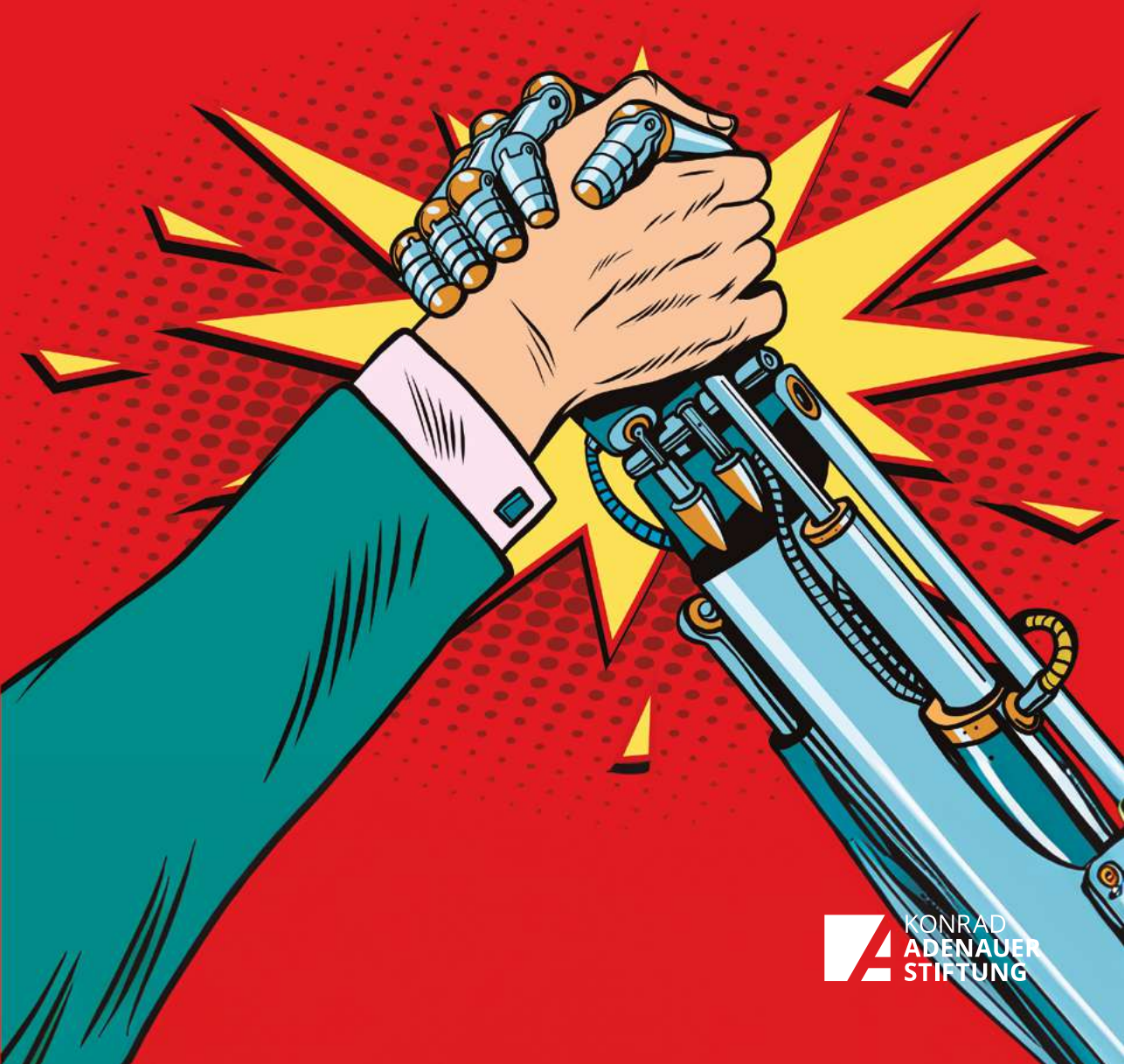


「DIÁLOGO POLÍTICO」

ARTIFICIAL DEMOCRACY





@dplatinoamerica



/dialogopoliticolatinoamerica



@dplatinoamerica

WWW

<http://dialogopolitico.org>

「DIÁLOGO POLÍTICO」



DIÁLOGO POLÍTICO
Año XL, n.º 1, 2025

DIRECTOR
Sebastian Grundberger

TRAINEE
Jonathan Neu

EDITORIAL COORDINATION
Ángel Arellano
Manfred Steffen

EDITORIAL ASSISTANCE
Agustina Lombardi

EDITORIAL BOARD
Fabiola Chambi (Bolivia)
Ana María Saavedra (Colombia)
Julio Castillo (México)
Gabriel Pastor (Uruguay)

COMMUNITY MANAGER
Yaldirmar Ruiz

ENGLISH TRANSLATION
Doris Filipovic
Marianela Santoveña
Manfred Steffen
Mirtha Tovar

IMAGES
Shutterstock

DESIGN AND SETUP
Yamila Murán Leivas
@cretayam

© Konrad-Adenauer-Stiftung
Plaza Independencia 749, oficina 201
11000 Montevideo, Uruguay
Tel.: +598 2902 0943

ISSN: 1688-9665



/dialogopoliticolatinoamerica



@dplatinoamerica



@dplatinoamerica



Fundación Konrad
Adenauer Montevideo



info@dialogopolitico.org



www.dialogopolitico.org



www.kas.de/parteien-lateinamerika

DIÁLOGO POLÍTICO is a platform for democratic dialogue among political opinionmakers about topics of relevance in Latin America, based on the values of freedom, solidarity, and justice. It connects the region with the great geostrategic debates in the world. It is a window for publicizing projects of the Konrad Adenauer Foundation in Latin America.

DIÁLOGO POLÍTICO is part of the Regional Political parties and Democracy in Latin America Program (kas Partidos). Its objective is to reduce political polarization through constructive and informed pluralist debate oriented to the common good, to strengthen the political center from its Social-Christian, Liberal and Conservative roots.

The texts published are the sole responsibility of their authors and do not necessarily express the publisher's opinion. The total or partial reproduction of the content is authorized, citing the source.

Subscribe to **DIÁLOGO POLÍTICO's**
weekly newsletter at
www.dialogopolitico.org

