internet*, 1–16. https://doi.org/10.1002/poi3.425
47. Gray, J.E., Hutchinson, J., Stilinovic, M. and Tjahja, N. (2024), The pursuit of 'good' Internet policy. *Policy Internet*, 16: 480-484. https://doi.org/10.1002/poi3.423
48. Pohle, J.; Santaniello, M. From multistakeholderism to digital sovereignty: Toward a new discursive order in internet governance. *Policy & Internet*, 2024. 1–20. https://doi.org/10.1002/poi3.426

49. Viano, C., Avanzo, S., Cerutti, M., Cordero, A., Schifanella, C. & Boella, G. (2022) Blockchain tools for socio-economic interactions in local communities. *Policy and Society*, 41, pp. 373–385. doi:10.1093/polsoc/puac007.
50. Karatzogianni, A., Tiidenberg, K., & Parsanoglou, D. (2022). The impact of technological transformations on the digital generation: Digital citizenship policy analysis (Estonia, Greece, and the UK). DigiGen Policy Brief, April 2022. DOI: 10.5281/zenodo.6457932.
51. Gerlich, Michael. 2024. Societal Perceptions and Acceptance of Virtual Humans: Trust and Ethics across Different Contexts. Social Sciences 13: 516. https://doi.org/10.3390/socsci13100516
52. Waldner, D. & Lust, E. (2018) Unwelcome change: Coming to terms with democratic backsliding. *Annual Review of Political Science*. 21(1): 93-113.
53. Roose, K. (2024) Available from: https://www.nytimes.com/2024/07/19/technology/ai-data-restrictions.html (Accessed on 1 Sept 2024).
54. Kolt, N. (2024) 'Governing AI Agents.' *Available at SSRN*. https://dx.doi.org/10.2139/ssrn.4772956.
55. Calzada, I. (2024c) Data (un)sustainability: Navigating utopian resistance while tracing emancipatory datafication strategies in Certomá, C., Martelozzo, F. and Iapaolo, F. (eds.) *Digital (Un)Sustainabilities: Promises, Contradictions, and Pitfalls of the Digitalization-Sustainability Nexus*. Routledge: Oxon, UK. doi:10.4324/9781003441311-11.
56. Benson, J. (2024) Intelligent Democracy: Answering The New Democratic Scepticism. Oxford, UK: Oxford University Press.
57. Coeckelbergh, M. (2024) Artificial intelligence, the common good, and the democratic deficit in AI governance. *AI Ethics*. doi:10.1007/s43681-024-00492-9.
58. García-Marzá, D. & Calvo, P. (2024) Algorithmic Democracy: A Critical Perspective Based on Deliberative Democracy. Cham, Switzerland: Springer Nature.
59. KT4Democracy. Available at: https://kt4democracy.eu/ (Accessed 1 January 2024).
60. Levi, S. (2024) Digitalización Democrática: Soberanía Digital para las Personas. Barcelona, Spain: Rayo Verde.
61. Poblet, M.; Allen, D. W. E.; Konashevych, O.; Lane, A. M.; Diaz Valdivia, C. A. From Athens to the Blockchain: Oracles for Digital Democracy. *Front. Blockchain* 2020, *3*, 575662. Available online: https://doi.org/10.3389/fbloc.2020.575662 (accessed on 18 November 2024).
62. De Filippi, P., Reijers, W. & Morshed, M. (2024) *Blockchain Governance*. Boston, USA: MIT Press.
63. Visvizi, A.; Malik, R.; Guazzo, G.M.; Çekani, V. The Industry 5.0 (I50) Paradigm, Blockchain-Based Applications and the Smart City. *Eur. J. Innov. Manag.* 2024, *4*. https://doi.org/10.1108/EJIM-09-2023-0826.
64. Roio, D., Selvaggini, R., Bellini, G. & Dintino, A. (2024) SD-BLS: Privacy preserving selective disclosure of verifiable credentials with unlinkable threshold revocation. *2024 IEEE International Conference on Blockchain (Blockchain)*, Copenhagen, Denmark, pp. 505-511. doi: 10.1109/Blockchain62396.2024.00074.
65. Viano, C., Avanzo, S., Boella, G., Schifanella, C. & Giorgino, V. (2023) Civic blockchain: Making blockchains accessible for social collaborative economies. *Journal of Responsible Technology*, 15, 100066. doi:10.1016/j.jrt.2023.100066.
66. Ahmed, S., et al. (2024) Field-building and the epistemic culture of AI safety. *First Monday*.
67. Tan, J. et al. 2024, Open Problems in DAOs. https://arxiv.org/abs/2310.19201v2
68. Petreski, Davor and Cheong, Marc, "Data Cooperatives: A Conceptual Review" (2024). ICIS 2024 Proceedings. 15. https://aisel.aisnet.org/icis2024/lit_review/lit_review/15
69. Stein, J., Fung, M.L., Weyenbergh, G.V. & Soccorso, A. (2023) Data cooperatives: A framework for collective data governance and digital justice', *People-Centered Internet*. Available from: https://myaidrivecom/view/file-ihq4z4zhVBYaytB0mS1k6uxy (Accessed 1 September 2024).
70. Dathatri, S. et al. Scalable watermarking for identifying large model outputs. *Nature* 2024, 634, 818-823. DOI:10.1038/s41586-024-08025-4.
71. Adler, et al. (2024) Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online, *arXiv*. Available from: https://arxiv.org/abs/2408.07892 (Accessed 1 September 2024).

