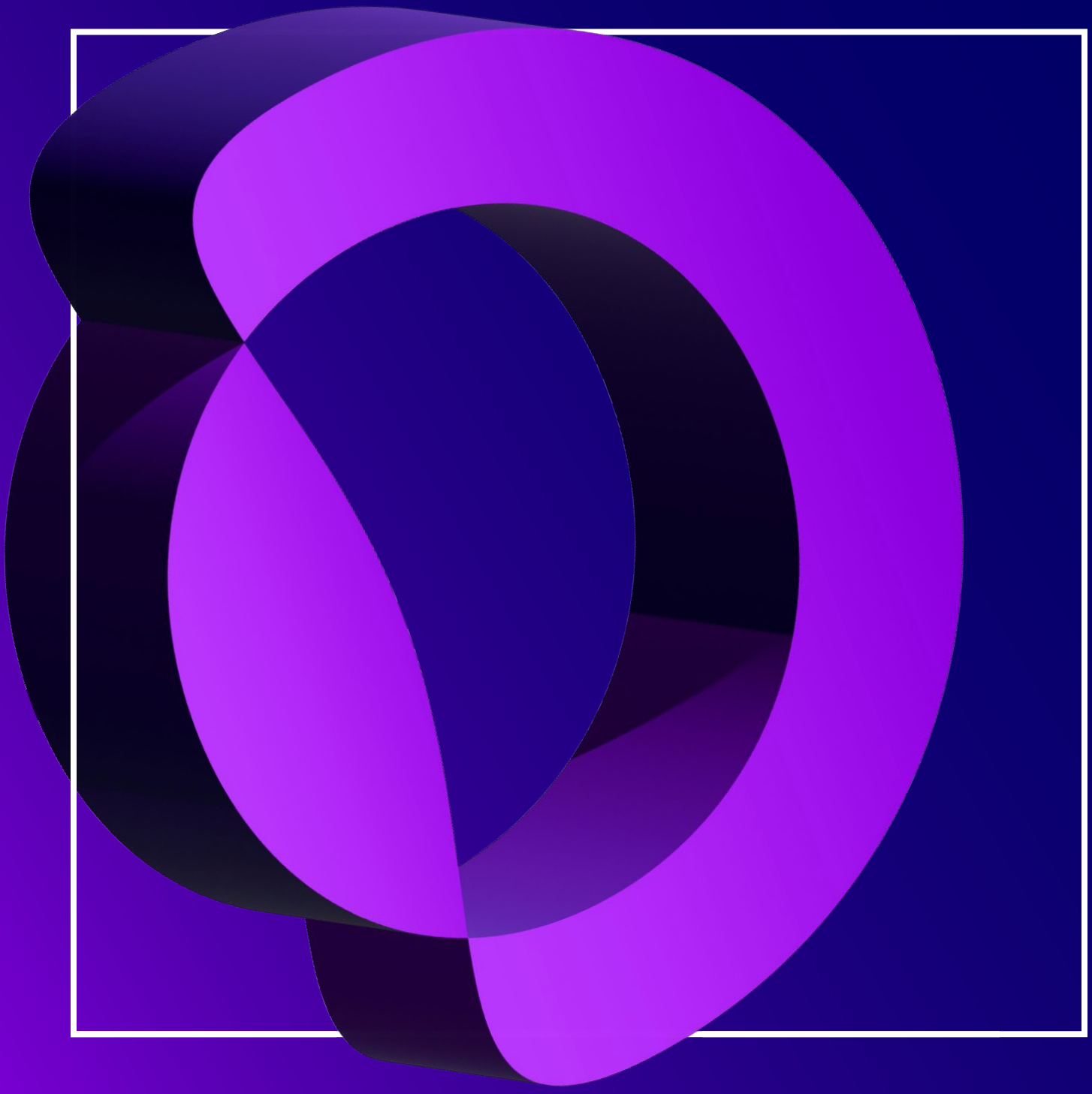
ie UNIVERSITY
CENTER FOR THE GOVERNANCE OF CHANGE

AI4 DEMOCRACY



DEMOCRACY RELOADED:
AI TO PROTECT AND PROMOTE DEMOCRATIC GOVERNANCE

JANUARY 2025

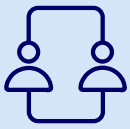
CONTENTS

WRITTEN BY
PRIMAVERA DE FILIPPI

AI4DEMOCRACY

The power of complex AI systems holds great promise to protect democracies against attacks and to make democratic processes more effective and participatory. AI4Democracy is a global research initiative to realize this promise. It is led by the Center for the Governance of Change at IE University, with Microsoft as strategic partner. AI4Democracy seeks to harness AI to defend and strengthen democracy through coalition-building, advocacy and intellectual leadership.

AI4DEMOCRACY IS COMPOSED OF TWO TRACKS:



- **AI4Democracy Action Coaalition:** aims at securing alliances with aligned international organizations and democratic governments to advocate tangible policy action. It has entailed the organization of high-level events and the participation in the key 2024 gatherings that have shaped the AI policy agenda to drive forward our recommendations.

- **Democracy-affirming AI intellectual leadership**: four policy papers have been produced by global AI experts to provide the academic and intellectual foundation for the positions of the Coalition. This research has driven the global conversation on AI and democracy and advanced specific, action-oriented policy recommendations for democratic governments and others.

  AI4Democracy is the continuation of Tech4Democracy—a global initiative by IE University, in partnership with the U.S. Department of State and with the strategic support of Microsoft, to study and promote democracy-affirming technologies worldwide.

RESEARCH:

#1

How AI can be used to inform policymaking **(June 2024), Deger Turan and Colleen McKenzie (AI Objectives Institute):**
This paper analyzes different paradigms under which policy development takes place and illustrate with case studies from 2023 and 2024 how AI tools have augmented civic capacity. It showcases AI's potential to support collective agency in ways that systematically feed back into AI governance and AI safety institutions, creating a virtuous circle of improving AI's impact on society.

#2

Depolarizing and moderating social media with AI **(July 2024), Pedro Ramaciotti (Sciences Po and CNRS):**
This paper proposes AI tools and guidelines for the enhancement of social media ecosystems, outlining concrete actions to improve compliance and moderation of the digital space. It explores the potential to provide platforms, regulators and researchers with a new framework for AI development that reconciles societal and business objectives.

#3

Enhancing legislative engagement with AI **(September 2024), Nathan Sanders and Matthew Victor (MAPLE) and Alicia Combaz and David Mas (Make.org):**
This paper explores specific ways in which AI can be used to increase engagement in the legislative process, making residents more informed about and active in policymaking while simultaneously making legislators more responsive and connected to their constituents. It outlines what challenges need to be overcome to deploy these tools equitably.

#4

Securing democratic infrastructures **(October 2024), Andrew Dwyer (Royal Holloway, University of London) and Roxana Radu (University of Oxford):**
This paper examines how AI can enhance the security of the materials and processes that enable democratic societies to function well. It focuses on two of these: parliamentary and electoral systems. For each, this paper explores how AI offers distinct advantages to protect our collective democratic infrastructures from adversarial attacks that seek to undermine democratic societies.

RESEARCH DIRECTOR

- **Primavera de Filippi**, Director of Research at the National Center of Scientific Research (CNRS) and Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard University

TEAM OF RESEARCHERS

- **Andrew Dwyer**, Lead of the UK Offensive Cyber Working Group, Associate Research Fellow at the Research Institute for Sociotechnical Cyber Security (RISCS) and Lecturer in Information Security at the university of London
- **Roxana Radu**, Associate professor of Digital Technologies and Public Policy at the Blavatnik School of Government, University of Oxford
- **Deger Turan**, President of AI Objectives Institute and CEO of Metaculus. Former CEO and Founder of Cerebra Technologies
- **Colleen McKenzie**, Executive Director at AI Objectives Institute and Co-founder of the Median Group
- **Pedro Ramaciotti**, Chair of AI in Social Sciences and Humanities at the National Center of Scientific Research (CNRS). Lead of the European Polarization Observatory of the CIVICA Consortium of European Universities in Social Sciences
- **Nathan Sanders**, Associate Editor of the Harvard Data Science Review and Member of the Board of Directors of the American Institute of Physics
- **Matthew Victor**, Associate at Bernstein Shur
- **David Mas**, Chief AI Officer at Make.org
- **Alicia Combaz**, Founder and CEO of Make.org

MEMBERS OF AI4DEMOCRACY ADVISORY BOARD

- **Sedef Akinli Kocak**, Director of AI Professional Development at the Vector Institute
- **Irene Blazquez Navarro**, Director of IE University Center for the Governance of Change
- **Nikki Freeman, Director**, AI Product Strategy & Partnerships of the Office of the CTO, at Microsoft
- **Dario García de Viedma**, Associate Director of IE University Center for the Governance of Change
- **Carlos Luca de Tena**, Executive Director of IE University Center for the Governance of Change
- **Gianluca Misuraca**, Founder and Vice President on Technology Diplomacy of Inspiring Futures
- **Manuel Muñiz**, Provost of IE University and Chair of its Center for the Governance of Change
- **Alex Roche**, Associate Director of IE University Center for the Governance of Change
- **Carlos Santiso**, Head of Division—Digital, Innovative and Open Government, OCDE
- **Ravi Shankar Chaturvedi**, Director of Research at Fletcher's Institute for Business in the Global Context (IBGC), Tufts University
- **Lena Slachmuijlder**, Executive Director of Digital Peacebuilding at Search for Common Ground
- **Jordan Usdan**, Sr. Director Strategy and Innovation, Office of the CTO at Microsoft

1. INTRODUCTION

1. INTRODUCTION

As technology is becoming an integral part of many people's lives, democracies around the world are facing new challenges. Declining voter turnout, diminished trust in institutions, growing political polarization, and the progressive weakening of civic engagement are only a few signals that are indicative of a crisis in political participation and democratic governance.

Citizens are often disengaged from the political process outside of elections, and meaningful public deliberation is becoming increasingly rare in light of disinformation, misinformation, and social media manipulation. These challenges are exacerbated by the globalized nature of information flows, which can be exploited by adversarial actors to undermine democratic processes. As a result, traditional democratic institutions, designed for the pre-Internet era, are struggling to maintain legitimacy and efficacy in the digital world.

Many of these challenges predate the development of artificial intelligence (AI). Yet, the advent of AI can significantly impact these systemic issues, bringing a whole new set of promises and perils. Indeed, when it comes to democracy, AI can be regarded as a double-edged sword: while it presents opportunities to safeguard and enhance democratic governance, it also introduces new risks.

On the one hand, if left unchecked, AI could exacerbate existing problems within democracies. The use of AI for disinformation campaigns and voter manipulation could undermine democratic values and further erode trust between citizens and governments. AI systems that are not transparent or accountable risk concentrating power in the

hands of a few public or private actors, exacerbating inequalities and sidelining marginalized groups. Additionally, the rise of "surveillance capitalism"—the commodification of personal data for profit—poses significant risks to privacy, especially when AI tools are used to monitor citizens under the guise of security.

On the other hand, if properly deployed, AI could enhance democratic governance by improving citizen engagement, facilitating more informed policy-making, and protecting democratic institutions against malicious attacks. For instance, AI can be employed to streamline public service delivery, detect disinformation, and even improve the quality of political discourse by analyzing public feedback and synthesizing diverse perspectives.