Index

- | | | | |
|----|--|----|--|
| 6 | Foreword
Artificial Democracy | 58 | The fight against
corruption from a new technological paradigm
<i>Denisse Rodríguez-Olivari</i> |
| 8 | Artificial Intelligence and elections:
premature threats?
<i>Sarah Kreps</i> | 66 | Artificial Intelligence and Public
Safety in Latin American Democracies
<i>Gustavo Flores-Macías y Bárbara
Hernández</i> |
| 20 | Governance of algorithms
<i>Daniel Innerarity</i> | 76 | Lars Zimmermann:
“Technology won’t relieve us of
democracy’s tasks”
<i>Interview</i> |
| 28 | Anticipatory governance
with the help of technology
<i>Lydia Garrido Luzardo</i> | 84 | Freedom and Truth Captured by
Artificial Intelligence
<i>Miguel Pastorino</i> |
| 38 | Is it possible to regulate AI?
Global experiences
<i>Ximena Docarmo</i> | 94 | Personalising fakes: towards the
disinformation apocalypse?
<i>Christoph Nehring</i> |
| 48 | From mass networks
to personalised voting
<i>Jesús Delgado Valery</i> | | |

Artificial Democracy

Artificial intelligence is ubiquitous, as promise or nightmare. It is associated with the prospects of solving tedious problems, making better use of material resources and saving time. At the same time, concerns arise that it will replace human beings, cause mass unemployment and kill creativity and freedom.

Many see society threatened by algorithms that can empower, but also manipulate and replace human intelligence. If we define democracy as a debate, a conversation about ideas, what impact will it have if our interlocutor is a machine, a device without feelings? How do we negotiate about government plans and allocate resources with an artifact with no past and no responsibility for the future?

Artificial intelligence is emerging in times marked by the threat of war in a context of global environmental, energy and civilizational problems. AI is presented as a tool for better management, but it requires an awareness of the risks involved in its application without institutional support.