72. Fratini, Samuele and Hine, Emmie and Novelli, Claudio and Roberts, Huw and Floridi, Luciano, Digital Sovereignty: A Descriptive Analysis and a Critical Evaluation of Existing Models (April 21, 2024). Available at SSRN: https://ssrn.com/abstract=4816020 or http://dx.doi.org/10.2139/ssrn.4816020
73. Hui, Yuk. Machine and Sovereignty for a Planetary Thinking. University of Minnesota Press: Minneapolis and London.
74. New America. *From Digital Sovereignty to Digital Agency*. New America Foundation, 2023. Available online: https://www.newamerica.org/planetary-politics/briefs/from-digital-sovereignty-to-digital-agency/ (accessed on 20 November 2024).
75. Glasze, G. et al. Contested Spatialities of Digital Sovereignty. *Geopolitics* 2023, 28(2): 919-958. DOI:10.1080/14650045.2022.2050070.
76. The Conversation (2023) Elon Musk's feud with Brazilian judge is much more than a personal spat – it's about national sovereignty, freedom of speech, and the rule of law. Available from: https://theconversation.com/elon-musks-feud-with-brazilian-judge-is-much-more-than-a-personal-spat-its-about-national-sovereignty-freedom-of-speech-and-the-rule-of-law-238264 (Accessed 20 September 2024).
77. The Conversation (2023) Albanese promises to legislate minimum age for kids' access to social media. Available from: https://theconversation.com/albanese-promises-to-legislate-minimum-age-for-kids-access-to-social-media-238586 (Accessed 20 September 2024).
78. Calzada, I. Data Co-operatives through Data Sovereignty. *Smart Cities* 2021, *4*, 1158–1172. doi: 10.3390/smartcities4030062
79. Belanche, D., Belk, R.W., Casaló, L.V. & Flavián, C. (2024) The dark side of artificial intelligence in services. *Service Industries Journal*, 44, pp. 149–172.
80. European Parliament. (2023). *EU AI Act: First Regulation on Artificial Intelligence*. Available online: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (accessed on 23 November 2024).
81. Yakowitz Bambauer, Jane R. and Zarsky, Tal, Fair-Enough AI (August 08, 2024). Forthcoming in the Yale Journal of Law & Technology, Available at SSRN: https://ssrn.com/abstract=4924588 or http://dx.doi.org/10.2139/ssrn.4924588
82. Dennis, C. et al. (2024). What Should Be Internationalised in AI Governance? *Oxford Martin AI Governance Initiative*.
83. Ghioni, R.; Taddeo, M.; Floridi, L. Open Source Intelligence and AI: A Systematic Review of the GELSI Literature. *SSRN*. Available online: https://ssrn.com/abstract=4272245 (accessed on 18 November 2024).
84. Bullock, S.; Ajmeri, N.; Batty, M.; Black, M.; Cartlidge, J.; Challen, R.; Chen, C.; Chen, J.; Condell, J.; Danon, L.; Dennett, A.; et al. Artificial Intelligence for Collective Intelligence: A National-Scale Research Strategy. 2024. Available online: https://ai4ci.ac.uk (accessed on 20 November 2024).
85. Alon, I.; Haidar, H.; Haidar, A.; Guimón, J. The future of artificial intelligence: Insights from recent Delphi studies. *Futures* 2024, *103514*. https://doi.org/10.1016/j.futures.2024.103514.
86. Ben Dhaou, S., Isagah, T., Distor, C., & Ruas, I.C. (2024). Global Assessment of Responsible Artificial Intelligence in Cities: Research and recommendations to leverage AI for people-centred smart cities. Nairobi, Kenya. United Nations Human Settlements Programme (UN-Habitat).
87. Narayanan, A. (2023) Understanding Social Media Recommendation Algorithms'. *Knight First Amendment Institute*, 1-49.
88. Settle, J.E. (2018) *Frenemies: How Social Media Polarizes America*. Cambridge University Press.
89. European Commission, Joint Research Centre, Lähteenoja, V., Himanen, J., Turpeinen, M., and Signorelli, S. *The landscape of consent management tools - a data altruism perspective.* Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/0852673, JRC137572.
90. Fink, A. (2024). Data cooperative. *Internet Policy Review*, *13*(2). https://doi.org/10.14763/2024.2.1752
91. Nabben, K. (2024). AI as a Constituted System: Accountability Lessons from an LLM Experiment.
92. Von Thun, M., Hanley, D.A. (2024) Stopping Big Tech from Becoming Big AI. Open Markets Institute and Mozilla.