This report explores practical uses of AI for democracy, presenting innovative solutions while also raising critical questions about how to manage the potential downsides of these technologies. By examining the opportunities and challenges posed by AI, the report aims to provide a roadmap for reloading democracy in the digital age. Reloading democracy, in this context, means leveraging AI to fix the systemic issues of existing democratic structures, but only doing so in ways that protect and promote democratic values. Indeed, as AI gets infused into our governance structures, we must ensure that these technologies remain a tool for enhancing, rather than undermining, democratic systems.

As such, this report explores the dual role of AI through two complementary lenses: as **both a protector and promoter of democratic governance.** Protecting democracy involves leveraging AI to combat the growing threats of disinformation, adversarial attacks, and election interference, while ensuring online spaces are kept safe and trustworthy. Promoting democracy means using AI to create more inclusive, participatory governance systems by enhancing the effectiveness of policy making, leading to greater civic engagement and making governments more responsive to the will (and needs) of the people.

By addressing this dual role of AI, this report seeks to contribute to the ongoing debates about the future of governance in the digital era.

First, with democracies facing existential threats from adversaries who exploit digital platforms to spread falsehoods and catalyse discord, it is essential to explore how AI can be harnessed to safeguard public trust and the integrity of electoral processes. AI-driven tools for identifying and combating disinformation, securing digital infrastructures, and ensuring election integrity have become crucial to promote meaningful deliberation and preserve democratic stability.

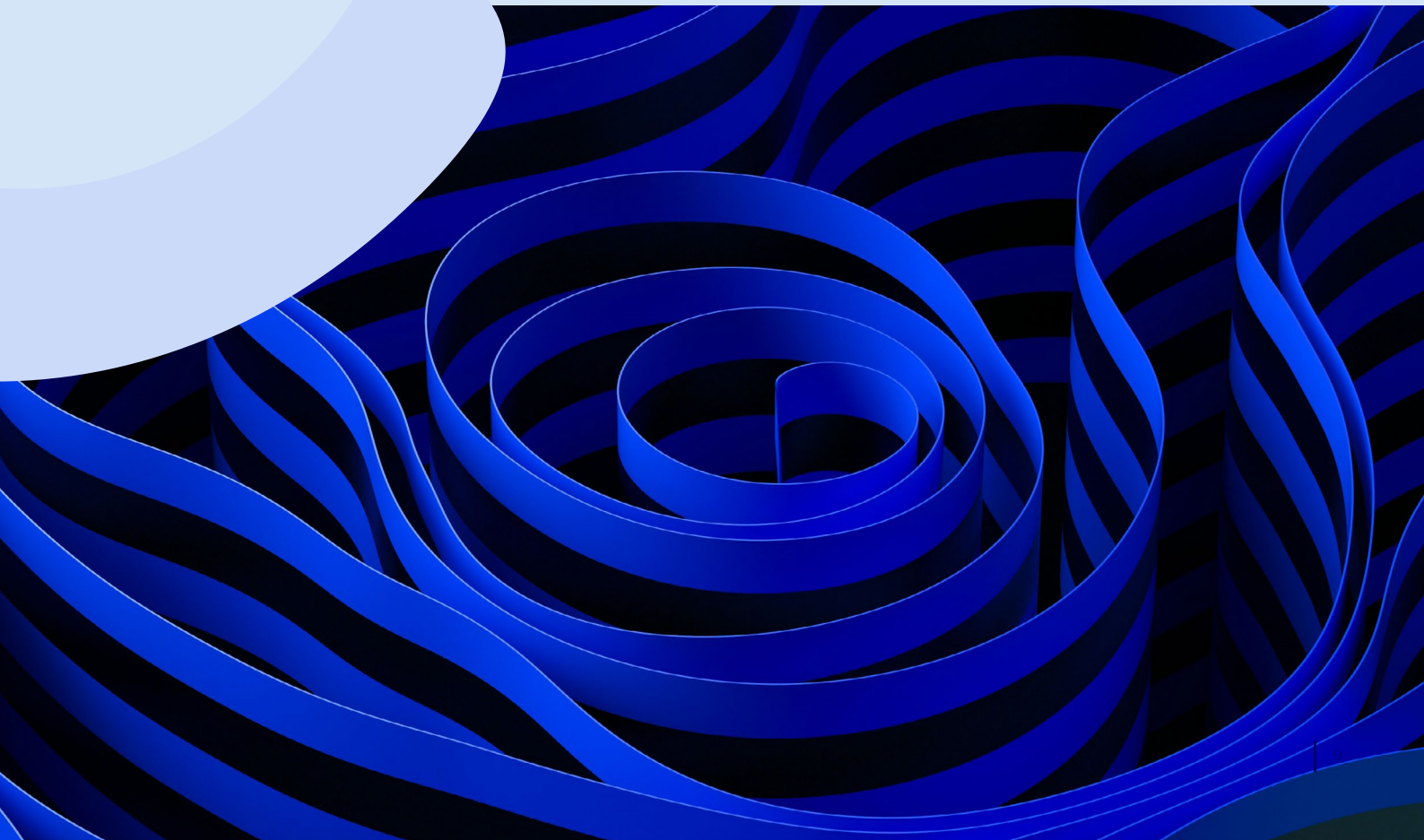
Second, as populations (and inequalities) grow, governance becomes more complex. AI can help process vast amounts of citizen input and create mechanisms for real-time citizen engagement in decision-making processes. In this way, AI offers new avenues to improve civic engagement and democratic participation, especially in countries where voter turnout and public trust in government have eroded.

Hence, **protection** and **promotion** are two essential components for democracy to thrive in the 21st century. Without robust defense mechanisms, democracies risk falling prey to the very same tools that were originally intended to support them. Without appropriate strategies to strengthen citizen engagement, democracies risk stagnating and becoming less representative of the people they serve.

By addressing these two complementary objectives, this report offers a framework for how AI can be integrated in existing political systems, in ways that not only preserve but enhance democratic institutions. Building upon the findings from our previous research papers (**RAMACIOTTI, DWYER, TURAN** and **SANDERS**), the report

critically examines the synergies and tensions that emerge from these studies, questioning the extent to which AI can deliver on its promises. Indeed, despite the opportunities of AI, we must remain vigilant about the risks of bias, exclusion, and misuse that it might be subject to.

Ultimately, the report aims to provide a holistic analysis of AI's role in supporting democracy, highlighting actionable strategies for simultaneously protecting and promoting democratic institutions. Indeed, these two dimensions are not mutually exclusive but deeply intertwined. This dual role of AI—as both a defense mechanism and a tool for democratic renewal—sets the stage for a deeper exploration of how these technologies can be scaled, adopted by big tech and governments alike, and integrated into democratic frameworks in a way that promotes transparency, accountability, and inclusivity. To be sure, the future of AI and democracy is not predetermined, it will depend on the decisions and policies we put in place today to ensure that AI effectively serves the public good.



2. AI AS A DEFENDER OF DEMOCRACY

2. AI AS A DEFENDER OF DEMOCRACY

A. DISINFORMATION, CONTENT PROVENANCE AND POLARIZATION

AI has proven highly effective in detecting disinformation, primarily through natural language processing (NLP) and machine learning models that can analyze vast amounts of data to identify misleading content (Oshikawa & al. 2020; Villela & al. 2023).

For instance, Facebook uses machine learning algorithms to detect misinformation by analyzing user behavior, post content, and network patterns. These solutions are effective in flagging content that does not comply with the company's policies and identifying fake users. Indeed, according to Facebook's Community Standards Enforcement Report, in Q1 2024, the company took action on over 14 million violent and graphic content, and shut down 1.2 billion fake accounts. The company claims that the increase in the amount of content that violates the platform's policy is mainly due to the improvement of its AI detection systems, which have better capacity at detecting nudity, violent content, and hate speech. Yet, the system struggles with nuanced misinformation and language barriers, requiring human oversight to improve accuracy.

To address this issue, MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) developed the "Detect Fakes" project, combining different machine learning techniques to analyze both the content (e.g. linguistic features) and metadata (e.g. published and author information) of news articles in order to assess the likelihood that they qualify as fake news. The system provides an explanation for its evaluation, so that it can be reviewed by humans. Another approach to tackling disinformation with AI Google's ClaimReview schema: a standardized format for fact-checks that can be displayed in Google's search results. This standard makes it possible for AI to process and match claims with relevant fact-checks in order to establish their rank in the search results. This has led to over 1 billion fact-checks appearing in Google search results every year. Yet, this requires widespread participation from fact-checkers and their ability to keep up with large volumes of (mis)information.

Despite the relative success of these systems, the volume of content produced across social media platforms on a daily basis presents a significant challenge. AI algorithms may struggle to process the sheer magnitude of posts, tweets, videos, and other forms of media generated in real-time. Moreover, even if LLMs can process multiple languages, they have more difficulty distinguishing between genuine and synthetic data in less commonly used languages (**OECD 2023**). Finally, even when these systems can flag misleading content, the continuous evolution of misinformation strategies—ranging from deepfakes to more subtle, context-specific disinformation—requires constant retraining and updating of models, which is resource-intensive and difficult to scale up globally. This is precisely the problem that the European project **vera.ai** is trying to address.

Besides, if AI can assist in moderating content and flagging misinformation, it is not a replacement for media literacy efforts. Users shall learn to critically engage with information and identify misinformation. Solely relying on AI could lead to complacency, where users trust platforms to manage content without understanding the nuances of misinformation.

Verifying the provenance of content also plays a key role for tracing the dissemination of disinformation and identifying bad actors. In that regard, the Content Authenticity Initiative (CAI) developed an industry standard for provenance metadata that can be incorporated in digital content, and that can be easily processed by AI systems. Yet, the creators of disinformation have no incentives to incorporate metadata into their fake content, requiring a more comprehensive mechanism to track the origin of online content. This highlights the importance of verifying online identity while respecting privacy. AI-driven identity verification systems, while potentially useful for distinguishing between real users and bots, can raise privacy concerns by requiring access to personal data. In particular, the tension between securing online identities and maintaining anonymity—especially in contexts where political expression might expose users to repression—makes the design of such systems particularly sensitive.

Besides, many of the tools implemented thus far are focusing on using AI to detect fake news, obscene content, hate speech, harassment or incitement to violence. They help mitigate the bad content, but they do not contribute to elevating the good content. Online platforms continue to drive people apart with more and more filter bubbles and increased polarization of opinions. To address this issue, Google Jigsaw has developed the Perspective API, leveraging AI tools to identify high-quality content in online discussions. These tools evaluate posts based on virtues such as nuance, evidence-based reasoning, personal stories, and human compassion. By assigning a numerical score (ranging from 0 to 1) to each post, the AI determines how likely it is to reflect these positive features. The Perspective API can be used by online platforms to rank content not by popularity (likes or comments), but by the quality of discussion, promoting more thoughtful, compassionate, and constructive conversations in digital spaces.

RAMACIOTTI presents another way in which AI can help improve the social media landscape by tackling the issue of polarization. In his paper, **RAMACIOTTI** explores how the emergence of social media as a digital public space has positioned algorithms as central mediators in how content is filtered, curated, and presented to users. With the growing overlap between online platforms and offline political behavior, concerns over polarization and the erosion of democratic processes have intensified. **RAMACIOTTI** examines the role of AI in mitigating these issues, proposing the use of representation learning spaces—a form of AI commonly employed in various applications—to address political segregation and polarization on social media. By drawing a parallel between *spatial models of politics* (from political science) and *representation learning spaces* (from computer science), the paper explores the use of AI as a unique technical opportunity for depolarization through the development of tools that can identify and disentangle the computational elements contributing to political polarization. Ultimately, **RAMACIOTTI** argues that the use of AI could help create better mechanisms for regulating and designing social media platforms to promote healthier, less divisive public discourse.

