Between the Algorithm and the People: Digital Populism and Legal Challenges in the Age of AI

[*]Tiago Matos

Abstract

This article examines the intersection between artificial intelligence (AI) and populism, arguing that the increasing appropriation of AI technologies by populist movements is reshaping the foundations of popular sovereignty in democratic societies. While populism claims to restore the direct will of "the people" against corrupt elites, its political strategies increasingly rely on algorithmic tools, such as recommender systems, automated social media bots, and predictive analytics, to influence public discourse, amplify polarisation, and circumvent traditional democratic institutions. These developments raise urgent legal questions about the integrity of democratic participation, the erosion of individual rights, and the role of law in regulating technological power. Through an interdisciplinary lens combining law, political theory and digital governance, the article explores how AI enables a new form of "algorithmic populism" that operates both as a rhetorical strategy and a governance technique. The study identifies key legal challenges, including gaps in accountability, manipulation of electoral processes, and threats to the democratic rule of law. It concludes by proposing regulatory principles capable of addressing these risks while safeguarding democratic values, stressing the need for legal frameworks that uphold human agency, transparency, and institutional oversight in the digital age.

Keywords: Digital populism, AI-generated disinformation, EU AI Act, democratic deliberation, algorithmic accountability.

I. Introduction

The convergence of artificial intelligence (AI) and digital populism represents a profound transformation in contemporary political dynamics, redefining the relationship between citizens, institutions, and technology. This phenomenon, termed digital populism, exploits the affordances of digital platforms, social media algorithms, generative AI, and data-driven

[*] Solicitor, Master's student in Law, specializing in Business Legal Sciences at Lusófona University of Porto, attending the Postgraduate Program in Real Estate Law at CIDP/FDUL, tiagojosematos@gmail.com

targeting, to propagate anti-elitist narratives, manipulate public opinion, and destabilize democratic processes. At its core lies the instrumentalization of liberal democracies' vulnerabilities, with technology serving as a tool to amplify divisive rhetoric, erode institutional trust, and distort the principles of popular sovereignty.[1]

Historically, populism has been defined as an ideology opposing "the pure people" to "the corrupt elites." In the digital age, this ideology has evolved into a movement characterized by direct emotional appeals, distrust in expertise, and the strategic use of disinformation. Social media platforms such as Facebook, Twitter/X, and YouTube have democratized access to information while simultaneously creating fertile ground for polarizing content. Algorithms optimized to maximize user engagement favor sensationalist, emotionally charged material over factual discourse, fostering information bubbles and echo chambers that reinforce ideological divides.[2]

AI exacerbates these dynamics by enabling unprecedented precision in public opinion manipulation. Tools such as large language models (LLMs) and deepfake technologies facilitate the production of hyper-realistic synthetic content—fabricated speeches, manipulated videos, and personalized propaganda—tailored to individual biases. For instance, during the 2024 U.S. presidential primaries, AI-generated content disseminated false claims about candidates' policies, exploiting pre-existing fears and prejudices. Similarly, the 2023 Mata v. Avianca case[3] highlighted how AI-generated legal citations infiltrated court proceedings, compromising judicial integrity when lawyers submitted unverified arguments produced by ChatGPT.

This phenomenon directly undermines popular sovereignty —the principle that ultimate political authority resides in the people—by replacing informed citizen participation with algorithmic manipulation. Populist leaders and movements leverage AI to microtarget voters, delivering disinformation that reinforces biases and obscures facts. The result is a distortion of

[1] In his paper "Digital Populism and Extreme Digital Populism" (original title: "Populismo digital e populismo digital extremista"), presented at the 22nd National Meeting of Journalism Researchers, Assunção (2025) analyzes these phenomena in the Brazilian context, demonstrating how populist actors instrumentalize algorithmic infrastructures to circumvent traditional gatekeepers and amplify divisive narratives.

[2] Social media algorithms prioritize emotionally charged content, creating "echo chambers" that reinforce ideological divides, as analyzed by Stanley (2008) and Mudde (2004).

[3] The 2023 Mata v. Avianca case, decided by the United States District Court for the Southern District of New York, highlights risks of AI misuse in legal contexts. In this case, attorney Steven Schwartz submitted a legal brief containing fabricated judicial citations generated by ChatGPT, including non-existent cases with fictitious quotes and legal holdings. Judge P. Kevin Castel imposed sanctions for bad faith and failure to verify the authenticity of legal sources, establishing an important precedent regarding professional responsibility in the use of generative AI tools (Mata v. Avianca, Inc., No. 22-cv-1461, S.D.N.Y. 2023).

democratic deliberation: instead of evidence-based debate, citizens are fed manipulated narratives validating their frustrations and demonizing the "other."

The Cambridge Analytica scandal exemplifies how psychographic profiling and AI-driven ad campaigns were used during the 2016 U.S. elections to manipulate voter behavior through psychological exploitation. These practices not only influence electoral outcomes but also erode public confidence in electoral systems, fostering perceptions of systemic manipulation by hidden forces rather than collective will. This climate of suspicion fuels a legitimacy crisis, where institutions such as governments and media are perceived as part of an elitist conspiracy, and even truth becomes a contested terrain.

In this context, the integration of AI into digital populism raises urgent legal and ethical challenges. While frameworks like the European Artificial Intelligence Act (Regulation (EU) 2024/1689) and the General Data Protection Regulation (Regulation (EU) 2016/679) aim to regulate high-risk AI systems and protect personal data, gaps persist in addressing the epistemic harms of AI-generated disinformation, such as deepfakes and fake news. Courts struggle to hold perpetrators accountable under traditional defamation laws, particularly when authorship is obscured by automated tools.

The tension between free speech and regulatory necessity complicates digital governance. Authoritarian regimes may impose draconian restrictions on AI misuse, but liberal democracies face the challenge of balancing technological innovation with civic safeguards. The AIA's risk-based approach prohibits biometric surveillance and subliminal manipulative techniques but lacks clarity on enforcing compliance across decentralized platforms. Meanwhile, the Digital Services Act (DSA) mandates the removal of illegal content but faces criticism for relying on self-regulation by tech giants.

This paper examines how AI-powered digital populism reconfigures popular sovereignty and explores legal solutions to safeguard democratic integrity in the era of algorithmic governance. The analysis will draw on case studies, regulatory frameworks, and comparative approaches. Central research questions include:

1. How does AI-driven digital populism—understood as the strategic use of algorithmic systems, automated content generation, and data-driven microtargeting to amplify anti-elitist narratives—redefine popular sovereignty, specifically the principle that legitimate political authority derives from the informed and autonomous will of the people?

2. What legal innovations—including transparency mandates, accountability frameworks, and rights-based safeguards—are necessary to protect democratic integrity (encompassing free and fair elections, pluralistic deliberation, and institutional trust) in the age of algorithmic governance, where automated systems increasingly mediate political communication and decision-making?

The analysis will be grounded in interdisciplinary scholarship, including Cas Mudde's work on populism's binary logic, Shoshana Zuboff's critique of surveillance capitalism,[4] and Mireille Hildebrandt's warnings about AI's "black box" accountability risks. Empirical evidence from recent elections, judicial rulings, and regulatory debates will underscore the urgency of reconciling technological progress with democratic resilience.