93. Rajamohan, R. (2024) Networked Cooperative Ecosystems. https://paragraph.xyz/@v6a/networked-ecosystems-2
94. Ananthaswamy, A. (2024) Why Machines Learn: The Elegant Math Behind Modern AI. London, UK: Penguin.
95. Bengio, Y. (2023) AI and catastrophic risk. *Journal of Democracy*. 34(4): 111-121.
96. European Parliament. *Social approach to the transition to smart cities*. European Parliament: Luxembourg, 2023.
97. Magro, A., (2024) Emerging digital technologies in the public sector: The case of virtual worlds. Luxembourg: Publications Office of the European Union.
98. Estévez Almenzar, M., Fernández Llorca, D., Gómez, E., & Martínez Plumed, F., 2022. *Glossary of human-centric artificial intelligence*. Publications Office of the European Union: Luxembourg. doi:10.2760/860665.
99. Calzada, I. & Almirall, E. (2020) Data Ecosystems for Protecting European Citizens' Digital Rights, *Transforming Government: People, Process and Policy (TGPPP)* 14(2): 133-147. DOI: 10.1108/TG-03-2020-0047.
100. **Calzada, I.;** Pérez-Batlle, M.; Batlle-Montserrat, J. People-Centered Smart Cities: An Exploratory Action Research on the Cities' Coalition for Digital Rights. *Journal of Urban Affairs 2021*, **43**, 1-26. DOI:10.1080/07352166.2021.1994861.
101. Mitchell, M., Palmarini, A.B. & Moskvichev, A. (2023) Comparing Humans, GPT-4, and GPT-4V on abstraction and reasoning tasks*. arXiv preprint*.
102. Gasser, U. & Mayer-Schönberger, V. (2024) *Guardrails: Guiding Human Decisions in the Age of AI*. Princeton, USA: Princeton University Press.
103. United Nations High-level Advisory Body on Artificial Intelligence (2024) *Governing AI for Humanity: Final Report*. United Nations, New York.
104. Vallor, S. (2024) The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking. NYC, USA: OUP.
105. Buolamwini, J. (2023) Unmasking AI: My Mission to Protect What is Human in a World of Machines. London, UK: Random House.
106. McCourt, F.H. Our Biggest Fight: Reclaiming Liberty, Humanity, and Dignity in the Digital Age. Crown Publishing: London, 2024.
107. Muldoon, J., Graham, M. & Cant, C. (2024) *Feeding the Machine: The Hidden Human Labour Powering AI*. Edinburgh, UK: Cannongate.
108. Burkhardt, S. & Rieder, B. (2024) Foundation models are platform models: Prompting and the political economy of AI. *Big Data & Society*, pp. 1–15. doi:10.1177/20539517241247839.
109. Finnemore, M. & Sikkink, K. (1998) International Norm Dynamics and Political Change. *International Organization*. 52: 887 - 917.
110. Lazar, S. (2024, forthcoming) Connected by Code: Algorithmic Intermediaries and Political Philosophy. Oxford: Oxford University Press.
111. Hoeyer, K. (2023) Data Paradoxes: The Politics of Intensified Data Sourcing in Contemporary Healthcare. Cambridge, MA, USA: MIT Press.
112. Hughes, T. (2024) The political theory of techno-colonialism. *European Journal of Political Theory*, pp. 1–24. doi:10.1177/14748851241249819.
113. Srivastava, S. Algorithmic Governance and the International Politics of Big Tech. Cambridge University Press: Cambridge, USA, 2021.
114. Utrata, A. (2024) Engineering territory: Space and colonies in Silicon Valley. *American Political Science Review*, 118(3), pp. 1097–1109. doi:10.1017/S0003055423001156.
115. Waldner, D. & Lust, E. (2018) Unwelcome change: Coming to terms with democratic backsliding. *Annual Review of Political Science*. 21(1): 93-113.
116. Guersenzvaig, A. & Sánchez-Monedero, J. (2024). AI research assistants, intrinsic values, and the science we want. *AI & Society*. https://doi.org/10.1007/s00146-023-01861-4.
117. Wachter-Boettcher, S. (2018) Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threat of Toxic Tech. London, UK: WW Norton & Co.