However, while depolarization is key for rebuilding trust in information ecosystems, AI alone cannot restore the trust deficit in online information. Media literacy requires human-centered interventions — education, training, and community engagement— that complement AI's technical solutions. Without broader social efforts to teach users how to critically evaluate information, AI tools may fall short in combatting disinformation and polarization.

B. RESILIENCE AGAINST ADVERSARIAL ATTACKS

AI is a double-edged sword, with the capacity to both undermine democratic practices through the dissemination of disinformation and manipulation, while simultaneously enhancing electoral integrity and citizen engagement through advanced monitoring and participatory tools. In particular, **DWYER** outlines how AI can strengthen the resilience of democratic institutions by detecting and mitigating cyber threats, inviting governments to proactively adopt AI technologies into the electoral and representative systems to identify malicious activities that might undermine public trust or interfere with electoral processes.

For instance, as discussed above, AI-enabled monitoring systems can analyze patterns of behavior on social media platforms to identify and neutralize bot-driven disinformation campaigns. Notable examples include the 2016 U.S. presidential election, where Twitter bots were used to influence political conversations on social media, spreading disinformation and increasing polarizations (Bessi & Ferrara 2016; Cohane 2021). Since then, a variety of machine learning techniques have been developed to help detect and address social media bot activities aimed at spreading misinformation (Ellaky & Benabbou 2024).

The European Union (EU) is addressing the challenges posed by the use of AI systems in political campaigns through comprehensive regulation, particularly with the introduction of the EU AI Act, which classifies AI systems used to influence voters as “high-risk”. Similarly, Brazil addressed the issue of electoral integrity by formally banning the use of AI in municipal electoral campaigns and requiring that any use of artificial intelligence for electoral purposes be accompanied by a clear public notice, with penalties for candidates who violate these regulations, including potential disqualification from running for office or rescission of their mandates if elected.

In January 2024, the U.S. Cybersecurity and Infrastructure Security Agency (CISA) has issued a report highlighting how generative AI could impact the security and integrity of election infrastructure, as malicious actors—including foreign nation state actors and cybercriminals—could leverage these capabilities for nefarious purposes, such as generating fake news and disinformation campaigns, impersonating political candidates or election office staffs to gain access to sensitive information, creating deep-fakes to harass or attack election officials, generating fake voter calls to overwhelm call centers, etc. At the same time, the CISA has developed a Roadmap for Artificial Intelligence, has deployed AI systems to promote the beneficial uses of AI to enhance cybersecurity capabilities, and ensure that AI systems are protected from cyber-based threats. This includes using AI tools to analyze data from various sources to identify potential threats to electoral infrastructure and enhance the security of voting systems in the processes leading up to elections.

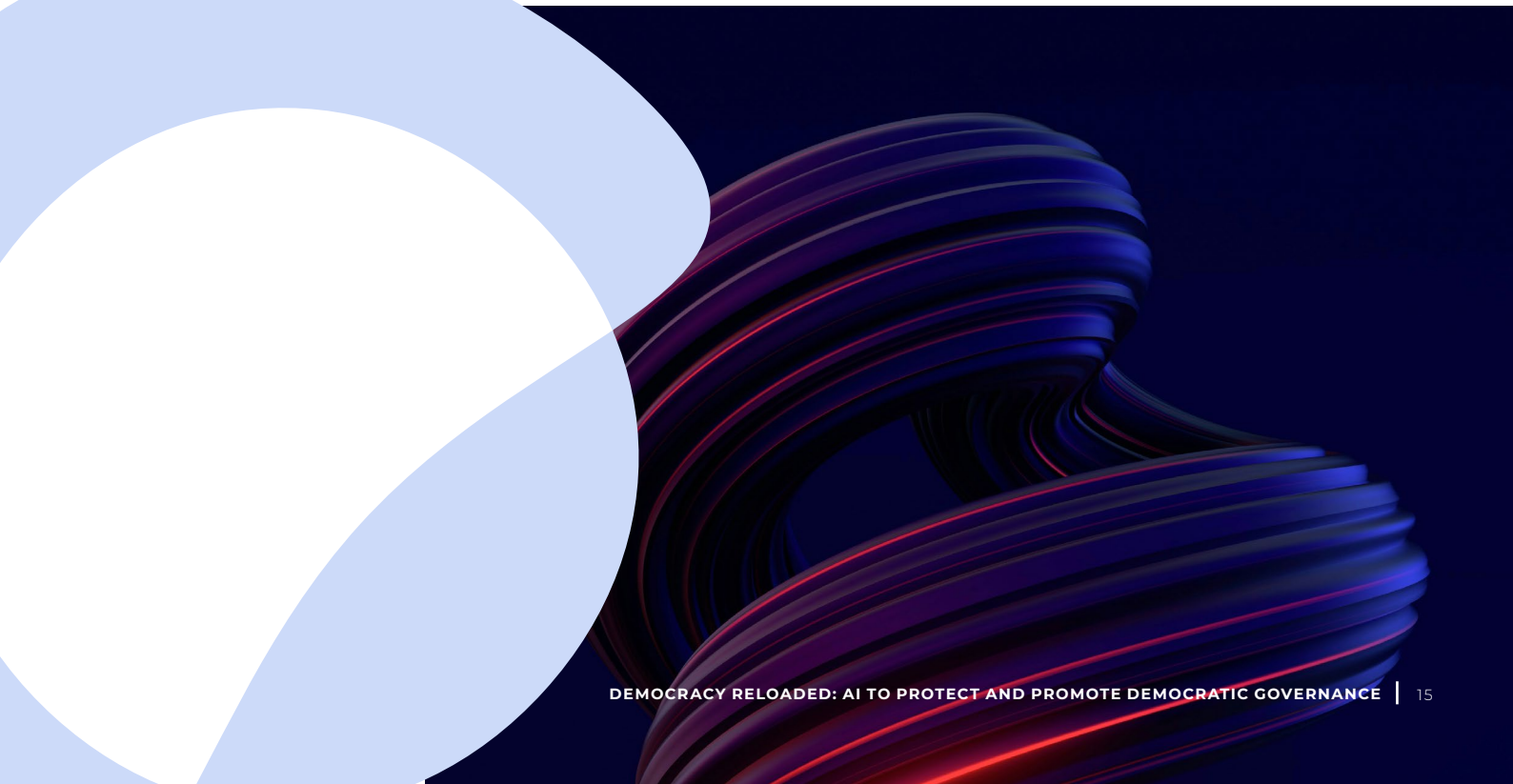
3. AI AS A TOOL TO STRENGTHEN DEMOCRACY

3. AI AS A TOOL TO STRENGTHEN DEMOCRACY

A. INFORMED POLICY MAKING

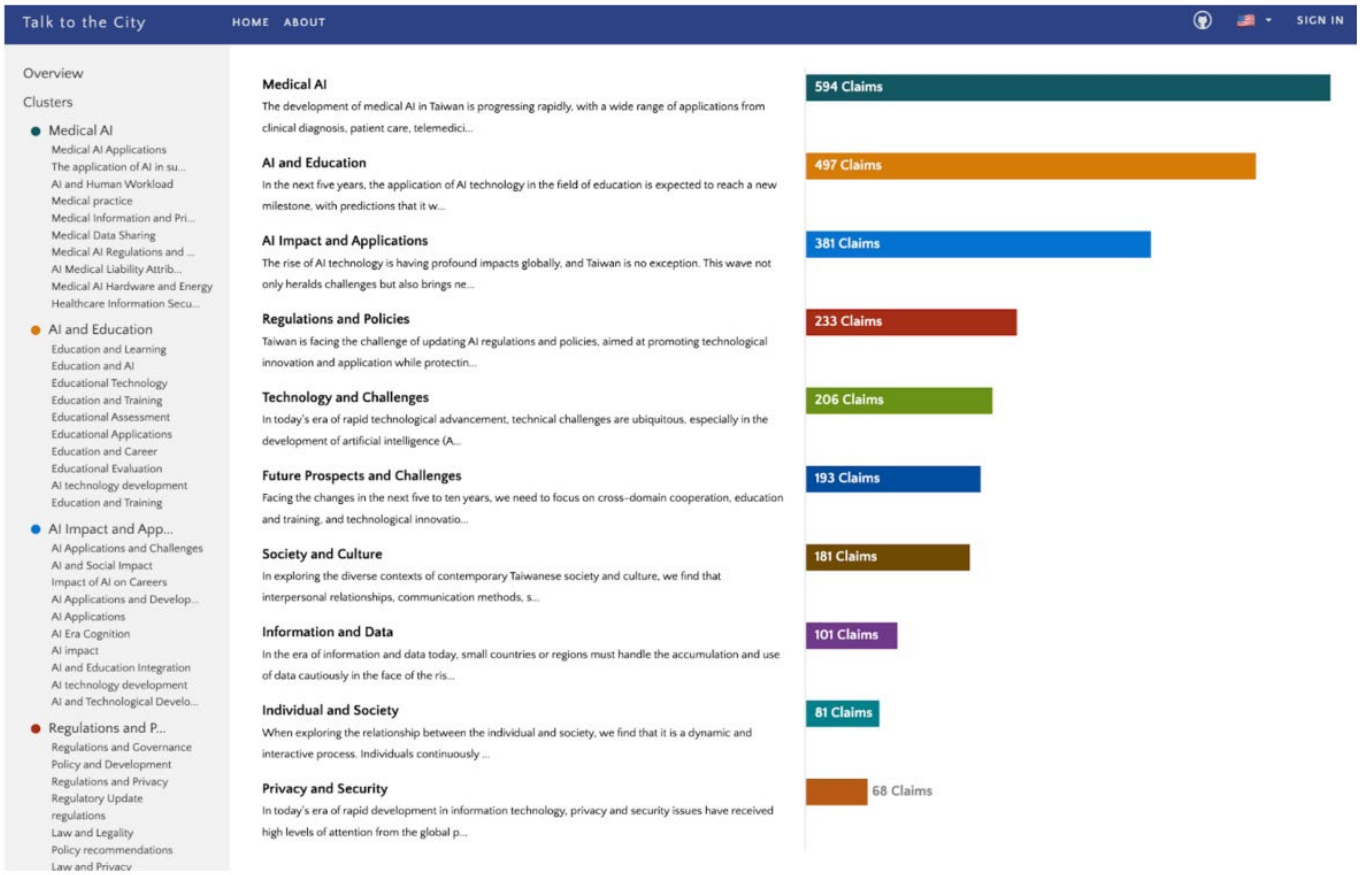
AI's potential to strengthen democratic governance hinges on its ability to act as a conduit between governments and citizens. As such, AI can significantly enhance the policymaking process by streamlining decision-making and improving access to information.