This article proceeds as follows: Section II establishes the theoretical and interdisciplinary framework, examining the conceptual foundations of digital populism, popular sovereignty, and algorithmic governance. Section III analyzes how AI functions as a vector of populist mobilization through case studies including Cambridge Analytica, deepfakes, and algorithmic content curation. Section IV identifies key legal challenges and regulatory gaps in addressing AI-driven populism, with comparative analysis of frameworks in the EU, UK, Canada, Portugal, and Cyprus. Section V concludes with policy recommendations and a normative vision for defending democratic sovereignty in the algorithmic age.

II. Theoretical and Interdisciplinary Framework

This section articulates the theoretical and conceptual underpinnings essential to examining the convergence of digital populism, popular sovereignty and algorithmic governance. Drawing upon interdisciplinary perspectives from political theory, digital ethics and legal scholarship, it seeks to illuminate how AI-driven populism challenges foundational democratic norms and reshapes the legal-political landscape in the digital age.

1. Digital Populism: From Anti-Elitist Discourse to Algorithmic Mobilization

Digital populism refers to the strategic appropriation of digital technologies, particularly social media platforms and AI-driven tools, to mobilize political support through disinformation,

[4] Zuboff (2019) explains how surveillance capitalism operates through the extraction of personal data to fuel targeted political manipulation, as digital platforms monetize psychographic profiles to deliver hyper-personalized messaging.

polarization and anti-institutional rhetoric. As Youngmi Kim observes, it constitutes “a new type of political behavior marked by the political use of the Internet as both a form of protest and a tool for power consolidation.” Unlike classical populism, which often relied on charismatic leadership and mass mobilizations, digital populism is characterized by its reliance on algorithmic amplification, generative AI and data analytics to craft and disseminate divisive narratives at scale.

This transformation is exemplified by political figures such as Donald Trump in the United States, Jair Bolsonaro in Brazil and Matteo Salvini in Italy. Trump’s prolific use of Twitter/X to propagate unverified claims and delegitimize mainstream media demonstrates how digital platforms facilitate direct and unmediated appeals to “the people,” circumventing traditional institutional gatekeepers. Likewise, Bolsonaro’s deployment of WhatsApp for the viral dissemination of misinformation during his electoral campaign underscores how encrypted, decentralized communication channels can be weaponized to avoid scrutiny and factual correction.

The advent of generative AI technologies, including large language models and deepfake tools, has exacerbated these dynamics. These technologies enable the mass production of hyper-realistic synthetic content, such as fabricated speeches, manipulated videos and bespoke propaganda, which reinforce confirmation biases and emotional resonance over factual integrity. During the 2024 U.S. presidential primaries, for instance, AI-generated content was employed to circulate falsehoods concerning candidates’ policy positions, exploiting public fears and prejudices. The result is a deeply destabilizing political environment in which democratic deliberation is supplanted by algorithmically engineered spectacles of outrage and suspicion.

2. Popular Sovereignty and the Fragmentation of the Public Sphere

The principle of popular sovereignty, which posits that ultimate political authority resides with the people, is increasingly imperiled by the logic of algorithmic curation. Democratic legitimacy presupposes a public sphere in which citizens have access to pluralistic, fact-based discourse and are able to deliberate freely and critically. However, contemporary digital infrastructures, optimized for engagement and monetization, prioritize sensationalist and emotionally provocative content over accuracy and nuance. Social media algorithms curate individual information ecosystems, producing so-called “filter bubbles” and “echo chambers” that entrench ideological divides.

The implications for democratic governance are profound. As articulated in *Between the Algorithm and the People*, algorithmic mediation fragments shared realities, rendering the formation of collective will increasingly elusive. Citizens become both targets and inadvertent amplifiers of mis- and disinformation, eroding trust in the institutions tasked with safeguarding public truth, namely governments, courts and independent media.

The Cambridge Analytica scandal serves as a paradigmatic example. By leveraging psychographic profiling and AI-driven behavioral targeting, private actors were able to manipulate electoral behavior through personalized psychological exploitation. As documented by Azgin and Kiralp (2024), the firm illicitly harvested data from approximately 87 million Facebook users and deployed the OCEAN personality model to craft hyper-targeted political advertisements during the 2016 U.S. presidential election, demonstrating how algorithmic systems can be weaponized to undermine electoral autonomy and informed consent.

Cas Mudde's conceptualization of populism as a moral binary between "the pure people" and "the corrupt elite" offers a critical lens through which to interpret this fragmentation. In the digital context, populist actors exploit algorithmic structures to amplify this dichotomy, framing dissenting voices as enemies of the people and institutional pluralism as betrayal. Populist actors strategically exploit this epistemic insecurity by framing AI-generated disinformation as "authentic" counter-narratives to institutional expertise, thereby positioning themselves as truth-tellers against allegedly corrupt elites while simultaneously undermining the shared factual basis necessary for democratic contestation. The digital sphere thus becomes a battleground where legitimacy is defined not by rational consensus but by emotional allegiance and algorithmic virality. In such a milieu, the conditions for authentic popular sovereignty are progressively undermined.

3. Algorithmic Governance: Legal and Ethical Challenges

The incorporation of algorithmic systems into governance structures raises significant legal and ethical dilemmas, particularly regarding transparency, accountability and democratic oversight. Algorithmic governance refers to the increasing reliance on AI and automated decision-making systems to mediate political, legal and social interactions. These systems, often opaque and proprietary, pose a structural challenge to the rule of law, as they operate through mechanisms that resist scrutiny and contestation.

Shoshana Zuboff's theory of "surveillance capitalism" underscores how the commodification of personal data fuels targeted political manipulation. Platforms such as Facebook and Google extract user data to construct psychometric profiles, which are then monetized through hyper-

personalized advertising. In the political arena, this translates into the ability to deliver micro-targeted messages that exploit cognitive and emotional vulnerabilities. The blurred boundary between persuasion and coercion calls into question the voluntariness and autonomy of the electorate, a cornerstone of democratic legitimacy.

From a legal perspective, the opacity of AI systems presents acute governance risks. As Mireille Hildebrandt argues, the legal order is ill-equipped to regulate systems that lack explainability and traceability. A telling example is the 2023 Mata v. Avianca case, in which legal counsel submitted fabricated judicial citations generated by ChatGPT. The court-imposed sanctions for bad faith, underscoring the dangers of uncritical reliance on AI outputs in sensitive legal contexts. Hildebrandt contends that AI must be governed by "legal-by-design" principles, which embed legal safeguards into the architectural fabric of the technology itself.

4. Regulatory Responses and Their Limitations

Regulatory frameworks designed to govern AI and digital platforms often struggle to keep pace with technological innovation. Key challenges include:

- **Enforcement Gaps**: While the EU's Digital Services Act (DSA) mandates the removal of illegal content, its reliance on platform self-regulation has been criticized for leading to inconsistent enforcement and failing to deter systemic abuses.
- **Jurisdictional Fragmentation**: The transnational nature of digital platforms complicates enforcement, as companies may operate from jurisdictions with more permissive regulatory environments, creating opportunities for regulatory arbitrage.
- **Regulatory Gaps**: Instruments such as the EU's Artificial Intelligence Act (AIA) and the General Data Protection Regulation (GDPR) represent initial attempts to govern high-risk AI systems. However, significant enforcement and interpretative gaps persist. Hildebrandt's advocacy for risk-based, context-sensitive regulation offers a promising path forward, emphasizing adaptability and resilience over rigid formalism.