118. D'Amato, K. (2024). *ChatGPT: towards AI subjectivity*. *AI & Society*. https://doi.org/10.1007/s00146-024-01898-z.
119. Shavit, Y., et al. (2023) Practices for governing agentic AI systems. OpenAI.
120. Bibri, S.E.; Allam, Z. The Metaverse as a Virtual Form of Data-Driven Smart Urbanism: On Post-Pandemic Governance through the Prism of the Logic of Surveillance Capitalism. *Smart Cities* 2022, *5*, 715-727. https://doi.org/10.3390/smartcities5020037
121. Bibri, S.E.; Visvizi, A.; Troisi, O. *Advancing Smart Cities: Sustainable Practices, Digital Transformation, and IoT Innovations*; Springer: Cham, Switzerland, 2024. https://doi.org/10.1007/978-3-031-52303-8.
122. Ayyoob Sharifi, Zaheer Allam, Simon Elias Bibri, Amir Reza Khavarian-Garmsir, Smart cities and sustainable development goals (SDGs): A systematic literature review of co-benefits and trade-offs, Cities, Volume 146, 2024, 104659, ISSN 0264-2751, https://doi.org/10.1016/j.cities.2023.104659.
123. Singh, A. Advances in Smart Cities: Smarter People, Governance, and Solutions. *Journal of Urban Technology*, 2019, 1-4. doi:10.1080/10630732.2019.1637606.
124. Reuel, A., et al. (2024) Open Problems in Technical AI Governance. *arXiv preprint*. https://arxiv.org/abs/2407.14981.
125. Aho, B. Data communism: Constructing a national data ecosystem. *Big Data & Society*, 2024, pp. 1-14. doi:10.1177/20539517241275888.
126. Valmeekam, K., et al. (2023). On the Planning Abilities of Large Language Models—A Critical Investigation. *arXiv preprint*. https://arxiv.org/abs/2305.15771.
127. Yao, S., et al. (2022) ReAct: Synergizing reasoning and acting in language models. *arXiv preprint*. https://arxiv.org/abs/2210.03629.
128. **Calzada, I.**, The Illusion of the Web3 Decentralization: Distributing Power or Creating a New Tech-Savvy Elite? *SSRN, 2024,* DOI: 10.2139/ssrn.5008910
129. Lazar, S. & Pascal, A. (2024) AGI and Democracy. *Allen Lab for Democracy Renovation*.
130. Ovadya, A. (2023) Reimagining Democracy for AI. *Journal of Democracy*. 34(4): 162-170.
131. Ovadya, A.; Thorburn, L.; Redman, K.; Devine, F.; Milli, S.; Revel, M.; Konya, A.; Kasirzadeh, A. Toward Democracy Levels for AI. *Pluralistic Alignment Workshop at NeurIPS 2024*. Available online: https://arxiv.org/abs/2411.09222 (accessed on 14 November 2024).
132. Alnabhan, M.Q.; Branco, P. BERTGuard: Two-Tiered Multi- Domain Fake News Detection with Class Imbalance Mitigation. Big Data Cogn. Comput. 2024, 8, 93. https://doi.org/10.3390/bdcc8080093
133. Gourlet, P., Ricci, D. and Crépel, M. (2024) Reclaiming artificial intelligence accounts: A plea for a participatory turn in artificial intelligence inquiries. *Big Data & Society*, pp. 1–21. doi:10.1177/20539517241248093.
134. Spathoulas, G., Katsika, A., & Kavallieratos, G. (2024) *Privacy preserving and verifiable outsourcing of AI processing for cyber-physical systems*. Norwegian University of Science and Technology, University of Thessaly.
135. Abhishek, T. & Varda, M. (2024) Data hegemony: The invisible war for digital empires'. *Internet Policy Review*. Available from: https://policyreview.info/articles/news/data-hegemony-digital-empires/1789 (Accessed 1 September 2024).
136. Alaimo, C. & Kallinikos, J. (2024) *Data Rules: Reinventing the Market Economy*. Cambridge, MA, USA: MIT Press.
137. OpenAI, GPT-4 Technical Report. 2023.
138. Dobbe, R. (2022) System safety and artificial intelligence.' in Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.
139. Bengio, Y., et al. (2024) International Scientific Report on the Safety of Advanced AI: Interim Report.
140. World Digital Technology Academy (WDTA). 2024. *Large Language Model Security Requirements for Supply Chain*. WDTA AI-STR-03, World Digital Technology Academy.
141. AI4GOV. Available at: https://ai4gov-project.eu/2023/11/14/ai4gov-d3-1/ (Accessed 1 January 2024).
142. Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E., & Tavares, M.M., 2024. *Gen-AI: Artificial Intelligence and the Future of Work*. IMF Staff Discussion Note SDN2024/001. International Monetary Fund, Washington, DC.