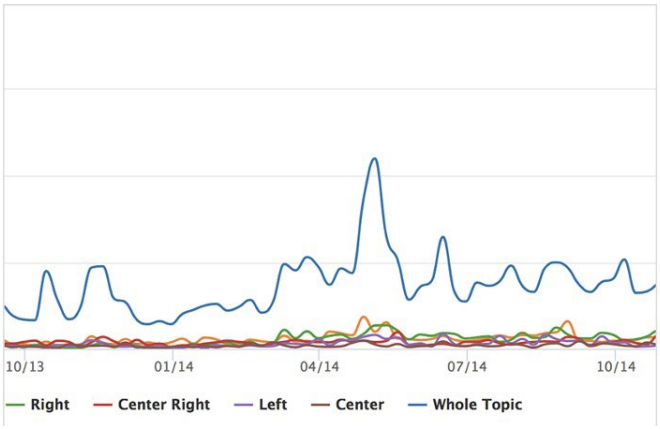
AI's data processing capabilities allow policymakers to detect emerging societal issues early on by analyzing vast datasets from diverse sources such as public opinion, social media, or economic indicators. Its ability to summarize complex problems efficiently can help identify trends or crises that might otherwise go unnoticed, thus enabling proactive policymaking. One of the most significant advancements in this regard involves leveraging AI in policy feedback systems, using AI tools to analyze and synthesize public opinion, transforming large volumes of qualitative inputs into actionable insights (UNESCO 2022).

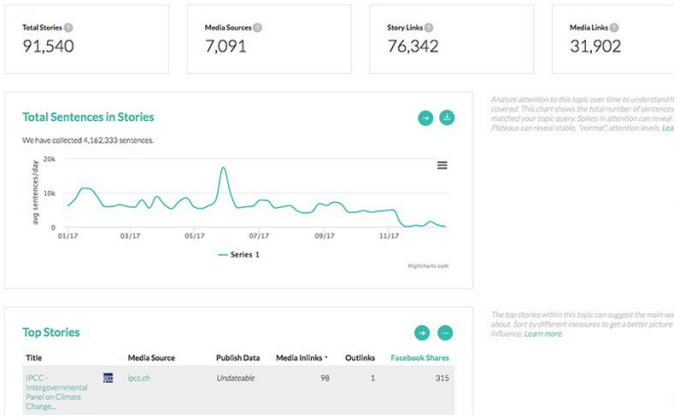


TURAN shows how AI-driven policy feedback systems can bridge the gap between citizens and governments by processing datasets that would otherwise be overwhelming to analyze. This makes it possible for decision-makers to fully tap into the collective intelligence of their constituents. Indeed, by distilling complex, diverse viewpoints into digestible, organized data, AI systems can help policymakers gain a deeper, more nuanced understanding of public needs and sentiments. To illustrate this point, **TURAN** uses the example of the **Talk to the City** (TttC) tool developed by the AI Objectives Institute, offering a glimpse into how AI can be used to support deliberative decision-making, by synthesizing large volumes of public opinion and producing actionable insights for policymakers. TttC aggregates input from various forms of public engagement—whether structured surveys or freeform content like interviews—and uses advanced clustering techniques to reveal patterns, common ground, and areas of polarization within a population's views, helping decision-makers understand the complexity of different perspectives. As such, TttC tool offers significant potential to create more targeted, responsive, and effective policies by providing policymakers with real-time, aggregated feedback from diverse populations. It also contributes to establishing a more inclusive and nuanced deliberative process, supporting well-informed and responsive decision-making that can address the needs and concerns of diverse stakeholders.

Similarly, Civic Signals is an initiative powered by MIT's Media Cloud platform, with research partners such as the *Reuters Institute for Journalism and the Global Disinformation Index*. It provides actionable insights to help policy-makers acquire a better understanding of Africa's media ecosystem and emerging civic technology sector. It provides a media explorer to get a quick overview of how a particular topic of interest is covered by digital news media, as well as source manager and a topic mapper to further dig into a particular issue at stake.







The EU-funded project AI4PublicPolicy provides another illustration of how AI can support policy-making, by facilitating the development of evidence-driven and data-driven policies. AI4PublicPolicy is a policy management environment using AI technologies to promote stakeholder engagement by processing and summarizing feedback from consultations, surveys, or public forums, helping policymakers gauge public sentiment more accurately.

More generally, with the growing use of sentiment analysis tools for social media platforms (such as AIM Insights, IBM Watson Natural Language Understanding, or Lexalytics), policy makers can monitor public opinion in real-time, gaining insights into what topics resonate most with citizens and which issues they are sensitive to. By analyzing vast amounts of online data, AI can detect shifts in public sentiment, allowing policy makers to craft their policies in ways that align with the concerns and interests of their constituents.

More sophisticated solutions also exist to promote collaboration between humans and AI agents for better policy making. For instance, a collaboration between the Citizens Foundation and The GovLab has led to the development of the Policy Synth software library, facilitating the creation of multi-scale AI agent logic flows, in order to help governments and citizens make better decisions together by integrating collective and artificial intelligence.

However, as promising as these tools are, they raise critical questions around inclusivity and bias. AI policy tools could unintentionally favor more digitally literate populations, reinforcing existing disparities. Furthermore, the insights AI generates are only as good as the data fed into the system. If datasets are skewed or incomplete, AI feedback mechanisms could misrepresent public opinion, potentially leading to misguided policies. For AI to genuinely strengthen democratic practices, these tools must be carefully designed to avoid exacerbating existing inequalities and ensure that they reflect the voices of marginalized and less digitally active groups.

Moreover, automated feedback systems pose significant risks, as they can be exploited to undermine the policy process by generating large volumes of deceptive or misleading input. A prime example occurred in 2017, when bots flooded the Federal Communications Commission (FCC) during a public comment period on net neutrality. More than a million fake comments were submitted, falsely representing public opposition to the rules. Regulators were able to detect the fraud because of the repetitive, similar nature of the comments, but the increasing sophistication of AI tools since then raises alarming concerns.

Today's AI systems can craft more convincing and varied submissions, making it harder to detect fraudulent attempts to sway policy decisions.

As AI-generated content becomes more indistinguishable from legitimate public feedback, policymakers may find it increasingly challenging to discern authentic civic engagement from manufactured input. This growing threat calls for urgent measures to safeguard the integrity of participatory government processes. Without robust detection mechanisms and oversight, automated feedback systems risk being manipulated, distorting public opinion, and undermining democratic governance.

Some may argue that *"the answer to the machine is in the machine"* (Arthur C. Clark). Effective solutions could involve advanced AI tools for detecting anomalies, identifying patterns of fraudulent or deceptive activity in large datasets. For example, machine learning algorithms can be trained to differentiate between genuine feedback and bot-generated responses, by analyzing submission patterns, metadata, and linguistic features, as well as patterns of unusual submission timing, frequency, or geographic clustering. AI-driven verification tools could also help policymakers validate the authenticity of submissions by cross-referencing them against publicly available data sources (Adam & Hocquard 2023)

AI solutions may also be used to help minimize bias by ensuring that input from diverse populations is equitably represented in policymaking processes. For instance, IBM has developed an open-source toolkit (AI Fairness 360) which can be used to examine, and mitigate discrimination and bias in machine learning models. AI solutions can also be designed to weight input based on demographic data, geographic location, or socioeconomic factors (Mensah 2023), adjusting for potential imbalances by prioritizing input from underrepresented groups, ensuring that their voices are amplified. Besides, to avoid reinforcing digital divides, AI tools could be integrated with multiple platforms to collect input from various channels, such as phone surveys, in-person consultations, or community meetings, in addition to online forms.

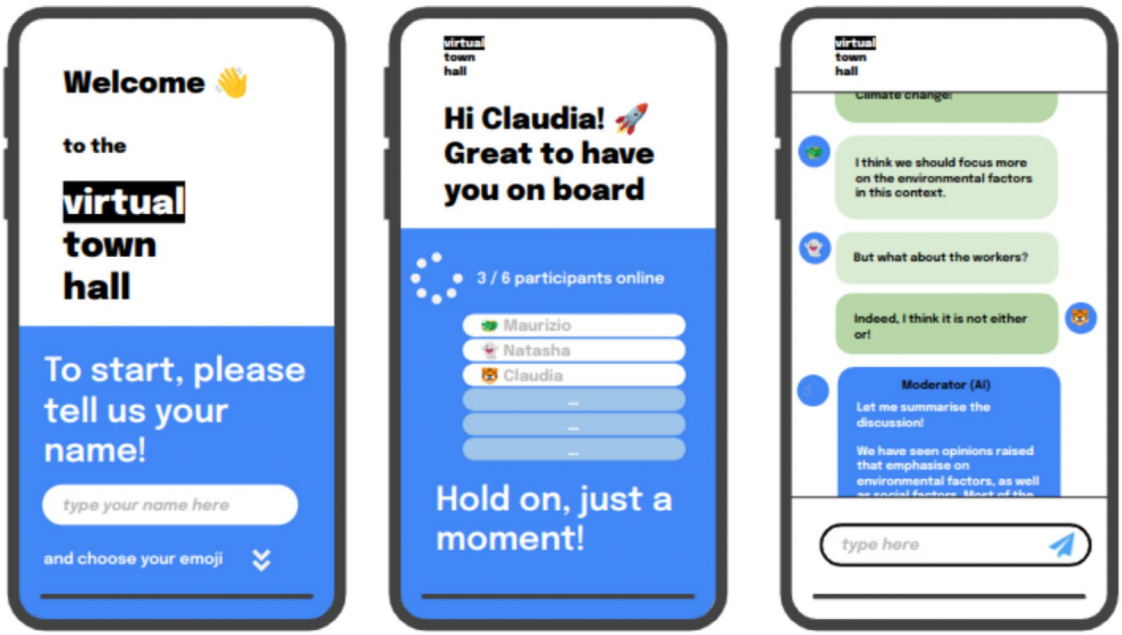
Of course, like with every technological solution, the effectiveness of AI systems is limited. Even if more (or better) AI solutions can be used to cope with the drawbacks or limitations of previous AI solutions, tech solutions cannot succeed in isolation. What's needed is a parallel focus on preparing people to critically engage with the outcomes and implications of AI. Indeed, even if AI can successfully detect biases and filter out disinformation, public education, human oversight and regulatory frameworks remain necessary to prevent these systems from being compromised over time.