5. Toward a Framework for Democratic Resilience

The theoretical and interdisciplinary analysis above reveals that digital populism, amplified by artificial intelligence and algorithmic infrastructures, poses a structural threat to the foundations of democratic governance. Recognizing these challenges, the European Union has developed the Democratic Shield initiative, a comprehensive framework designed to protect democratic processes from foreign interference, disinformation, and hybrid threats. Launched by the European Commission, the Democratic Shield encompasses measures to enhance media

literacy, strengthen electoral integrity, counter foreign information manipulation, and support independent journalism. This initiative operates in conjunction with the Digital Services Act, the AI Act, and the European Democracy Action Plan to create a multi-layered defense of democratic institutions.

Building upon the Democratic Shield's institutional architecture, and complementing the Council of Europe's standards on freedom of expression and media pluralism, a robust framework for democratic resilience must address both the technological and normative dimensions of AI-driven populism. To operationalize this vision and foster democratic resilience, a multi-pronged institutional response is required, grounded in the adaptation of traditional rule of law principles to the algorithmic age.

The first pillar is **transparency**, a principle deeply rooted in liberal constitutionalism and administrative law. Traditionally, transparency has required that governmental actions be open to public scrutiny, enabling citizens to hold power accountable. In the digital context, this principle must extend to algorithmic systems that mediate political communication. Legal mandates must enforce the clear disclosure and labelling of AI-generated content, especially in political communication, ensuring citizens are aware of synthetic or manipulative information. This represents an evolution of the traditional "notice" requirement in administrative procedure, adapted to address the opacity of automated decision-making systems. The EU's AI Act partially addresses this through transparency obligations for certain AI systems, but enforcement remains inconsistent, and coverage of political communication tools is incomplete.

The second pillar is **accountability**, which corresponds to the rule of law principle that power must be subject to legal constraint and that wrongdoers must face consequences. In traditional legal systems, accountability operates through clearly defined chains of responsibility, judicial review, and sanctions for misconduct. However, AI systems complicate this framework due to their distributed nature, emergent behaviors, and the involvement of multiple actors (developers, deployers, users). Developers, platforms, and political actors must be held legally and ethically responsible for the harms caused by deceptive or polarizing uses of AI technologies. This requires updating liability frameworks—such as the Product Liability Directive and the emerging AI Liability Directive—to address algorithmic harms, including those resulting from generative models used for disinformation. Unlike traditional tort law, which presumes identifiable human agency, AI accountability must grapple with questions of

foreseeability, causation, and the attribution of responsibility across complex socio-technical systems.

The third pillar is **digital literacy**, which builds upon the democratic principle of an informed citizenry. Classical democratic theory, from John Stuart Mill to Jürgen Habermas, has emphasized that legitimate self-government depends on citizens' capacity for rational deliberation and critical evaluation of information. In the algorithmic age, this capacity must be actively cultivated through civic education initiatives expanded to equip citizens with the analytical tools necessary to navigate algorithmic environments, recognize manipulation, and engage in informed political deliberation. This goes beyond traditional civic education by incorporating technical literacy (understanding how algorithms curate information), epistemic resilience (distinguishing credible from manipulated sources), and participatory competencies (engaging meaningfully in digitally mediated public spheres). Countries such as Estonia and Finland have pioneered such programs, integrating digital citizenship into national curricula and demonstrating measurable improvements in resistance to disinformation.

By integrating insights from political theory, digital ethics, and constitutional law, this framework offers more than regulatory fixes: it affirms a normative foundation for democratic resilience grounded in the active defense of transparency, pluralism, informed participation, and the rule of law. In the face of AI-driven populism, safeguarding democracy requires not only innovation in governance, but also the reaffirmation of constitutional values in the design, deployment, and oversight of algorithmic systems.

III. Artificial Intelligence as a Vector of Populist Mobilization

This chapter investigates how AI functions as a tool of populist influence through detailed examination of three emblematic case studies: first, the Cambridge Analytica affair, which exemplifies psychographic manipulation; second, the proliferation of AI-generated disinformation through deepfakes and social bots; and third, the operational logic of algorithmic content delivery on platforms like YouTube and Facebook. Each case study illustrates distinct mechanisms through which AI enables and amplifies populist mobilization, collectively demonstrating the multifaceted nature of this phenomenon.

1. Cambridge Analytica and the Weaponization of Psychographic Data

The Cambridge Analytica scandal remains a paradigmatic example of how AI-driven psychometric techniques may be appropriated for political manipulation. Through the illicit

acquisition of personal data from millions of Facebook users, the firm developed psychographic profiles based on the OCEAN personality model (openness, conscientiousness, extraversion, agreeableness, neuroticism). These models enabled the creation of finely tuned persuasive content designed to exploit specific psychological predispositions of voters.[5]

Such profiling facilitated the strategic deployment of emotionally provocative messaging, particularly in the context of the 2016 United States presidential election. Political actors associated with the Trump campaign harnessed these insights to disseminate divisive narratives, framing political opponents as embodiments of corruption and presenting populist candidates as authentic representatives of the "real people." Crucially, these tactics operated by reinforcing the fundamental populist dichotomy between the virtuous populace and the allegedly corrupt elite.

As Zuboff (2019) argues in *The Age of Surveillance Capitalism*, such practices exemplify how personal data is extracted, analyzed, and weaponized to shape behavior at scale, transforming citizens into objects of prediction and manipulation rather than autonomous political agents. Mittelstadt et al. (2016) further demonstrate how algorithmic profiling creates ethical challenges related to consent, transparency, and the potential for discrimination, particularly when deployed in high-stakes political contexts.

2. Synthetic Media and the Crisis of Epistemic Authority

The second case study examines the rise of synthetic media, particularly deepfakes, as a tool of populist disinformation. Technological advances in generative AI have rendered the production of synthetic media both technically feasible and economically accessible. Deepfakes—a term combining "deep learning" and "fake," referring to AI-generated videos or audio clips that depict individuals saying or doing things they never did—are increasingly used to fabricate political events and undermine public trust. For example, during the 2024 general elections in India, AI-generated voice messages purporting to be from political candidates were disseminated across rural constituencies, misleading voters and sowing confusion.[6]

In parallel, bot networks operating on social media platforms serve to simulate grassroots support for populist agendas. These automated agents amplify partisan messaging, create the

[5] Azgin & Kiralp (2024) illustrate how the Cambridge Analytica scandal exemplifies the use of psychographic profiling in microtargeting voters, reinforcing the populist dichotomy between "the people" and "elites."
[6] Rădulescu (2024) highlights how AI-generated voice calls targeting rural voters during India's 2024 elections underscore the urgent need to regulate synthetic media.

illusion of widespread consensus, and suppress critical voices through coordinated disinformation campaigns. The 2018 Brazilian elections offer a salient case: large-scale bot activity was deployed to propagate anti-corruption rhetoric in favor of then-candidate Jair Bolsonaro, thereby marginalizing dissent and consolidating populist appeal.

These phenomena contribute to a broader epistemological crisis. As the capacity to distinguish between authentic and fabricated information deteriorates, citizens increasingly rely on emotive heuristics and partisan cues rather than empirical validation. Populist actors exploit this epistemic insecurity by presenting AI-generated content as "genuine" revelations suppressed by dominant elites, thereby delegitimizing institutional sources of truth.[7]

3. The Architecture of Algorithmic Populism: Platform Design as Political Infrastructure

The third case study shifts focus from specific actors or content types to the structural features of digital platforms themselves. Social media platforms have become central arenas for political communication, and their algorithmic architectures are not ideologically neutral. Rather, as this case study demonstrates, the design choices embedded in recommendation algorithms, engagement metrics, and content moderation systems create systematic affordances that favor populist communication strategies.