143. ENFIELD (2024) Available from: https://www.enfield-project.eu/about (Accessed 1 September 2024). Call: oc1-2024-TES-01. SGA: oc1-2024-TES-01-01. Democracy in the Age of Algorithms: Enhancing Transparency and Trust in AI-Generated Content through Innovative Detection Techniques (PI: Prof Igor Calzada). Grant Agreement Number: 101120657. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/competitive-calls-cs/6083?isExactMatch=true&status=31094502&order=DESC&pageNumber=1&pageSize=50&sortBy=start Date

144. Palacios S. et al. (2022). AGAPECert: An Auditable, Generalized, Automated, Privacy-Enabling Certification Framework with Oblivious Smart Contracts. *Journal of Defendable and Secure Computing*, X,Y.

145. GPAI Algorithmic Transparency in the Public Sector (2024) *A State-of-the-Art Report of Algorithmic Transparency Instruments*. Global Partnership on Artificial Intelligence. Available from www.gpai.ai. (Accessed 1 September 2024).

146. Lazar, S. & Nelson, A. (2023) AI safety on whose terms? *Science*, 2023. 381(6654): 138-138.

147. HAI (2024) Artificial Intelligence Index Report 2024. Palo Alto, USA: HAI.

148. Nagy, P. & Neff, G. (2024) Conjuring algorithms: Understanding the tech industry as stage magicians. *New Media & Society*, 26(9), pp. 4938–4954.

149. Kim, E., Jang, G.Y. & Kim, S.H. (2022) How to apply artificial intelligence for social innovations. *Applied Artificial Intelligence*, 36(1). doi:10.1080/08839514.2022.2031819.

150. **Calzada, I.**; Cobo, C. Unplugging: Deconstructing the Smart City. *Journal of Urban Technology*, 2015, *22*, 23–43. doi: 10.1080/10630732.2014.971535

151. Visvizi, A.; Godlewska-Majkowska, H. Not Only Technology: From Smart City 1.0 through Smart City 4.0 and Beyond (An Introduction). In *Smart Cities: Lock-In, Path-dependence and Non-linearity of Digitalization and Smartification*; Visvizi, A., Godlewska-Majkowska, H., Eds.; Routledge: London, UK, 2025; pp. 3–16. Available online: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003415930-2/technology-anna-visvizi-hanna-godlewska-majkowska (accessed on).

152. Troisi, O.; Visvizi, A.; Grimaldi, M. The Different Shades of Innovation Emergence in Smart Service Systems: The Case of Italian Cluster for Aerospace Technology. *J. Bus. Ind. Mark.* 2024, *39*, 1105–1129. https://doi.org/10.1108/JBIM-06-2023-0302.

153. Visvizi, A.; Troisi, O.; Corvello, V. Research and Innovation Forum 2023: Navigating Shocks and Crises in Uncertain Times—Technology, Business, Society; Springer Nature: Cham, Switzerland, 2024. https://doi.org/10.1007/978-3-031-44721-1.