B. CITIZEN ENGAGEMENT

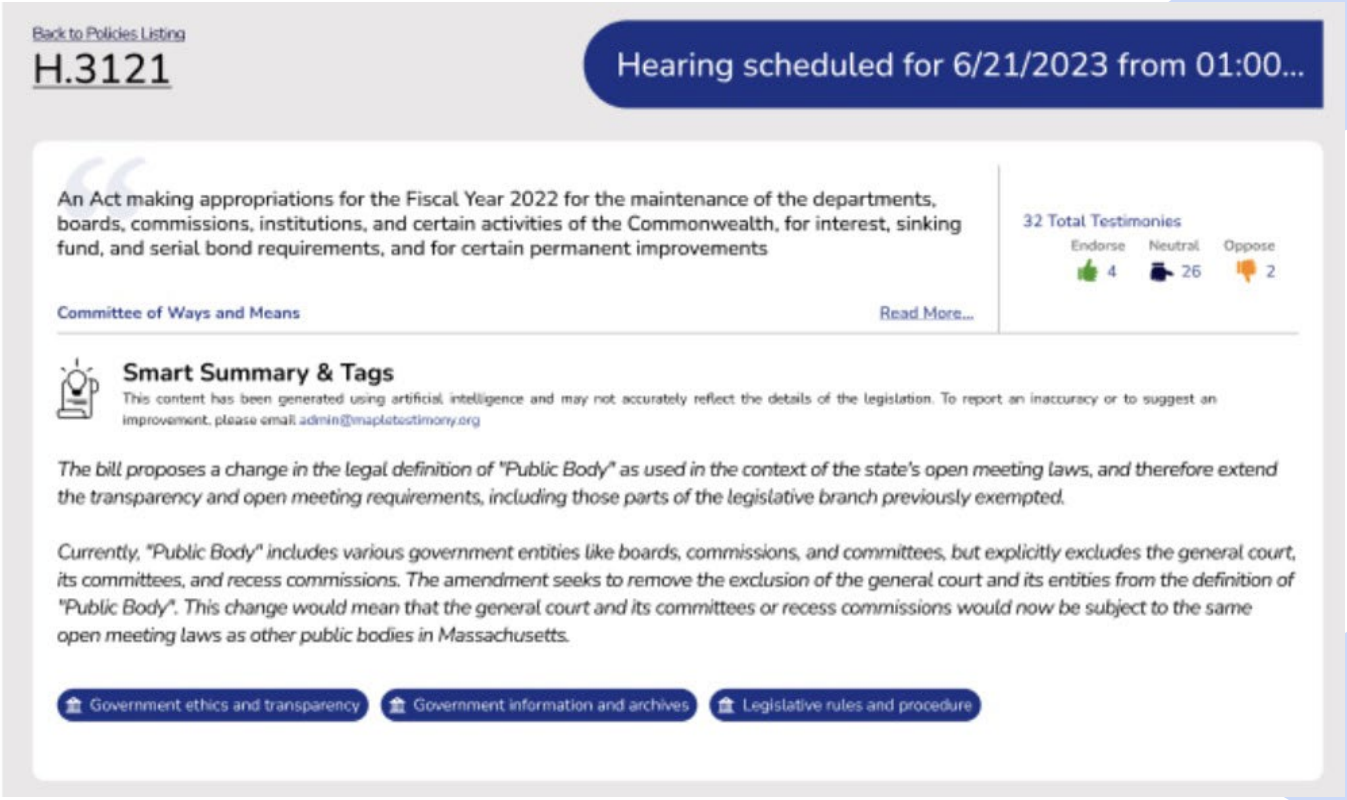
In the past few decades, a global trend of political apathy (Zhelnina 2020) has emerged, resulting from a growing sense of disengagement from politics, often fueled by disillusionment with political systems (Jacoby & Soron 2001, Norris 2004). This poses a series of challenges to initiatives aimed at enhancing democratic participation. AI can help address political apathy by providing more personalized forms of engagement on interactive civic platforms, simplifying complex information, and encouraging citizen participation through gamification. AI can also serve as a catalyst for collective input, enabling more active and participatory policymaking by breaking down barriers to access and allowing citizens to engage with complex policy debates in a structured manner.

SANDER highlights the potential of AI tools to promote citizen engagement via initiatives like **Make.org**'s public consultation platform and **MAPLE**'s open-source legislative engagement system. Both rely on AI to make democratic processes more accessible, facilitating public participation in policy discussions and organizing the diverse opinions submitted by citizens.

Make.org facilitates citizen engagement around open questions, inviting citizens to make proposals and vote on the proposals of others. It comes along two other platforms: **Dialog** (which has already been used by the French Ministry of Economy and the German Ministry of Interior) to connect relevant stakeholders and get them to collaborate and design impactful projects; and **Panoramic**, using AI to make it easier for everyone to access complex content such as the expansive deliberations within citizen assemblies. All of these platforms create channels for engagement between citizens and policymakers, promoting collective decision-making on a whole new scale. This enhances participatory policymaking by reducing barriers to entry and helping citizens express their views on pressing societal issues. For example, the Agenda of Hope that was held in the run-up to the European elections witnessed over 1.5 million votes and collected more than 5,000 proposals, demonstrating the ability of AI tools to manage collective input.



MAPLE (the Massachusetts Platform for Legislative Engagement) aims to promote citizen education and public engagement with the Massachusetts legislature by providing educational materials, facilitating public comment on legislation. It also contributes to increasing transparency around the Massachusetts legislative processes by providing an AI-powered searchable database of bills and a repository of public testimonies. MAPLE also makes the legislative process more comprehensible by leveraging AI to summarize complex legal and policy documents, helping users understand legislative data and synthesizing testimonies. As such, MAPLE can promote a more constructive civic dialogue by avoiding the pitfalls of performative social media engagement—creating instead a more substantive, focused discourse around legislative issues. This is particularly important in an era where much of online discourse has been reduced to polarized or performative behavior. Additionally, by having human moderators review all submissions, MAPLE ensures that civic discourse remains productive and civil. This human-in-the-loop model helps prevent the spread of misinformation and promotes a more trustworthy environment for users to engage with legislative processes.



Other citizen engagement platforms include Your Priorities, Go Vocal, and Assembl, which use toxicity screening AI tools to flag inappropriate inputs. For instance, Your Priorities uses Jigsaw's Perspective API for content curation, and AI-based anomaly detection to identify content or users that should be removed from the platform. It also leverages LLM-based systems to provide written summaries of users' inputs. Platforms like deliberAIde use LLMs to enhance and scale up the operations of citizens assemblies and deliberations with the use of AI facilitators to support the deliberative process (Argyle & al. 2023). These AI systems can facilitate multiple assemblies simultaneously, ensuring that every participant has a say, and requesting clarifications when needed (McKinney 2024) AI can

also be used to amplify collective intelligence. For instance, Swarm from Unanimous AI helps large groups converge on AI-driven decisions, predictions or insights; whereas platforms Thinkscape, Harmonica AI and Common Good AI leverage AI to facilitate productive real-time conversations in large human groups by subdividing them into smaller groups and using multiplayer AI agents to build engagement. Thanks to AI translation tools, deliberation can be done in a multilingual process with automated translation (Kalampokis & al. 204).

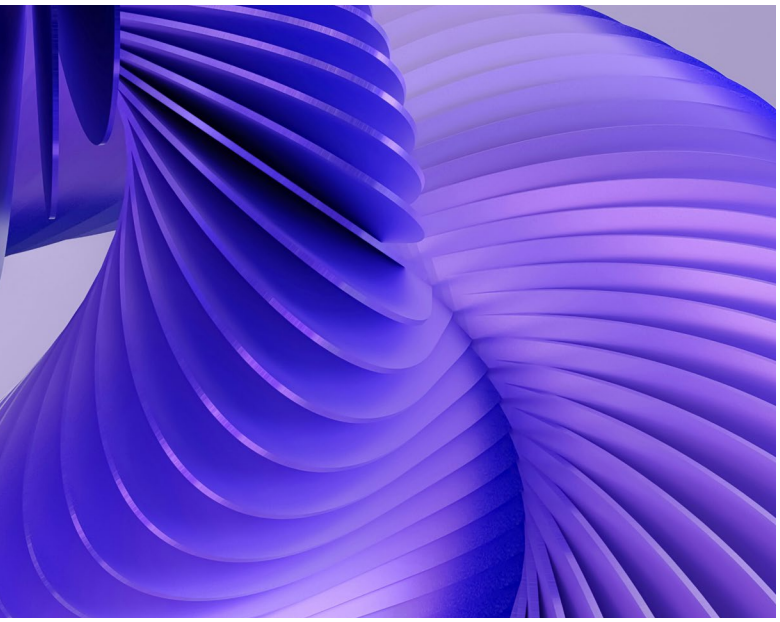
However, while the potential benefits are clear, there are critical concerns about the equity and inclusivity of these AI-driven engagement tools. One of the central challenges is ensuring that AI-driven civic education and engagement platforms resonate across diverse populations, including those with different levels of digital literacy, education, and access to technology. In a state like Massachusetts, where inequality in digital access exists, the question of whether AI-powered legislative engagement platforms like MAPLE can effectively engage all citizens becomes particularly relevant. If not designed with inclusivity in mind, these platforms may inadvertently favor more tech-savvy and digitally literate populations, leaving behind marginalized groups with limited access to digital tools. Similarly, in the context of citizen engagement, even though Make.org facilitates public engagement in large-scale debates, the representation of voices might not be as broad as the numbers suggest. Platforms must incorporate robust measures to ensure that they successfully reach marginalized populations and voices from underrepresented communities, such as rural populations or those less familiar with digital platforms. This is particularly critical when using AI to synthesize feedback on public policies, since there is a risk that the algorithms used to summarize citizen proposals may overemphasize certain viewpoints, potentially skewing policy outcomes toward the preferences of more vocal or engaged groups.

Moreover, it remains unclear whether AI-driven tools can bridge the gap between governments and citizens who may already be skeptical of technology and governance institutions. While AI can enhance civic dialogue, there is a risk that AI systems might become perceived as opaque or manipulative, particularly if citizens feel that their inputs are being filtered or synthesized by algorithms they don't understand. In particular, AI's ability to summarize opinions can sometimes flatten nuanced perspectives, making it harder for policymakers to fully grasp the depth of public concerns. The challenge for platforms like Make.org is to ensure that complexity is not sacrificed for simplicity, and that diverse opinions are adequately represented in policy deliberations.

The Democratic Commons project, a research initiative initiated by Make.org, in partnership with leading institutions like Sciences Po and Sorbonne University, is an important step toward addressing some of these concerns. By developing frameworks to assess and mitigate AI biases, this project seeks to ensure that AI tools align with democratic principles and reflect a wide array of citizen voices. The involvement of organizations like Hugging Face and Mozilla.ai further highlights the importance of ethical AI in civic engagement.

Finally, as noted by **SANDERS**, even if these projects succeed in broadening participation, the *quality* of that participation may not necessarily improve, nor may it lead to tangible changes in legislative outputs. Indeed, even if a platform attracts a large number of users, the long-term success of these projects depends on maintaining broad, diverse, and sustained participation. The concern with political apathy is that even if people engage initially, sustaining their interest and involvement over time remains difficult. To have a real impact, citizen engagement platforms must get participants to engage meaningfully —through informed dialogue, critical thinking, and valuable input that can influence policy. On their side, governments must

demonstrate a commitment to seriously engaging with the outcomes of deliberative processes, showing how citizen participation will lead to tangible societal impact. This includes outlining how public input will be processed and evaluated, and acting transparently in how recommendations are assessed and integrated into policy discussions.



C. MORE EFFICIENT SERVICE DELIVERY

In addition to promoting citizen engagement and more informed policy making, AI also has the potential to transform the way public administrations provide services to citizens. This can have a significant impact on democracy by improving government accessibility and responsiveness to citizens.

For instance, in 2018 Estonia launched a task force to develop a national strategy for Artificial Intelligence (published in 2019) focused on the adoption of AI in the public sector. One of the most prominent examples is the *Bürokratt* system, a network of AI-powered chatbots that help answer queries about government services 24/7, reducing the workload on human staff and enabling citizens to receive accurate information faster. These chatbots are integrated across various government services, allowing citizens to access a multiplicity of services through voice commands and conversational interfaces. Other examples pioneered by Estonia include the use of AI for tax fraud detection systems, traffic management, emergency response and assistance. Finally, Estonia's e-governance model also uses AI to proactively offer services based on a citizen's data profile, simplifying interactions and reducing bureaucratic delays. These include tools for monitoring and profiling risk groups (such as young people who are not in education, employment or training), as well as machine learning software to match job seekers with employers, or to predict the healthcare needs of patients with chronic illnesses. These initiatives have positioned Estonia as a global leader in the application of AI in public administration, making service delivery more efficient, responsive, and transparent. With its Digital Agenda 2030 (adopted in 2021), Estonia aims to build a more proactive digital state, where public services cater to citizens without the need for citizens to take initiative.