Empirical studies have demonstrated that such algorithms can lead users down "radicalization pathways" by continuously recommending increasingly extreme content. On YouTube, for instance, exposure to mildly controversial videos has been shown to trigger a cascade of suggestions culminating in conspiratorial or extremist material. This process reinforces existing cognitive biases, deepens ideological silos, and curtails exposure to dissenting viewpoints.

The political implications are far-reaching. In the Italian context, the League party led by Matteo Salvini capitalized on Facebook's targeting tools to disseminate xenophobic imagery, optimizing campaign messages for maximum virality.[8] The cumulative effect is a feedback loop wherein populist actors adapt their communications to the affordances of the algorithm,

[7] Momoc (2018) explores how AI-generated deepfakes intensify "epistemic insecurity," a condition in which citizens struggle to distinguish truth from falsehood, fragmenting shared realities and enabling the spread of digital populism.

[8] Battista and Mangone (2025) examine how Matteo Salvini's League party employed Facebook's targeting tools to circulate xenophobic imagery, strategically optimizing content for virality and emotional resonance among specific voter segments.

and in turn, the algorithm privileges their narratives. This reciprocal relationship constitutes what may be termed algorithmic populism, wherein technological infrastructures become co-conspirators in the erosion of pluralistic dialogue.[9]

4. Mechanisms of Populist Engagement in the Digital Age

Populist movements increasingly circumvent legacy media and institutional mediators, opting instead for direct communication channels mediated by AI-curated platforms. This strategy enhances the immediacy and emotionality of political messaging. Donald Trump's prolific use of Twitter/X exemplifies this approach: despite recurrent breaches of platform guidelines, his content consistently achieved high levels of algorithmic visibility, fostering a sense of personal connection with his electoral base.

Similarly, the Bolsonaro campaign in Brazil deployed AI-driven data analytics to segment audiences across WhatsApp groups and deliver hyper-localized content.[10] This form of micropropaganda framed Bolsonaro as a defender of the "ordinary citizen" against opaque elite conspiracies, effectively mobilizing disaffected voters.

At the epistemological level, populists portray AI-generated content not as synthetic distortions but as revelatory disclosures of hidden truths. By reframing disinformation as resistance to hegemonic narratives, populist leaders position themselves as epistemic insurgents, validating their claims through virality rather than veracity. This strategy not only undermines public trust in conventional media and expertise but also licenses authoritarian interventions under the guise of democratic renewal.

Having established how AI functions as a vector of populist mobilization, the analysis now turns to the legal and regulatory challenges posed by these developments. The following section examines the structural gaps in existing legal frameworks and explores comparative approaches to governing AI-driven populism across multiple jurisdictions.

[9] Hildebrandt (2020) advocates for 'legal-by-design' principles, embedding constitutional safeguards (e.g., transparency, due process) into AI systems. This approach contrasts with the EU AI Act's reactive model and could mitigate risks like AI-generated disinformation in elections.

[10] Brussino and Alonso (2021) analyze how Bolsonaro's campaign leveraged WhatsApp to disseminate misinformation through encrypted messaging, thereby circumventing fact-checking mechanisms and intensifying societal polarization.

IV. Legal Challenges and Regulatory Gaps: Addressing AI-Driven Populism

The integration of artificial intelligence (AI) into populist strategies presents significant legal and regulatory challenges, particularly concerning the preservation of democratic integrity, the establishment of accountability for AI-induced harms, and the reconciliation of free speech protections with the imperative to mitigate manipulative technologies. This chapter critically examines three principal legal issues, erosion of democratic deliberation, accountability gaps, and the free speech versus regulation dilemma, while analyzing regulatory approaches across the European Union (EU), the United Kingdom (UK), Canada, and member states such as Portugal and Cyprus. These two member states are examined as illustrative examples of smaller EU jurisdictions facing distinct implementation challenges: Portugal represents a Southern European context with growing digital infrastructure but limited regulatory capacity, while Cyprus exemplifies the complexities of a small island state with significant cross-border digital flows and jurisdictional vulnerabilities to external disinformation campaigns.

1. Erosion of Democratic Deliberation: Fragmentation and Epistemic Crisis

The proliferation of algorithmic populism has led to the fragmentation of public discourse, undermining the foundational principle of democratic deliberation: informed, rational debate among citizens. Social media platforms, driven by engagement-optimized algorithms, prioritize sensationalist and emotionally charged content over factual accuracy, fostering filter bubbles and echo chambers that entrench ideological polarization. This dynamic is further exacerbated by AI-generated disinformation, including deepfakes and synthetic text, which distort reality and erode shared understandings of truth.

For instance, the case of Mata v. Avianca in the United States highlighted how AI-generated legal citations infiltrated judicial processes, compromising the integrity of legal reasoning when attorneys submitted unverified outputs from ChatGPT. Similarly, during India's 2024 elections, AI-generated voice calls impersonating politicians inundated rural constituencies with false promises and divisive rhetoric, exploiting low digital literacy to manipulate voter behavior.

2. Accountability Gaps in AI Governance

Establishing accountability for harms caused by AI-driven populism is a formidable legal challenge. The opacity of algorithmic systems, coupled with the distributed nature of digital platforms, complicates the attribution of responsibility. The EU's regulatory framework, while ambitious, reveals significant gaps in this regard.

2.1 Limitations of the EU Regulatory Framework

The EU's Digital Services Act (DSA), while strengthening due diligence obligations in relation to online platforms, has raised concerns regarding its efficacy in addressing emerging threats such as disinformation and deepfakes. Rather than imposing binding obligations on social media companies, the EU largely delegates monitoring and moderation duties to the platforms themselves, reinforcing a model of self-regulation that has proven insufficient in curbing algorithmic abuses.[11]

This regulatory gap is especially problematic for so-called "limited-risk" systems—such as generative AI models used to produce disinformation—that fall outside the Act's strictest compliance requirements but still produce significant societal harm.[12] Furthermore, existing legal instruments such as the Product Liability Directive remain ill-equipped to handle the complexities of accountability in AI contexts. In the widely cited case of Mata v. Avianca, sanctions were imposed on legal practitioners who relied on fabricated AI-generated citations, but the creators of the underlying technology were not held responsible.[13] This example illustrates the structural challenges of attributing liability when the causal chain involves autonomous and widely distributed systems. In addition, it is crucial to consider complementary legislative initiatives recently adopted by the European Union, such as the Artificial Intelligence Liability Directive (Directive (EU) 2024/2853). This directive aims to strengthen accountability mechanisms in cases involving harm caused by AI systems, introducing significant procedural innovations, including a presumption of fault under certain conditions and improved access to evidence in litigation involving autonomous technologies.

[11] Monti (2021) critiques the EU's overreliance on platform self-regulation, arguing that it leads to uneven enforcement and undermines the deterrent effect of digital legislation.

[12] The principle of proportionality in AI regulation requires balancing democratic safeguards with technological innovation. As Stanley (2008) argues, populist movements often reject deliberative norms, framing any regulation as elitist overreach. This complicates risk-based frameworks like the EU AI Act, which must reconcile algorithmic harms with free speech protections.