154. Federico Caprotti, Federico Cugurullo, Matthew Cook, Andrew Karvonen, Simon Marvin, Pauline McGuirk & Alan-Miguel Valdez (27 Mar 2024): Why does urban Artificial Intelligence (AI) matter for urban studies? Developing research directions in urban AI research, Urban Geography, DOI: 10.1080/02723638.2024.2329401

155. Federico Caprotti, Catalina Duarte, Simon Joss, The 15-minute city as paranoid urbanism: Ten critical reflections, Cities, Volume 155, 2024, 105497, https://doi.org/10.1016/j.cities.2024.105497.

156. Cugurullo, F.; Caprotti, F.; Cook, M.; Karvonen, A.; McGuirk, P.; Marvin, S. The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence. *Urban Studies*, 2024. DOI: DOI: 10.1177/00420980231203386.

157. Sanchez, T.W.; Fu, X.;Yigitcanlar, T.; Ye, X. The Research Landscape of AI in Urban Planning: A Topic Analysis of the Literature with ChatGPT. Urban Sci. 2024, 8, 197. https://doi.org/10.3390/urbansci8040197

158. Kuppler, A.; Fricke, C. Between innovative ambitions and erratic everyday practices: urban planners' ambivalences towards digital transformation. DOI: 10.3828/tpr.2024.41

159. Eubanks, V. (2019) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. London: Picador.

160. Lorinc, J. Dream States: Smart Cities, Technology, and the Pursuit of Urban Utopias. Toronto: Coach House Books, 2022.

161. Benjamin Leffel, Ben Derudder, Michele Acuto, Jeroen van der Heijden, Not so polycentric: The stratified structure & national drivers of transnational municipal networks, Cities, Volume 143, 2023, 104597, ISSN 0264-2751, https://doi.org/10.1016/j.cities.2023.104597.

162. Luccioni, S., Jernite, Y. & Strubell, E. (2024) Power hungry processing: Watts driving the cost of AI deployment? in *The 2024 ACM Conference on Fairness, Accountability, and Transparency*.
163. Gohdes, A.R. (2023) Repression in the Digital Age: Surveillance, Censorship, and the Dynamics of State Violence. Oxford, UK: Oxford University Press.
164. Seger, E., et al. (2020) Tackling threats to informed decision-making in democratic societies: Promoting epistemic security in a technologically-advanced world.
165. Burton, J.W., Lopez-Lopez, E., Hechtlinger, S. *et al.* (2024) How large language models can reshape collective intelligence. *Nat Hum Behav* 8, 1643–1655. https://doi.org/10.1038/s41562-024-01959-9
166. Lalka, R. (2024) *The Venture Alchemists: How Big Tech Turned Profits into Power*. New York, NY, USA: Columbia University Press.
167. Li, F.-F. (2023) The Worlds I See: Curiosity, Exploration, and Discovery and the Dawn of AI. London, UK: Macmillan.
168. Medrado, A. & Verdegem, P. (2024) Participatory action research in critical data studies: Interrogating AI from a South–North approach. *Big Data & Society*, 11(1).
169. Mejias, U.A.; Couldry, N. Data Grab: The New Colonialism of Big Tech (and How to Fight Back). WH Allen: London, 2024.
170. Murgia, M. (2024) *Code Dependent: Living in the Shadow of AI*. London, UK: Henry Holt and Co.
171. Johnson, S. & Acemoglu, D. (2023) Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity. London, UK: Basic Books.
172. Ludovico Rella, Kristian Bondo Hansen, Nanna Bonde Thylstrup, Malcolm Campbell-Verduyn, Alex Preda, Daivi Rodima-Taylor, Ruowen Xu & Till Straube (22 Oct 2024): Hybrid materialities, power, and expertise in the era of general purpose technologies, Distinktion: Journal of Social Theory, DOI: 10.1080/1600910X.2024.2414312
173. Merchant, B. (2023) Blood in the Machine: The Origins of the Rebellion Against Big Tech. London, UK: Little, Brown and Company.
174. Sieber, R., Brandusescu, A., Adu-Daako, A. & Sangiambut, S. (2024) Who are the publics engaging in AI? *Public Understanding of Science*. https://doi.org/10.1177/09636625231219853.
175. Tunç, A. (2024). Can AI determine its own future? *AI & Society*. https://doi.org/10.1007/s00146-024-01892-5.
176. Floridi, Luciano, Why the AI Hype is Another Tech Bubble (September 18, 2024). Available at SSRN: https://ssrn.com/abstract=4960826
177. Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013.
178. Batty, M. *Inventing Future Cities*; MIT Press: Cambridge, MA, USA, 2018.
179. Batty, M. Urban Analytics Defined. *Environment and Planning B: Urban Analytics and City Science* 2019, *46*, 403–405. doi:10.1177/2399808319842119.
180. Marvin, S.; Luque-Ayala, A.; McFarlane, C. *Smart Urbanism: Utopian Vision or False Dawn?*; Routledge: New York, NY, USA, 2016.
181. Marvin, S.; Graham, S. Splintering Urbanism: Networked Infrastructures, Technological Mobilities, and the Urban Condition; Routledge: London, UK, 2001.
182. Marvin, S.; Bulkeley, H.; Mai, L.; McCormick, K.; Palgan, Y.V. Urban Living Labs: Experimenting with City Futures. *European Urban and Regional Studies* 2018, *25*, 317–333. doi:10.1177/0969776418787222.
183. Kitchin, R. *Code/Space: Software and Everyday Life*; MIT Press: Cambridge, MA, USA, 2011.
184. Kitchin, R.; Lauriault, T.P.; McArdle, G. Knowing and Governing Cities through Urban Indicators, City Benchmarking, and Real-Time Dashboards. *Regional Studies, Regional Science* 2015, *2*, 6–28. doi:10.1080/21681376.2014.983149.
185. Calzada, I. (2020) Platform and data co-operatives amidst European pandemic citizenship. *Sustainability*, 12(20), p. 8309. doi:10.3390/su12208309.
186. Monsees, L. Crypto-Politics: Encryption and Democratic Practices in the Digital Era. Routledge: Oxon, UK, 2020
187. Visvizi, A.; Kozlowski, K.; **Calzada, I.;** Troisi, O. *Multidisciplinary Movements in AI and Generative AI: Society, Business, Education.* Edward Elgar: Chentelham, UK, 2025.
188. **Calzada. I.** *Datafied Democracies Unplugged*, Springer: Cham, Switzerland, 2025.