Similar initiatives have been undertaken, albeit with a narrower scope, by several countries and municipalities around the world. For instance, the city of Barcelona in Spain integrates AI into its *smart city* initiative, where AI is used for traffic management, energy efficiency in public buildings, and urban planning. Similarly, Singapore integrated AI in its *Smart Nation* initiative, using AI tools for public safety through predictive policing, real-time analysis of surveillance footage, and traffic flow management. AI also powers chatbots that assist citizens in accessing government services.

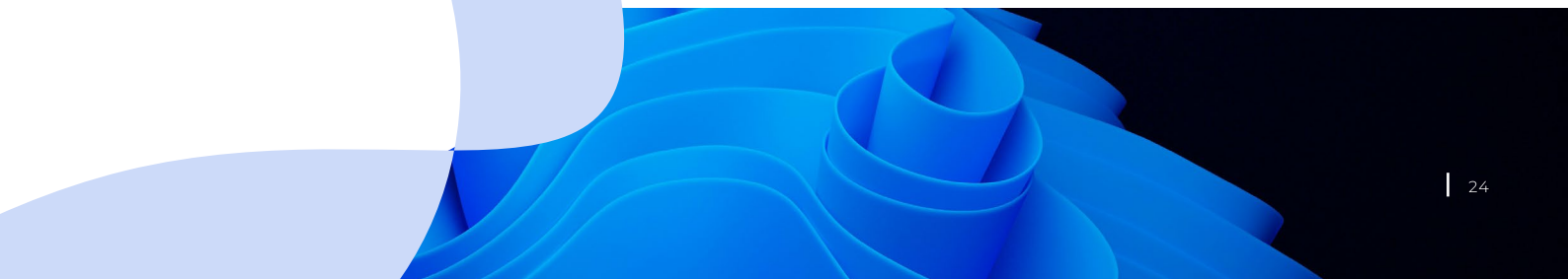
In addition to promoting citizen engagement and more informed policy making, AI also has the potential to transform the way public administrations provide services to citizens.

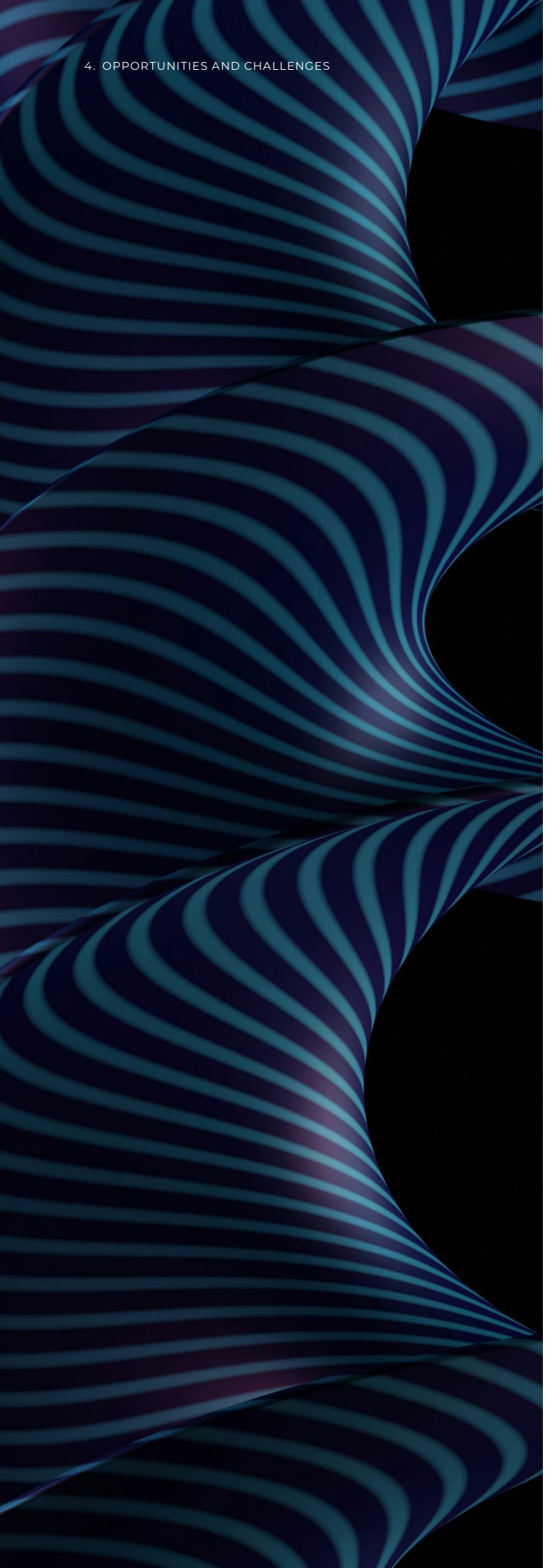
4. OPPORTUNITIES AND CHALLENGES

4. OPPORTUNITIES AND CHALLENGES

As discussed in the previous sections, AI can play a role in both protecting and promoting democracy. In fact, AI can be leveraged as both a tool for resilience and as a mechanism to engage and mobilize the public. Yet, these two functions, albeit substantially distinct, are closely interrelated through a complex network of interdependence that emerge between democratic resilience and citizen engagement.

To begin with, effective defense mechanisms are essential for creating an environment where democracy can thrive. AI's role in combating disinformation and polarization (**Section 2a**) is crucial, as misinformation not only distorts public discourse but also erodes trust in democratic institutions. A healthy information landscape also supports the prevention of cyberattacks (**Section 2b**), as cyber vulnerabilities often exploit informational weaknesses. By identifying and mitigating the spread of false narratives, AI also helps to maintain a more accurate and reliable information ecosystem. This, in turn, is vital for informed policy-making, as decisions rooted in accurate data and collective understanding enable more effective governance. By providing accessible tools for large-scale feedback analysis (**Section 3a**), AI can then strengthen the democratic process by offering policymakers insights that represent a broader segment of society. In particular, the feedback analysis tools provided by AI can substantially enhance citizen engagement (**Section 3b**), creating a deeper, more reflective form of participation that ultimately can lead to more informed and participatory policy-making.





Moreover, the combination of using AI to protect democratic institutions from external attacks (**Section 2b**) and misinformation (**Section 2a**) is fundamental for the uninterrupted functioning of democratic processes. Cyber-attacks can disrupt elections, manipulate information, and undermine public confidence in government institutions. By fortifying these systems, AI allows for a stable political landscape where citizens feel safe to express their views and engage in the democratic process. This stability further encourages citizen engagement (**Section 3b**), as individuals are more likely to participate in political discourse when they believe that their rights and the integrity of their democratic institutions are protected.

Conversely, the strengthening of democracy through informed policymaking and active citizen engagement creates a feedback loop that reinforces the defensive aspects of democracy. An informed public is less susceptible to manipulation by disinformation campaigns, and a civically engaged citizenry actively participates in safeguarding democratic values. Through platforms that facilitate citizen input, AI not only gathers insights but also empowers citizens to voice their perspectives, leading to more responsive and accountable governance.

This holistic view highlights that defending democracy and strengthening it are not isolated activities—they are mutually reinforcing processes that require an integrated and multi-faceted strategy for democratic resilience. By integrating AI effectively into both of these domains, policymakers can cultivate a resilient democratic environment that promotes active citizenship while simultaneously protecting against threats. Ultimately, this synergy is essential for the establishment of a robust democracy capable of adapting to the challenges of the modern media landscape.

However, these opportunities are not devoid of challenges. Concerns about bias, privacy and transparency have become critical issues. AI requires careful scrutiny to ensure that it serves the interests of society. We list below the key challenges that must be accounted for:

Fighting Against Disinformation and Polarization (Section 2a):

- **Bias and Accuracy:** AI systems trained on historical data can perpetuate biases in the dataset, leading to the misidentification of disinformation or even the flagging accurate information as false (Leiser 2022).
- **Information Overload:** The vast amount of information available online complicates the detection of disinformation. As demonstrated during the Covid pandemic, AI may struggle to distinguish between nuanced perspectives and outright falsehoods, potentially resulting in oversimplification of complex issues.

Resilience Against Cyber-Attacks (Section 2b):

- **Evolving Threat Landscape:** Cyber-threats are constantly evolving, requiring AI systems to adapt rapidly. In particular, polymorphic malware poses challenges in maintaining up-to-date AI tools able to detect increasingly sophisticated attacks.
- **Data Privacy and Security:** Strengthening resilience against cyber-attacks often involves extensive data collection and monitoring. The EU ban of facial recognition technologies in public spaces represents an attempt to balance the need for security with the protection of individual privacy rights.

More Informed Policy Making (Section 3a):

- **Quality of Data:** Effective policy-making relies on high-quality data. If the data is flawed or unrepresentative, the AI recommendations may be misguided—e.g, AI-driven predictive policing systems trained on historical arrest data have been shown to reinforce patterns of racial profiling.
- **Political Will and Implementation:** Translating AI-generated insights, into actionable policies requires political will and commitment from decision-makers, which can be lacking in polarized political environments (Yar & al. 2024)

Citizen Engagement (Section 3b):

- **Accessibility and Inclusivity:** AI tools may create barriers for under-represented populations or marginalized groups that do not have access to technology or have low digital literacy. This was illustrated by the Aadhaar biometric ID system in India, which created an unintended digital divide, excluding marginalized groups from accessing services like welfare benefits, food distribution, or healthcare.
- **Trust in AI:** For citizens to engage meaningfully, they must trust the platforms and systems powered by AI. Concerns over data privacy, algorithmic transparency, and the manipulation potential can hinder engagement efforts (Bedue & Fritzsche 2022)

Hence, although AI can bolster both the protective and participatory functions of democracy, there are clear limits and risks to the use of these tools. These risks must be addressed through ongoing research and regulatory oversight to ensure that AI genuinely serves the democratic interest without exacerbating existing challenges.

5. ACTIONABLE TAKEAWAYS FOR INDUSTRY AND POLICY

5. ACTIONABLE TAKEAWAYS FOR INDUSTRY AND POLICY

A. INDUSTRY RECOMMENDATIONS

Scalable solutions through big tech require high levels of transparency and oversight

To effectively safeguard democratic governance, AI solutions must be scalable and designed to address specific challenges on social media and public consultation platforms. For instance, large social media platforms like Facebook and Tiktok have deployed AI systems that can detect harmful content and flag disinformation in real-time. Yet, these mechanisms should come along with specific guarantees to ensure transparency and accountability in the process. This requires developing auditable algorithms that allow users to see why content has been flagged, and how moderation decisions are made. Collaboration with third-party fact-checkers, including organizations such as Snopes and Full Fact (which Facebook has already partnered with) is key to maintaining both transparency and trust.

Big tech companies can also leverage AI for early detection of cyber-attacks on democratic infrastructure, such as phishing campaigns targeting election officials or critical online government services. For example, Google's Project Shield is a service that defends news, human rights, and elections-related sites from DDoS attacks. It leverages AI to analyze large volumes of data for patterns signaling a potential attack. Tech companies should make these cybersecurity tools available to other companies, and potentially also to governments, with dedicated task forces to ensure that critical democratic systems remain secure.

To promote citizen engagement, large online operators can deploy AI-powered public consultation platforms that analyze vast amounts of feedback from citizens. For example, Pol.is is an open source engagement platform that uses AI to map opinions in large-scale discussions that could be adopted by governments or public administrations to understand public sentiment on policy proposals. This would ensure an inclusive and diverse collection of citizen input, reducing biases that traditional surveys may present. Companies developing these tools should also commit to open-sourcing these technologies to encourage wider government adoption, especially in smaller countries or municipalities with limited resources.