[13] Raji et al. (2020) discuss structural barriers to holding AI developers accountable, especially in contexts where harm results from emergent system behavior rather than direct human intent.

These instruments constitute important advances in addressing structural barriers to the effective liability of AI developers and operators. However, their applicability in political contexts, particularly in addressing algorithmic disinformation and electoral manipulation, warrants further examination. Understanding how these legal tools can be mobilised to hold actors accountable for the deceptive or polarising use of AI, especially when such practices compromise the integrity of democratic institutions, remains a pressing concern.[14]

2.2 Challenges in Transnational Enforcement

The cross-border nature of AI platforms presents a fundamental obstacle to effective regulation. Although the EU asserts the extraterritorial scope of its AI governance regime, practical enforcement is often inconsistent. Non-EU companies may fail to comply due to legal conflicts, uneven implementation or limited capacity to adapt their systems to EU norms. Prominent platforms such as TikTok and Facebook have repeatedly resisted or delayed content moderation obligations under the Digital Services Act, raising questions about the Union's reliance on self-regulatory commitments.[15]

In contrast, jurisdictions like New Zealand have adopted stricter accountability measures, including legal sanctions for the misuse of AI in contexts such as judicial proceedings or governmental communication.[16] These approaches offer valuable models for more coercive regulatory designs that go beyond voluntary compliance. Meanwhile, the proliferation of deepfakes during events such as the 2024 United States primaries has underscored the ease with which malicious actors can exploit jurisdictional blind spots. Content produced outside the EU often falls outside its regulatory reach, even when it circulates widely within the Union. Addressing these vulnerabilities may require the introduction of extraterritorial compliance requirements, mandating that foreign AI providers who disseminate content in EU jurisdictions adhere to minimum transparency and due diligence obligations.

Finally, disparities in enforcement capacity between Member States further undermine the EU's normative ambitions. While some regulators adopt proactive stances, others lack the

[14] As highlighted by Raji et al. (2020), the difficulties in attributing responsibility in AI contexts stem precisely from the complexity and distributed nature of technological systems. In this regard, the AI Liability Directive represents a promising normative development, as it introduces presumptions of fault and mechanisms facilitating access to evidence in cases of algorithmic harm—features that are particularly relevant in scenarios involving politically mediated manipulation through AI systems.

[15] Monti (2021) highlights the enforcement disparities across Member States, noting that regulatory capacity and political will vary widely, enabling inconsistent application of EU tech law.

[16] Raji et al. (2020) point to New Zealand as an example of a jurisdiction with concrete legal penalties for AI misuse, particularly in high-stakes domains such as legal or governmental processes.

institutional resources or political will to enforce existing rules evenly across the digital landscape. This fragmentation reinforces asymmetries that allow powerful actors to forum-shop for the most permissive environments.

3. Free Speech vs. Regulation: Balancing Innovation and Democratic Safeguards

The tension between free speech protections and the necessity to regulate manipulative AI systems remains a contentious issue. While authoritarian regimes like China impose stringent controls on AI-generated content, liberal democracies grapple with balancing innovation with safeguards against disinformation.

3.1 EU's Risk-Based Approach vs. UK/Canada's Sector-Specific Models

The EU's AIA exemplifies a precautionary, harmonized framework, banning high-risk AI systems outright (e.g., social scoring, emotion recognition in law enforcement) while imposing transparency obligations on "limited-risk" systems like chatbots. In contrast, the UK's pro-innovation regulatory strategy emphasizes sector-specific guidelines, encouraging self-regulation by tech firms. For example, the UK's Information Commissioner's Office (ICO) requires AI developers to assess risks to human rights but stops short of banning specific technologies.

Canada's approach lies between these extremes. While lacking a comprehensive AI law akin to the AIA, its Algorithmic Impact Assessment mandates federal agencies to evaluate biases in automated decision-making systems. However, critics argue this framework lacks enforceability, as compliance is largely voluntary.

3.2 Portugal and Cyprus: Disparities in Implementation and the Challenge of Regulatory Capacity

Smaller EU member states like Portugal and Cyprus face unique structural challenges in enforcing the AIA uniformly, highlighting broader questions about the feasibility of harmonized AI governance across jurisdictions with vastly different resources and institutional capacities. Limited technical capacity, budgetary constraints, and competing political priorities hinder their ability to audit AI systems, investigate algorithmic harms, or penalize violations effectively.

Portugal: A Case Study in Resource Constraints

Portugal provides an instructive example of the challenges facing mid-sized EU member states. While the National Data Protection Commission (CNPD) has demonstrated competence in enforcing GDPR, issuing significant fines for data protection breaches, it lacks the specialized resources necessary to monitor AI-generated disinformation in real-time or to conduct forensic audits of complex algorithmic systems deployed during electoral campaigns. The 2024 Portuguese legislative elections saw increased circulation of synthetic media on social platforms, yet no systematic regulatory response was mounted due to resource limitations and the absence of clear legal mandates under national law. This gap is particularly concerning given Portugal's growing role as a digital hub in Southern Europe, with increasing numbers of tech companies establishing operations in Lisbon and Porto. The Portuguese government has announced plans to develop a national AI strategy, but implementation remains in early stages, and coordination between the CNPD, the media regulator (ERC), and electoral authorities (CNE) is fragmented. Moreover, Portugal faces linguistic and cultural challenges in addressing disinformation. Content moderation systems deployed by major platforms are often optimized for English, Spanish, and other major languages, leaving Portuguese-language content—particularly from Brazil and African lusophone countries—inadequately monitored. This creates vulnerabilities that can be exploited by both domestic and foreign actors seeking to manipulate Portuguese public discourse.

Cyprus: Vulnerabilities of a Small Island State

Cyprus faces even more acute challenges. As a small island state with a population under one million, Cyprus's regulatory bodies operate with minimal staff and technical expertise in AI governance. The Office of the Commissioner for Personal Data Protection, Cyprus's equivalent to the CNPD, has a staff of fewer than 20 individuals responsible for all data protection and emerging AI-related issues. This severely limits the state's capacity to conduct proactive monitoring, investigate complex cases, or impose meaningful sanctions on large multinational platforms. Moreover, Cyprus's strategic location and its role as a financial and digital services center make it particularly vulnerable to cross-border disinformation campaigns, including those originating from non-EU actors. The island's divided status, with ongoing political tensions between the Republic of Cyprus and the Turkish Republic of Northern Cyprus, creates additional vectors for information manipulation. During the 2023 presidential elections, Cypriot authorities documented instances of coordinated inauthentic behavior on social media,

including bot networks amplifying divisive narratives about reunification and migration. However, they lacked the legal tools and technical capacity to attribute responsibility or impose sanctions. The absence of a comprehensive national AI strategy further exacerbates these vulnerabilities. While Cyprus has participated in EU-level discussions on AI regulation, domestic implementation has been slow. The country's small size means it lacks the economies of scale necessary to develop indigenous AI expertise or to attract significant investment in digital governance infrastructure. This creates a dependency on external actors—including the very platforms it seeks to regulate—for technical assistance and capacity building.