189. Palacios, S.; Ault, A.; Krogmeier, J.V.; Bhargava, B.; Brinton, C.G. AGAPECert: An Auditable, Generalized, Automated, Privacy-Enabling Certification Framework with Oblivious Smart Contracts. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 3269–3286. https://doi.org/10.1109/TDSC.2022.3192852.

190. Hossain, S.T.; Yigitcanlar, T. Local Governments Are Using AI without Clear Rules or Policies, and the Public Has No Idea. *QUT Newsroom*. Available online: https://www.qut.edu.au/news/realfocus/local-governments-are-using-ai-without-clear-rules-or-policies-and-the-public-has-no-idea (accessed on 9 January 2025).

191. Gerlich, M. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* **2025**, *15*, 6. https://doi.org/10.3390/soc15010006&#8203;:contentReference[oaicite:0]{index=0}.

192. Bousetouane, F. Agentic Systems: A Guide to Transforming Industries with Vertical AI Agents. *arXiv* **2025**, [cs.MA] 2501.00881. Available online: https://arxiv.org/abs/2501.00881 (accessed on 9 January 2025).

193. Fontana, S.; Errico, B.; Tedesco, S.; Bisogni, F.; Renwick, R.; Akagi, M.; Santiago, N. AI and GenAI Adoption by Local and Regional Administrations. European Union, Commission for Economic Policy, 2024. ISBN: 978-92-895-3679-0; doi: 10.2863/6007868.

194. Hossain, S.T.; Yigitcanlar, T.; Nguyen, K.; Xu, Y. Cybersecurity in Local Governments: A Systematic Review and Framework of Key Challenges. *Urban Governance*, Article in Press. Available online: https://doi.org/10.1016/j.ugj.2024.12.010 (accessed on 9 January 2025).

195. Laksito, J.; Pratiwi, B.; Ariani, W. Harmonizing Data Privacy Frameworks in Artificial Intelligence: Comparative Insights from Asia and Europe. *PERKARA – Jurnal Ilmu Hukum dan Politik* **2024**, *2*(4), 579–588. DOI: 10.51903/perkara.v2i4.2229.

196. Nature. Science for Policy: Why Scientists and Politicians Struggle to Collaborate. *Nature*, 2024. Available online: https://www.nature.com/articles/science4policy (accessed on 9 January 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.