From our commitment to democracy in **DIÁLOGO POLÍTICO**, we want to encourage an exchange based on diverse perspectives. For this purpose, we have invited a dozen international experts.

“To thrive in the digital age, democracies must foster a culture of critical thinking and media literacy among their citizens,” says U.S. political scientist **Sarah Kreps**, who analyses in this issue the impact of AI in elections through the systematic use of deepfakes. The executive director of Transparencia Electoral, **Jesús Delgado Valery**, adds that the use of AI can transform campaigns through tools that allow the identification of groups of voters and the appropriation of their personal data.

The philosopher **Daniel Innerarity** doubts that “algorithms can take over the entire political process, with the efficiency they sometimes promise and with the legitimacy that should justify this new regime”. To resolve this we need to examine the expectations of algorithmic governance and its limits.

“The transformative power and disruptive potential of AI requires ethical governance. Responsible anticipation and future-literate use are key elements in policymaking,” says expert **Lydia Garrido**.

AI is part of a transition that needs to be regulated. **Ximena Docarmo** of Innovalab looks at some possible regulatory models. On the other hand,

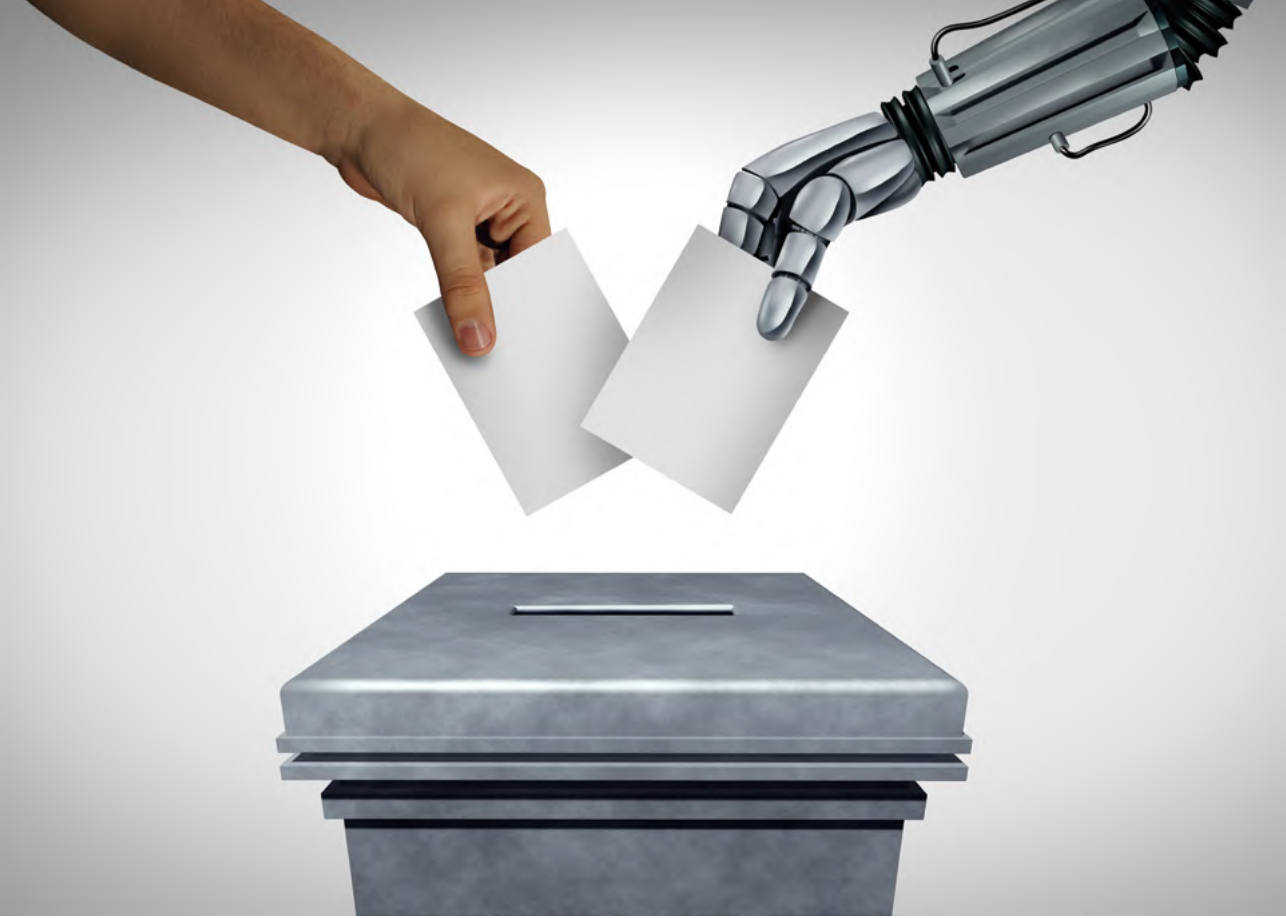
Denisse Rodriguez-Olivari, presents the potential of AI in the fight against corruption that afflicts democratic institutions. The efficient analysis of large amounts of data makes it possible to quickly identify anomalies and act accordingly.