Implications for EU Harmonization

These disparities underscore a fundamental tension in the EU's approach to AI regulation: while the AIA aspires to harmonization, its effectiveness depends on the capacity of individual member states to implement and enforce its provisions. The experiences of Portugal and Cyprus reveal that formal legal harmonization does not automatically translate into substantive regulatory convergence when member states possess vastly different institutional resources. Without targeted capacity-building initiatives, technical assistance programs, and potentially centralized enforcement mechanisms at the EU level, smaller member states risk becoming weak links in the Union's regulatory architecture. This not only undermines the overall integrity of the framework but also creates opportunities for regulatory arbitrage, where actors seeking to evade scrutiny can exploit jurisdictional gaps by routing operations through under-resourced member states. Addressing these challenges will require a multi-faceted approach: financial support through EU structural funds for regulatory capacity building; technical assistance through agencies such as the European Union Agency for Cybersecurity (ENISA); knowledge-sharing networks connecting regulators across member states; and consideration of supranational enforcement mechanisms for cases involving cross-border algorithmic harms. Only through such coordinated efforts can the EU ensure that its ambitious regulatory framework for AI governance is implemented effectively across all member states, regardless of size or resources.

4. Comparative Perspectives and Pathways Forward

4.1 The Need for Agile and Inclusive Regulation

The rapid evolution of artificial intelligence demands a regulatory approach that is both adaptive and participatory. Scholars and policy institutions increasingly argue for governance

frameworks capable of responding to emerging risks without inhibiting innovation. The EU's risk-based model, exemplified by the Artificial Intelligence Act, aims to calibrate obligations according to the potential harm posed by specific systems. However, this approach often underestimates systemic risks associated with so-called "low-risk" technologies, particularly in public discourse and democratic processes.

In contrast, Brazil's General Data Protection Law (LGPD) incorporates a more participatory and impact-focused framework. Through its requirement for Algorithmic Impact Assessments in the public sector, the LGPD mandates evaluations of potential bias and discriminatory outcomes before the deployment of AI systems. This pre-emptive scrutiny enables greater alignment with constitutional values such as equality and non-discrimination.[17] Assunção analyzes how this model has been applied across Latin America, highlighting both its effectiveness and the political challenges involved in ensuring compliance in countries with uneven institutional capacities.[18]

Beyond national frameworks, the Organisation for Economic Co-operation and Development (OECD) has advanced a set of AI Principles that promote human-centred, transparent and accountable AI. These principles offer a foundation for global harmonization, especially in areas where synthetic media and generative models transcend borders. By encouraging interoperable standards - such as cross-jurisdictional labelling requirements for AI-generated content - the OECD framework helps address fragmentation and facilitate mutual recognition of regulatory safeguards.[19]

4.2 Lessons from Comparative Jurisdictions

International responses to AI-related risks reveal valuable lessons for the EU's regulatory trajectory. As mentioned previously, Brazil's emphasis on algorithmic transparency exemplifies a rights-based approach that goes beyond mere risk classification. Meanwhile, New Zealand has taken bold steps by imposing judicial sanctions on legal professionals who misuse generative AI, thus reinforcing the norm of professional responsibility.[20]

[17] Brussino & Alonso (2021) discuss the emergence of Algorithmic Impact Assessments in Latin America as a tool for safeguarding constitutional rights in digital governance.
[18] Assunção (2025) explores the implementation of Brazil's LGPD and its emphasis on algorithmic accountability in the public sector, arguing that this model offers a scalable alternative to the EU's risk-based framework.
[19] Brussino & Alonso (2021) point to the OECD AI Principles as a benchmark for creating interoperable legal standards, particularly in domains like synthetic media regulation.
[20] Assunção (2025) highlights the role of judicial enforcement in deterring professional misconduct involving AI tools, using New Zealand's legal sector as a case study.

These developments signal a broader shift toward accountability mechanisms that are both sector-specific and context-sensitive. They also underscore the importance of including civil society, professional bodies and affected communities in the co-creation of AI governance norms. Participatory structures not only increase legitimacy but also enable regulations to remain agile in the face of technological disruption.

5. Toward a Resilient Democratic Framework: Integrating Legal, Institutional, and Civic Responses

The legal challenges posed by AI-driven populism necessitate a multifaceted response that integrates regulatory reform, institutional capacity-building, and civic empowerment. No single intervention can adequately address the complex interplay of technological, political, and epistemic factors that enable algorithmic populism. Instead, a resilient democratic framework must operate simultaneously across multiple domains, each reinforcing the others.

Strengthening Accountability: From Product Liability to Algorithmic Responsibility

First, strengthening accountability requires comprehensive revision of liability frameworks to address the unique challenges posed by autonomous and generative AI systems. The existing Product Liability Directive, designed for physical products with clear chains of causation, is ill-suited to algorithmic harms that emerge from distributed systems, emergent behaviors, and the interaction of multiple actors. The recently adopted AI Liability Directive represents a significant step forward by introducing rebuttable presumptions of fault and facilitating access to evidence in AI-related litigation. However, its application to political contexts—particularly cases involving disinformation, deepfakes, and electoral manipulation—remains untested.

Legal frameworks must explicitly recognize that AI developers, platform operators, and political actors who deploy these technologies bear responsibility for foreseeable harms, including those that undermine democratic processes. This may require sector-specific liability regimes that account for the heightened public interest in electoral integrity and political communication. For instance, platforms that host political advertising could be held to stricter standards of due diligence regarding the authenticity of content and the transparency of targeting practices. Similarly, developers of generative AI models used to create synthetic media could face liability when their tools are deployed for electoral manipulation, particularly if they fail to implement reasonable safeguards such as watermarking or provenance tracking.

The challenge lies in calibrating liability rules to incentivize responsible innovation while avoiding chilling effects on legitimate speech and technological development. Drawing on comparative experiences, jurisdictions such as New Zealand have imposed professional sanctions on legal practitioners who misuse AI tools, demonstrating that accountability can be enforced without stifling innovation. The EU should consider analogous approaches for political actors and campaign operatives who knowingly deploy AI-generated disinformation.

Enhancing Transparency: Mandating Disclosure and Enabling Scrutiny

Second, enhancing transparency demands both technical and legal interventions. Mandating labeling of AI-generated content, particularly deepfakes and synthetic media used in political communication, is essential to preserving citizens' epistemic autonomy. Such requirements should be harmonized across jurisdictions to prevent regulatory arbitrage and should be accompanied by technical standards for provenance tracking and content authentication. The Coalition for Content Provenance and Authenticity (C2PA), an industry-led initiative, offers a promising model for embedding cryptographic metadata in digital content to verify its origin and modification history. However, voluntary adoption has been slow, and regulatory mandates may be necessary to achieve widespread implementation.

Additionally, algorithmic audits must become routine practice for platforms whose recommendation systems shape political discourse. These audits should be conducted by independent third parties with technical expertise and should assess not only compliance with formal regulations but also the substantive impacts of algorithmic curation on democratic deliberation, including effects on polarization, filter bubbles, and the amplification of extremist content. The EU's Digital Services Act provides a foundation for such audits, but implementation has been uneven, and enforcement mechanisms remain weak. Strengthening these provisions—through clearer audit standards, public reporting requirements, and meaningful penalties for non-compliance—is essential.

Transparency must also extend to political microtargeting practices. Current regulations, including the GDPR and the DSA, impose some constraints on data-driven political advertising, but significant gaps remain. Voters often have no meaningful way to know why they are being targeted with specific messages, what data profiles underpin those decisions, or who is funding the campaigns. Comprehensive transparency registries, modeled on initiatives such as the EU Transparency Register and enhanced by real-time disclosure requirements,

could provide citizens and researchers with the information necessary to scrutinize political communication practices.