Private companies developing AI should abide by specific principles and standards

Private companies should make a public commitment to ethical AI development through self-imposed standards, such as Google's AI principles or Microsoft's Responsible AI Standards, to ensure their systems align with democratic values like fairness, inclusivity, and respect for individual rights. By incorporating these standards into internal workflows and open-sourcing critical components of their systems, companies can increase the credibility of their AI tools and help set industry-wide best practices for safeguarding democracy.

Besides, as many of the leading AI companies are concentrated in countries like the U.S. or China, to facilitate trust across borders, these companies must adopt internationally recognized standards for AI transparency or accountability, such as the ISO/IEC standards on AI or the OECD AI Principles.

Third-party audits by independent global organizations, like the European Union Agency for Cybersecurity (ENISA) could provide unbiased evaluations of AI tools, ensuring governments worldwide have confidence in their reliability and fairness.

B. PUBLIC-PRIVATE PARTNERSHIPS

Combining private sector innovation with public sector oversight and accountability

Public-private partnerships can bridge the gap between the private sector's innovation potential and the public accountability that governments and civil society offer. AI companies, with their vast resources and technological expertise, can develop tools and solutions to enhance democratic processes, but without collaboration with governments and civil society, these innovations may fail to address broader societal needs or uphold democratic values.

The private sector could partner with governments and civil society to co-create specific AI tools that address ethical and social challenges, while preserving democratic processes. For example, Google's AI for Social Good initiative is designed to help underserved communities by partnering with specific organizations to build data-driven, AI tools and solutions that tackle pressing social challenges. Similar initiatives could be done in collaboration with national governments, to encourage the development of AI-driven platforms that respond to the needs of specific governments or civil society organizations. Governments could even incentivize the ethical deployment of AI for social good through tax breaks or grant programs for companies that successfully deploy democratic-friendly AI solutions (more below).

For instance, large social media platforms that already enjoy a wide-ranging user base could collaborate with governmental authorities to develop AI-driven public consultation platforms that aggregate and analyze citizen feedback in real-time, allowing policymakers to make more responsive decisions. Of course, these platforms would need to be developed alongside human rights organizations to ensure that AI systems used to analyze public opinion reflect the voices of diverse communities and to avoid reinforcing biases. The establishment of ethics boards made up of civic organizations and marginalized groups could even provide an additional layer of oversight to these systems.

Regulatory sandboxes to experiment with AI technology in a low-risk environment

AI companies could engage in public-private partnerships with governments and civil society organizations to develop regulatory sandboxes allowing AI innovations to be tested while remaining compliant with democratic safeguards. The UK Financial Conduct Authority (FCA) has already pioneered a regulatory sandbox for fintech, allowing companies to trial financial innovations while remaining compliant with regulation. This model can be extended to AI, where companies like OpenAI or Anthropic could work with governments to test new AI tools and technologies in a low-risk environment, allowing for rapid and on-going adjustments before widespread implementation. Civil society organizations could help provide insights into what qualifies as ethical and sound AI design, as has already been done by Mozilla's AI transparency initiative aimed at providing insights on how to implement public oversight mechanisms that will help citizens understand how AI systems impact their lives.

Development of novel AI solutions in collaboration with the public and private sectors

Policymakers could establish public-private innovation programs like the Defense Advanced Research Projects Agency (DARPA) in the U.S., where private companies collaborate with government researchers to co-develop AI tools aimed at protecting or promoting democracy— as it was done with **Pol.is** in Taiwan or **Citizen Space** in the UK.

C. POLICY RECOMMENDATIONS

Harmonized regulatory frameworks to promote AI innovation with legal safeguards

Policymakers should prioritize the creation of **regulatory frameworks** that encourage AI innovation while embedding key democratic safeguards. For instance, the European Union's AI Act sets up robust regulations governing the deployment of AI in various sectors, with an emphasis on transparency and accountability (Taeihagh 2021). Yet, because AI applications often have a global scope, more harmonized and interoperable regulatory frameworks are needed to create alignment between regions—such as bridging EU's General Data Protection Regulation (GDPR) with other privacy laws to ensure consistent safeguards in data use. Specifically, to ensure data privacy standards are upheld, governments should establish clear regulations for how personal data collected by public and private AI systems is handled. In the EU, the GDPR already provides a solid model for data protection. Additional safeguards could include using privacy-preserving techniques like differential privacy or federated learning to process data without compromising individuals' identities.

Alternatively, policymakers could **require mandatory audits** for AI systems used in public governance. For instance, Estonia developed a comprehensive digital governance model that uses AI for public service delivery (Robinson & al. 2021). Policymakers could mandate routine audits to ensure that these AI systems remain unbiased, transparent, and fair, requiring companies to provide clear documentation of how AI algorithms operate and make decisions. This would significantly contribute towards building public trust and ensuring that these tools do not inadvertently exacerbate biases or discrimination.

Public investments in AI tools and infrastructure to preserve digital sovereignty

Beyond regulations, preserving digital sovereignty also requires **public investments** for the creation of AI systems that are owned and managed by local and accountable public bodies. Public AI infrastructure is especially relevant for Europe, whose approach to AI centers on data sovereignty, trust, and transparency (as reflected by regional regulations such as the GDPR and the AI Act). Public investments in AI could contribute to both reducing dependency on non-EU tech giants (preventing potential domination by outside corporate powers, particularly from the U.S. or China) and promoting the development of ethical and transparent AI systems that align with European values of privacy, inclusivity, and fairness. Initiatives like Gaia-X, a project aimed at creating a federated data infrastructure based on European standards, and the proposed creation of the European Distributed Institute for AI in Science (EDIRAS)—also referred to as the "CERN for AI"— are important steps towards promoting European sovereignty in the field of AI.

Nationally, the UK Department for Science, Innovation and Technology announced in 2023 the launch of the Foundation Model Taskforce, with £100M to develop AI infrastructure and public service procurement. The goal is to ensure sovereign capabilities and establish the UK as a world leader in AI innovation, while serving as a global standard bearer for AI safety.

Another prominent example of publicly funded AI infrastructure is the **Falcon** foundational model, funded by the UAE and developed by the Technology Innovation Institute (TII) in Abu Dhabi. Falcon was released as open source for any public or private institutions to use.

Encouraging democratic AI innovation through grants, subsidies, and tax benefits

In addition to developing their own AI infrastructure, governments could also take proactive measures to encourage the development of AI solutions that promote and protect democratic systems by offering tax incentives, grants, and subsidies for companies working on technologies that strengthen democratic systems. These incentives could, for instance, target companies developing AI algorithms to counter the spread of false information on social media platforms, especially during critical moments like elections or referendums. Governments could even encourage the adoption of existing tools such as Factmata or Logically by offering matching funds or grants for companies that integrate these solutions into their platforms, ensuring they remain financially viable and scalable across large networks.

Creation of research hubs or centers of excellence to promote democratic uses of AI

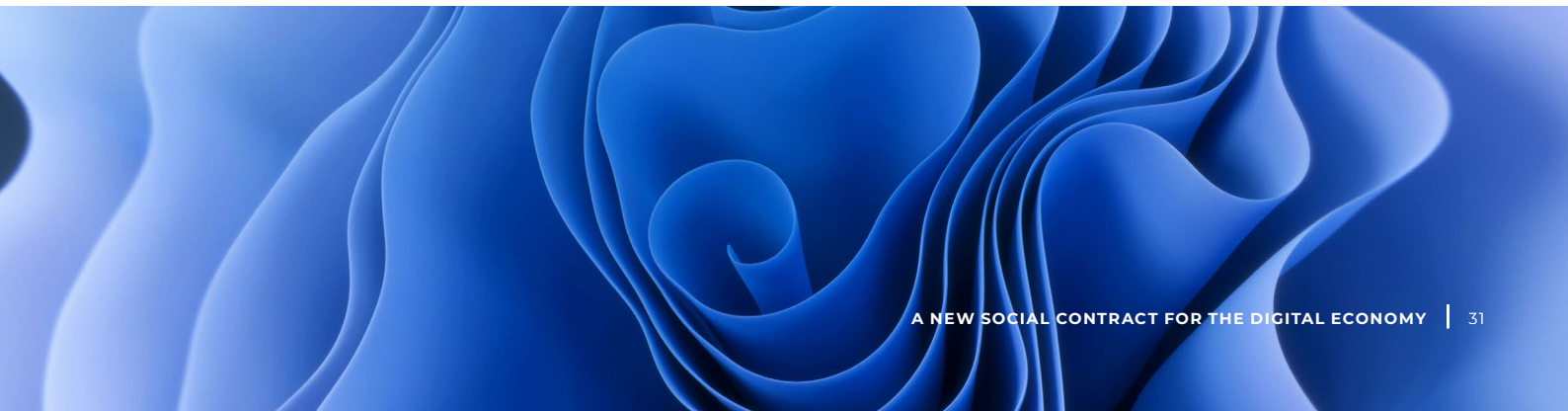
Governments could also establish AI research hubs or centers of excellence that focus on using AI for civic engagement, providing grant funding for companies or academic institutions that contribute to creating tools aimed at promoting transparent public discourse, moderating civic discussions, and increasing participation in legislative processes.

Cross-sector interdisciplinary working groups to provide advices to the public sector

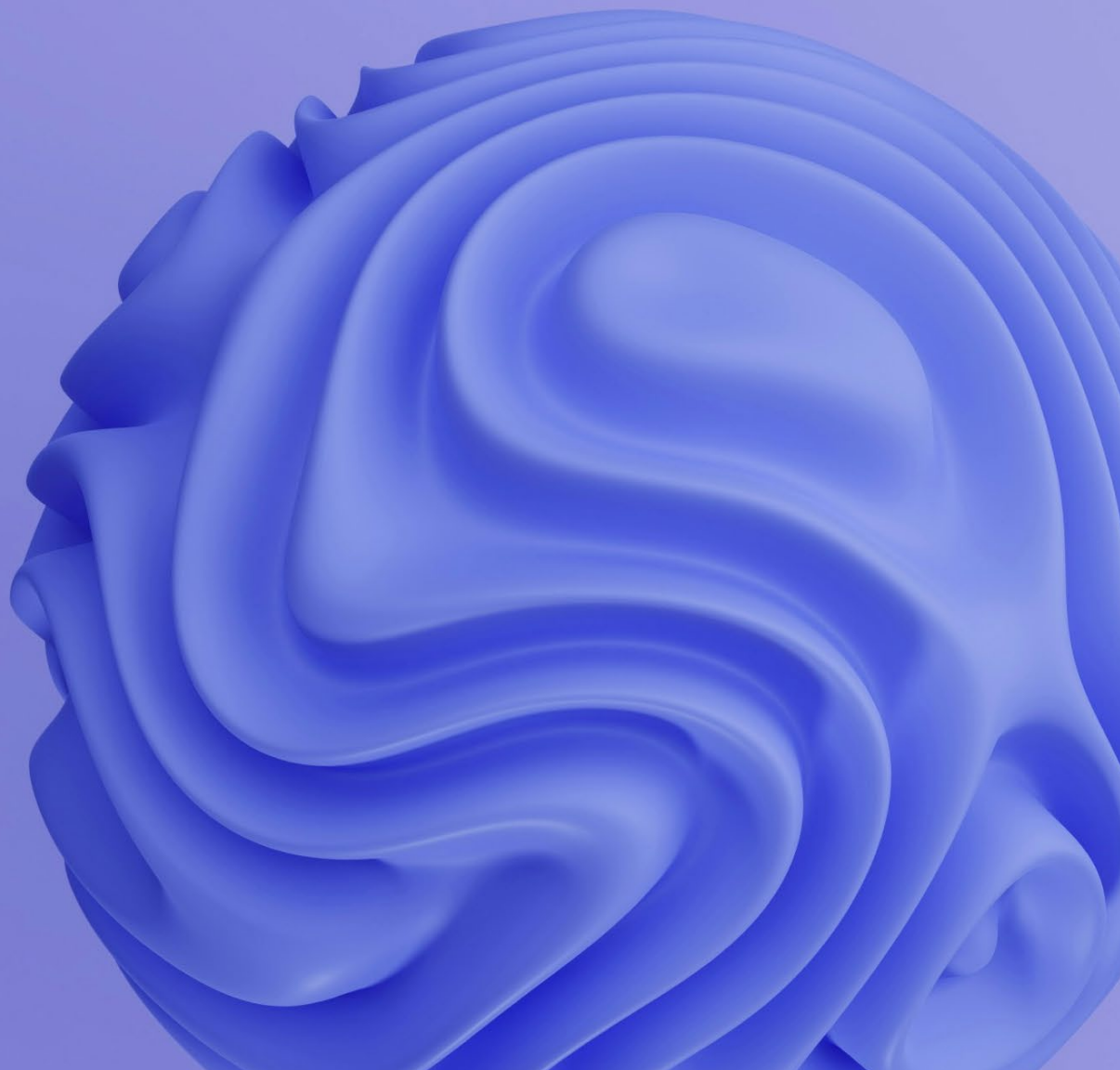
Lastly, cross-sector working groups composed of policy-makers, AI developers, and ethicists could be created to continuously update guidelines and best practices for AI in governance, ensuring they keep pace with rapid technological advancements. These working groups could develop open-source standards for AI accountability—much like the efforts from Mozilla's Trustworthy AI project, which offers actionable frameworks for creating AI systems that are transparent and accountable. By requiring these systems to be auditable and to adhere to clear ethical guidelines, policymakers can ensure that AI applications contribute positively to democratic governance rather than threatening it.