Another issue affecting democracies in the continent is security. **Gustavo Flores-Macías** and **Bárbara Hernández** highlight in their chapter the need for the responsible and ethical use of AI to maximize its benefits and mitigate the risks it poses to civil liberties in Latin American democracies.

Philosopher **Miguel Pastorino** warns about the potential for hijacking the truth and freedom of human beings. In the same vein, **Christoph Nehring**, expert of the KAS disinformation program, explores the effects and implications of AI-generated disinformation and the special role of influencers and journalists in identifying and combating disinformation.

Democracy learns from mistakes and creates a framework for settling differences and conflicts. Its legitimacy is based on transparency in the search for solutions, decision-making processes and accountability. This requires an informed and active citizenry. As our interviewee, expert **Lars Zimmermann**, says: “Technology will not relieve us of the tasks of democracy.”

Sebastian Grundberger and Manfred Steffen



Artificial Intelligence and elections: premature threats?

SARAH E. KREPS

The new technological paradigm and deepfakes influence democratic processes. Manipulated products can mislead people into believing that certain contents are real. Democracies need to foster media literacy among their citizens to ensure that the founding values of democracy among their citizens endure in the face of new challenges.

In 2020, one of the major parties in India, the Bharatiya Janata Party (BJP), used deepfake technology to create videos of one of its [politicians](#), Manoj Tiwari. The videos portrayed Tiwari speaking languages that he does not speak, such as Haryanvi and English, with a goal of targeting different linguistic demographics. The video was relatively benign, an attempt to portray the individual in a positive light, but made manifest the potentially less benign consequences were the technology to be misused. The prospect was not just hypothetical. Just a year before in Gabon, the Gabon presidential office [released](#) a video of its infirm leader, Ali Bongo, who had suffered a stroke, aiming to dispel rumors about his health and political stability. Skepticism about its authenticity fueled political unrest and led to an attempted coup among military officers seeking to restore democracy and stability in the country.

With the rise of artificial intelligence and the growing examples of deepfakes insinuating themselves into the democratic process, researchers have warned about more extreme misuse cases. In 2023, the warnings about deepfakes mounted as 2024, a historic year for elections—with “more voters than ever in history” voting, 64 countries and a combined population

of 49% of the world—[approached](#). *Newsweek* [warned](#) that “deepfakes could destroy the 2024 election.” More generally, the proliferation of generative Artificial Intelligence such as ChatGPT means that it’s not just images that can be inauthentic, but also text such as the news stories individuals read about politics ([Kreps et al 2020](#)). This has prompted scholars to warn about the democracy-eroding effects of AI-generated text ([Kreps and Kriner 2023](#)).

And yet, election after election in 2024 showed that these warnings were overblown or at least premature. In most of the elections, deepfakes or AI-generated content was largely absent, seen by relatively few individuals, and certainly not consequential enough to sway elections.