Promoting Digital Literacy: Cultivating Civic Resilience

Third, promoting digital literacy requires sustained investment in educational programs that equip citizens to critically evaluate algorithmically curated information. This is not merely a technical skill but a democratic competency essential to informed participation in contemporary public life. Programs should be integrated into formal education systems, as in Estonia and Finland, where digital citizenship is taught from primary school onward. These curricula should cover not only basic digital skills but also critical media literacy, understanding of algorithmic curation, recognition of synthetic media, and awareness of psychological manipulation techniques.

Adult education is equally important, particularly for populations disproportionately targeted by disinformation. Community-based initiatives, partnerships with civil society organizations, and public awareness campaigns can reach citizens who are no longer in formal education. Special attention must be directed toward vulnerable populations, including rural communities, elderly citizens, and linguistic minorities, who may face heightened exposure to disinformation combined with lower levels of digital literacy.

Digital literacy initiatives should be evidence-based, continuously evaluated for effectiveness, and adapted to evolving technological threats. Research from Finland's fact-checking and media literacy programs demonstrates that sustained, curriculum-integrated approaches produce measurable improvements in citizens' ability to identify misinformation and resist manipulation. Such programs should be scaled across the EU, with funding and technical support provided to member states that lack the resources to develop them independently.

Integrating Legal, Institutional, and Civic Dimensions

Importantly, these three pillars—accountability, transparency, and digital literacy—are mutually reinforcing. Transparency enables accountability by making algorithmic harms visible and attributable. Digital literacy empowers citizens to demand both transparency and accountability from platforms and political actors. Accountability mechanisms, in turn, create incentives for platforms to be more transparent and for political actors to refrain from manipulative practices. A resilient democratic framework must therefore pursue these objectives in concert, recognizing that partial or siloed interventions are unlikely to succeed.

V. Recommendations and Conclusions

The intersection of artificial intelligence (AI) and digital populism represents a profound and multifaceted challenge to the structural integrity of democratic governance. As algorithmically amplified populist narratives exploit the vulnerabilities of digital communication infrastructures, urgent legal and institutional reforms are imperative to mitigate these emergent risks. Drawing upon recent case studies, comparative regulatory frameworks, and interdisciplinary scholarship, this section articulates a set of policy recommendations designed to counteract the pernicious effects of algorithmic populism while reinforcing the normative pillars of liberal democracy.

1. Legal Reforms to Address AI-Induced Populist Threats

1.1 Mandatory Impact Assessments for Political AI Systems

To ensure a proactive approach to the democratic risks posed by AI, it is essential to mandate comprehensive impact assessments for AI systems deployed within political and electoral contexts. These assessments should not merely evaluate technical parameters but must scrutinise potential adverse effects on democratic integrity, including the amplification of disinformation, algorithmic bias, political polarisation, and the erosion of public trust. The European Union's Artificial Intelligence Act (AIA) currently provides a framework for assessing high-risk AI applications; however, it omits "limited-risk" systems, such as generative models used for political communication and voter targeting.

To close this regulatory lacuna, the AIA ought to be expanded to encompass all AI applications with plausible political consequences. This extension would resonate with the Brookings Institution's advocacy for a "layered governance" model, which emphasises regulatory adaptability in the face of rapidly evolving technological capabilities. The experience of the 2024 United States primary elections, wherein AI-generated deepfakes misled voters through fabricated political messaging, underscores the necessity of pre-deployment scrutiny. Such assessments could have imposed obligations on developers to incorporate mechanisms for content verification, provenance tracking, or disclaimers for synthetic media.

Furthermore, Brazil's implementation of Algorithmic Impact Assessments (AIA) for federal agencies serves as an illustrative precedent, demonstrating the efficacy of structured evaluation processes in identifying and curbing biases embedded in automated decision-making tools.

Although jurisdictionally confined, the Brazilian model exemplifies a regulatory architecture capable of fostering algorithmic accountability.

1.2 Transparency Obligations: Labelling Requirements and Algorithmic Audits

The principle of transparency constitutes a cornerstone in the normative framework for AI governance. In the specific context of political communication, mandatory labelling of AI-generated content, whether in the form of synthetic video, text, or automated social media posts, would materially enhance the epistemic autonomy of citizens. While the AIA imposes certain transparency duties on providers of chatbots and recommender systems, inconsistent enforcement and the absence of a harmonised labelling standard weaken its efficacy.

Platforms such as TikTok and Facebook, which have been repeatedly implicated in the dissemination of manipulated content, must be subject to uniform labelling protocols. The OECD AI Principles offer a viable basis for transnational harmonisation, urging jurisdictions to adopt interoperable standards for transparency and disclosure.

Equally essential are mandatory algorithmic audits, particularly for platforms whose recommender systems operate as de facto curators of political discourse. These audits should be conducted by independent, third-party experts with full access to relevant data and models. The audit process must not only assess compliance with formal legal requirements but also evaluate the substantive impact of algorithmic systems on democratic deliberation, including their role in amplifying extremist content, fostering polarisation, and creating filter bubbles. The Digital Services Act (DSA) provides a foundational framework for such audits, but its provisions must be strengthened through clearer standards, public reporting obligations, and meaningful penalties for non-compliance.

1.3 Strengthening Accountability and Liability Frameworks

Revising liability frameworks to address AI-induced harms is a critical component of a resilient democratic response. The EU's AI Liability Directive, which introduces a rebuttable presumption of causality in cases of AI-related harm, represents a significant step forward. However, its application to the nuanced and often indirect harms of political disinformation requires further clarification. Legal frameworks must explicitly recognise that developers, deployers, and political actors who use AI systems to manipulate public opinion can be held accountable for the resulting damage to democratic processes.

This may necessitate the creation of sector-specific liability regimes for political communication, imposing a heightened duty of care on platforms and campaign organisations.

Drawing inspiration from New Zealand's approach, where legal professionals face sanctions for misusing AI, the EU could develop a code of conduct for political actors, with clear penalties for the deliberate deployment of deceptive AI technologies during election campaigns.

2. Institutional Strengthening to Foster Democratic Resilience

2.1 Digital Literacy as a Democratic Imperative

A legally robust response to algorithmic populism must be complemented by institutional investments in digital literacy. Citizens must be empowered to understand, interrogate, and resist manipulative algorithmic content. Digital literacy programmes should focus on the identification of synthetic media, recognition of algorithmic bias, and cultivation of critical media consumption skills. Brazil's General Data Protection Law (LGPD), which mandates public education on data protection rights, provides a foundational model that could be adapted to encompass AI-specific competencies.

Estonia's integration of digital citizenship education into school curricula also exemplifies a proactive approach to cultivating a digitally resilient electorate. Special attention should be directed towards communities disproportionately targeted by disinformation campaigns. In the Indian context, AI-generated voice calls disseminated false narratives to rural voters, demonstrating the necessity for culturally and linguistically localised digital literacy initiatives. These efforts should be state-supported, community-based, and continuously evaluated for efficacy.

2.2 Cross-Jurisdictional Coordination and Regulatory Convergence

Given the transnational dynamics of AI-driven populism, unilateral regulatory efforts are structurally inadequate. Effective governance in this domain requires robust cross-border coordination to deter regulatory arbitrage and uphold democratic norms in digital spaces. Institutions such as the OECD and the EU–US Trade and Technology Council serve as platforms for developing joint standards and interoperable enforcement mechanisms capable of responding to algorithmic manipulation at scale.