Investments in public education to strengthen citizens's awareness and resilience

Of course, it comes without saying that these technologically-driven initiatives should not come at the expense of more investment in public education, equipping citizens with the skills to detect misinformation and manipulation early on. Technological solutions must be complemented by efforts to strengthen public resilience through media literacy programs, critical thinking initiatives, and educational tools that help individuals become more discerning consumers of information. Investing in these public measures ensures that AI is not just a defensive tool, but part of a broader, proactive strategy to empower citizens to become active contributors and informed participants to the media and information landscape.



6. FUTURE RESEARCH DIRECTIONS



6. FUTURE RESEARCH DIRECTIONS

This report has shown how AI can assume a dual role in both defending and strengthening democracy, when used to fight disinformation, enhance cybersecurity, improve policymaking, and boost citizen engagement. However, there are still significant challenges ahead. In particular, as the use of AI progressively expands in every field of society, significant efforts are needed to integrate it responsibly within national and international laws. Indeed, if AI is to become a transformative force for democratic societies, it must be continually refined and embedded within robust ethical and regulatory frameworks to ensure that it is developped and deployed in a rightful manner.

One important area for research is the interplay between AI, **privacy and surveillance**. While AI has the potential to defend democracy by combatting disinformation and strengthening civic engagement, its deployment also raises significant privacy concerns (Ergashev 2023). There is a growing need to critically assess how AI solutions might inadvertently pave the way for surveillance capitalism (Zuboff 2023), where data collection and monitoring by corporations or governments could erode personal freedoms. Future research should focus on identifying the balance between leveraging AI for democratic resilience and ensuring robust safeguards against intrusive surveillance practices.

Similarly, in the context of **Intellectual Property** (IP), in order for AI to be widely adopted for public or commercial use, particularly in fields like art, music, and software development, there must be a clear regulatory framework that balances innovation with respect for creators' rights. Yet, given the current legal uncertainty surrounding the application of IP regulations to AI, the generative AI landscape has become a battleground, raising a multiplicity of ethical and legal concerns with regard to the use

of copyrighted materials without proper licensing or attribution. On the one hand, artists, writers, and content creators have accused AI developers of infringing on their works by using data collected from the internet without compensation or consent. On the other hand, several companies have openly admitted to scraping and using copyrighted data to train their model without obtaining prior permission from the right holders. This resulted in several lawsuits against major firms like OpenAI and Google. Without a clear regulatory framework, the rights of content creators are at risk of being undermined, while the companies training or using generative AI models may incur significant legal penalties if their activities were to eventually qualify as copyright infringement. Yet, the applicability of existing copyright regulations to the creation of training datasets and to the training of generative AI models is a complicated issue that has not yet been fully resolved. Further research is needed to help governments and regulatory authorities establish clear guidelines and enforcement mechanisms to ensure that AI development and deployment align with existing IP laws, without excessively hindering innovation. In the meantime, private ordering solutions are being developed by the private sector. These include, amongst others, the Fairly Trained initiative, certifying AI companies that rightfully obtained a license for their training data; Adobe's Content Authenticity solution enabling content creators to add a "*do not train*" tags on their works to protect them from being used in AI models without consent; and initiatives like Story Protocol and Alias.studio providing novel technological solutions for automated IP management in generative AI.

Beyond the issues related to the global regulation of AI (Miazi 2023), the **role of AI in global governance** also presents a promising field of study (Sapignoli 2021). In that regard, the Global Governance Institute has recently launched the AI and Global Governance programme to address the critical challenges and opportunities posed by AI on global governance. Future research could explore how AI can support international governance frameworks, by facilitating decision-making in global governance bodies (Truby 2020) and bringing more transparency in international agreements (Igbinenikaro & Adewusi 2024). Exploring these possibilities will be essential to understanding how AI can contribute to a more equitable and efficient global governance system, enhancing democratic principles on an international scale (Daly & Hagendorff 2022).

Lastly, one promising avenue for further research is the study of AI in political philosophy, to explore the ways in which AI could fundamentally alter democratic institutions in both theory and practice. With the increasing role of AI in decision-making processes, traditional political structures shift to accommodate AI-driven governance. Researchers could examine whether AI could contribute to redefining principles like representation, participation, and authority, potentially leading to new models of democratic systems that better integrate algorithmic decision-making alongside human deliberation. Together, these research directions promise to deepen our understanding of AI's role in shaping not just the future of democracy, but the future of governance itself at both national and global levels.

REFERENCES

Adam, M., Hocquard, E. (2023) Artificial Intelligence, democracy, and elections. European Parliamentary Research Service. Available online.

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., ... & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences,* 120(41), e2311627120.

Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management, 35*(2), 530-549.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First monday*, 21(11-7).

Cohane, K. (2021). The Role of AI Twitter Bots Used During US Elections: A study of how malicious Twitter bots play a role in increasing digital conflict and potentially influence voter perceptions leading up to a US election.

Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., & Wei Wang, W. (2022). AI, Governance and Ethics: Global Perspectives.

Ellaky, Z., & Benabbou, F. (2024). Political Social Media Bot Detection: Unveiling Cutting-edge Feature Selection and Engineering Strategies in Machine Learning Model Development. *Scientific African*, e02269.

Ergashev, A. (2023). Privacy concerns and data protection in an era of ai surveillance technologies. *International Journal of Law and Criminology*, 3(08), 71-76.

Igbinenikaro, E., & Adewusi, O. A. (2024). Policy recommendations for integrating artificial intelligence into global trade agreements. *International Journal of Engineering Research Updates*, 6(01), 001-010.

Jacoby, R., & Soron, D. (2001). The end of utopia: politics & culture in an age of apathy. *Labour,* (47), 203.

Kalampokis, E., Karacapilidis, N., Karamanou, A., & Tarabanis, K. (2024). Fostering Multilingual Deliberation through Generative Artificial Intelligence.

Leiser, M. R. (2022) Bias, journalistic endeavours, and the risks of artificial intelligence. In T. Pihlajarinne & A. Alén-Savikko (Eds.), Artificial Intelligence and the Media: Reconsidering Rights and Responsibilities (pp. 8-32). Cheltenham: Edward Elgar Publishing.

McKinney, S. (2024). Integrating Artificial Intelligence into Citizens' Assemblies: Benefits, Concerns and Future Pathways. *Journal of Deliberative Democracy,* 20(1).

Mensah, G. B. (2023). Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in AI Systems. *Preprint, November,* 10.

Miazi, M. A. N. (2023). Interplay of Legal Frameworks and Artificial Intelligence (AI): A Global Perspective. *Law and Policy Review, 2*(2), 01-25.

Norris, P. (2004, January). The evolution of election campaigns: Eroding political engagement. In *Conference on Political Communications in the 21st Century* (pp. 1-27).

OECD (2023) Working Party on Artificial Intelligence Governance. AI LANGUAGE MODELS: TECHNOLOGICAL, SOCIO-ECONOMIC AND POLICY CONSIDERATIONS, available at https://one.oecd.org/document/DSTI/CDEP/AIGO(2022)1/FINAL/en/pdf

Oshikawa, R., J. Qian, and W. Y. Wang (2020) "A survey on natural language processing for fake news detection." Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020) pp. 6086-6093. Available at http://arxiv.org/abs/1811.00770

Robinson, N., Hardy, A., & Ertan, A. (2021). Estonia: A curious and cautious approach to artificial intelligence and national security.

Sapignoli, M. (2021). The mismeasure of the human: Big data and the 'AI turn'in global governance. *Anthropology Today, 37*(1), 4-8.

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and society, 40*(2), 137-157.

Truby, J. (2020). Governing artificial intelligence to benefit the UN sustainable development goals. *Sustainable Development, 28*(4), 946-959.

UNESCO (2022) Elections in Digital Times: A guide for electoral practicionners.

Villela H. F., F. Corrêa, J. S. d. A. N. Ribeiro, A. Rabelo, and D. B. F. Carvalho, (2023) "Fake news detection: a systematic literature review of machine learning algorithms and datasets," Journal on Interactive Systems, Porto Alegre, RS. Vol. 14, no. 1, pp. 47–58, number: 1. Available at: https://sol.sbc.org.br/journals/index.php/jis/article/view/3020

Yar, M. A., Hamdan, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). Governing with Intelligence: The Impact of Artificial Intelligence on Policy Development. *Information, 15*(9), 556.

Zhelnina, A. (2020). The apathy syndrome: How we are trained not to care about politics. Social Problems, 67(2), 358–378. https://doi.org/10.1093/socpro/spz019

Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (pp. 203-213). Routledge

IE UNIVERSITY
CENTER FOR THE GOVERNANCE OF CHANGE

Written by:
Primavera De Filippi

This is the final report of AI4Democracy, a global research within AI4Democracy, a global research and outreach initiative led by the Center for the Governance of Change at IE University, with Microsoft as strategic supporter. AI4Democracy seeks to harness AI to defend and strengthen democracy through coalition-building, advocacy, and intellectual leadership.

Suggested citation:
Primavera de Filippi. (2025). *Democracy Reloaded: AI to protect and promote democratic governance*, AI4Democracy, IE Center for the Governance of Change.

Images: Unsplash and some images were generated by different AI tools.
Design: epoqstudio.com

FOR MORE INFORMATION ON THE AI4DEMOCRACY INITIATIVE, VISIT:

IE.EDU/CGC/RESEARCH/AI4DEMOCRACY

cgc.ie.edu