Priority areas for international collaboration should include:

- **Extraterritorial compliance mechanisms**, ensuring that AI developers and platforms based outside the EU adhere to obligations under the AIA and the Digital Services Act.

Recent interventions by the European Commission in response to TikTok's failure to remove illegal content illustrate the urgency of operationalizing these powers in practice.

- **Shared incident databases** for AI-related harms, allowing regulators to identify patterns, anticipate risks and coordinate timely responses. A collective intelligence model—based on real-time data-sharing and co-analysis—could enhance the effectiveness of risk mitigation across jurisdictions.
- **Interoperability of regulatory models**, bridging the EU's risk-based approach with sector-specific regimes such as the UK's pro-innovation framework or Brazil's LGPD. This alignment is essential to prevent fragmentation and ensure legal certainty for developers and users alike.

However, achieving true convergence also requires targeted support for member states with limited technical capacity. Smaller jurisdictions, such as Cyprus or Malta, may lack the institutional infrastructure needed to conduct independent algorithmic audits or enforce compliance robustly. In this context, the EU should establish algorithmic audit units funded at the supranational level to assist national regulators in reviewing high-risk AI deployments and disinformation campaigns.

Additionally, technical capacity-building programs, inspired by Estonia's national digital literacy initiatives, should be extended to lower-capacity member states, equipping civil servants, legal professionals and educators with the skills necessary to engage meaningfully with algorithmic governance. These efforts are not merely technocratic upgrades; they are constitutional investments in the resilience of democratic institutions.

3. Defending Democratic Sovereignty in the Algorithmic Age: Towards a New Social Contract

The convergence of artificial intelligence and digital populism does not merely pose new challenges to democratic institutions, it compels a fundamental redefinition of sovereignty and political agency in the algorithmic era. Traditional models of popular sovereignty, grounded in deliberative transparency and participatory legitimacy, are increasingly displaced by opaque algorithmic systems that mediate political information, structure discourse, and subtly reconfigure collective will.

This reality calls for the articulation of a new social contract for algorithmic governance, a normative framework capable of embedding democratic values directly into the design, deployment, and oversight of AI systems. Rather than treating technology as an externality to

be regulated post hoc, such a contract must assert the principle of "legal-by-design," ensuring that AI architectures are developed in alignment with fundamental rights and constitutional safeguards from the outset.[21]

Central to this democratic reconfiguration is the recognition that algorithmic systems are not neutral. They reflect and reproduce societal asymmetries unless consciously shaped otherwise. Participatory governance structures, especially those that incorporate the voices of marginalized communities disproportionately affected by disinformation and algorithmic harms, are essential to restoring democratic agency in digital contexts.[22]

This involves inclusive design processes where diverse stakeholders co-create the normative parameters guiding AI development, ensuring that technologies serve the public interest rather than undermining it. Drawing on Zuboff's critique of surveillance capitalism, this new social contract must institutionalize mechanisms for algorithmic accountability, public transparency, and civic empowerment. These include citizens' assemblies on AI ethics, independent oversight bodies with investigatory powers, and open-source frameworks for auditing algorithmic decisions in political contexts.

Defending democracy in the digital age also entails anticipating the evolving risks posed by AI-powered disinformation, microtargeting, and epistemic fragmentation. Agile, adaptive regulatory regimes, responsive to technological innovation and capable of enforcing democratic constraints across both public and private actors are indispensable. But so too are civic infrastructures that foster digital literacy, epistemic resilience, and participatory deliberation, cultivating a public capable not only of resisting manipulation but of shaping the ethical trajectories of technological progress.

Ultimately, safeguarding democratic sovereignty in an algorithmic world requires more than institutional reform. It demands a transformative political vision - one that reclaims technological development as a democratic project and ensures that artificial intelligence becomes a vector of empowerment rather than a mechanism of control. Only by embedding democratic values into the very logic of AI systems can societies uphold the foundational promise of self-government in the 21st century.

[21] Hildebrandt (2020) emphasizes the importance of "legal-by-design" principles to align AI systems with human rights and constitutional safeguards from the start.
[22] Momoc (2018) argues for participatory design processes in AI governance that include marginalized groups disproportionately impacted by algorithmic disinformation and disinformation and harms.

References

Aslanidis, P. (2016). Is populism an ideology? A refutation and a new perspective. *Political Studies*, 64, 88–104.

Assunção, A. (2025, January). *Populismo digital e populismo digital extremista*. Paper presented at the 22º Encontro Nacional de Pesquisadores em Jornalismo, Vol. 22, 2024.

Azgin, B., & Kiralp, S. (2024). Surveillance, disinformation, and legislative measures in the 21st century: AI, social media, and the future of democracies. *Social Sciences*, 13(10), 510.

Bartlett, J., Birdwell, J., & Littler, M. (2011). *The new face of digital populism*. Demos.

Battista, D., & Mangone, E. (2025). Technological culture and politics: Artificial intelligence as the new frontier of political communication. *Societies*, 15(4), 75.

Brundage, M., et al. (2018). *The malicious use of artificial intelligence*. arXiv. https://arxiv.org/abs/1802.07228

Brussino, S., & Alonso, D. (2021). Citizens and democracy: Political legitimacy processes in Latin American democracies. In C. Zúñiga & W. López-Lópe (Eds.), *Political psychology in Latin America* (pp. 11–34). American Psychological Association.

Bruzzone, A. (2021). *Ciberpopulismo: Política e democracia no mundo digital*. Editora Contexto.

Citron, D. K., & Pasquale, F. (2014). The scored society. *Washington Law Review*, 89(1).

de la Torre, C. (2018). Populism revived: Donald Trump and the Latin American leftist populists. *The Americas*, 75, 733–753.

Durham, L. (2009). *Punishing the poor: The neoliberal government of social insecurity*. Duke University Press.

European Commission. (2023). *Ethics guidelines for trustworthy AI*.

Hildebrandt, M. (2020). *AI for the rule of law?* Springer.

Howard, P. N., et al. (2018). *Junk news and bots during the U.S. election*. Oxford Internet Institute.

Mittelstadt, B., et al. (2016). The ethics of algorithms. *Big Data & Society*.

Momoc, A. (2018). Populism 2.0, digital democracy and the new ‘enemies of the people’. *Communication Today*, 9(1), 58–77.

Monti, M. (2021). *Online disinformation and digital populism in the EU and the USA: Western constitutionalism at the crossroads of internet regulation*. STALS Research Paper, 1.

Mudde, C. (2004). The populist zeitgeist. *Journal of International Affairs*, 58(1).

Prior, H. (2021). *Digital populism and disinformation in «post-truth» times*.

Rădulescu, B.-G. (2024). The threat of algorithmic populism: Intelligence strategies for safeguarding democracy. *Intelligence Info*, 4(1), 14–26.

Raji, I., et al. (2020). *Closing the AI accountability gap*. ACM Digital Library.

Stanley, B. (2008). The thin ideology of populism. *Journal of Political Ideologies*, 13(1), 95–110.

Velez, T. (2022). *Populismo e media: O impacto das atitudes populistas no consumo de media em Portugal* (Master's thesis, ISCTE-Instituto Universitário de Lisboa).

Wacquant, L. (2009). *Punishing the poor: The neoliberal government of social insecurity*. Duke University Press.

Weyland, K. (1999). Neoliberal populism in Latin America and eastern Europe. *Comparative Politics*, 31(4), 379–401.

Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.