The question then is why? With such potential to destroy elections in ever-different ways—given that the number and use of deepfakes or other AI-generated content is almost infinite—why has AI either not been created or not had consequential impacts on the elections consistent with the pessimistic predictions?

This essay first defines deepfakes and AI-generated content and the reason the technology has been predicted to undermine democracy, particularly elections. It

then takes stock of how AI has been used in different 2024 elections, pointing to the dearth of significant impacts relative to the theoretical prospects. The essay then offers ideas for why AI has not had the consequential impact in line with its potential and suggests why these past experiences may not be prologue. It closes with reflections about potential future misuses and how democratic polities must remain vigilant and digitally literate.

What is Generative AI and What is the Potential Threat to Democracy?

Generative Artificial Intelligence [refers](#) to a subset of artificial intelligence that is capable of creating new content, such as text, images, audio, and video. The technology relies on *machine learning* models, particularly those involving deep learning, to generate outputs that mimic real-world data. Although the rise of consumer-facing ChatGPT has led to the proliferation of text-based generative AI and concerns about democratic disruptions, deepfakes have already created realistic but fake videos of political candidates or public figures and been used for misinformation.

Deepfakes [are](#) synthetic media in which a person in an existing image or video is replaced with someone else's likeness using artificial intelligence (AI) and machine learning techniques. The term "deepfake" comes from the combination of "deep learning," which is a subset of machine learning, and "fake," which implies something that is inauthentic. Deepfakes are created using a type of artificial intelligence called Generative Adversarial Networks (GANs). Think of GANs as a pair of digital artists. One artist, the "generator," tries to create fake images or videos that look real, while the other artist, the "discriminator," tries to spot the fakes. Through this back-and-forth

process, the generator gets better at making realistic-looking fake content. Another type of AI, called autoencoders, helps by learning the patterns in real images or videos and then using that knowledge to recreate similar, but fake, content.

The fundamental concern with deepfakes is that their verisimilitude between reality and the manipulated audio, video, or images can mislead individuals into thinking that the content is real. Deepfakes stand to pose a theoretical threat to democratic elections due to their potential to manipulate public perception, spread misinformation, and undermine trust in the electoral process.

For example, they could be used to create false narratives by making it appear that political figures have said or done things they never actually did. This can mislead the public and sway public opinion based on fabricated information. The [Gabon deepfake](#) is a good example, having both eroded trust in public figures and the media and undermining the integrity of the democracy.

Indeed, scholars have warned and shown that people cannot discern AI-generated news content from actual news content ([Kreps et al 2020](#)), that AI-generated propaganda is persuasive to individuals in ways that could accomplish the goal of psychological manipulation (Goldstein et

» The fundamental concern with deepfakes is that their verisimilitude between reality and the manipulated audio, video, or images can mislead individuals into thinking that the content is real. «

» [Deepfakes] could be used to create false narratives by making it appear that political figures have said or done things they never actually did. «

al [2024](#)), and that members of Congress cannot distinguish between AI and human-generated constituent messages, which can potentially affect the legislative agenda given the potential to write advocacy messages at scale ([Kreps and Kriner 2023](#)). Although these studies are about AI-generated content in general rather than specifically deepfakes, the points are the same. Individuals cannot tell the difference between AI and human-generated content, the content can manipulate individuals, and these technologies have the capacity to proliferate and scale.

Although propaganda and misinformation have long been available, generative AI is different in several ways. For example, deepfakes can create highly realistic audio, images, and videos that are much more convincing than traditional methods that might seek to misrepresent reality, which makes it more difficult for individuals to flag the inauthenticity. Relatedly, the auditory or visual nature of some types of generated content can have a stronger emotional impact on individuals compared to text-based misinformation, which makes the message more memorable and persuasive. The proliferation of generative AI technology also makes it faster, cheaper, and easier—even in real-time—for anyone to produce sophisticated content that was previously only possible for well-resourced organizations or state actors. These differences make deepfakes a particularly

potent tool for [disinformation and manipulation](#), posing new challenges for maintaining the integrity of public discourse and democratic processes.

Beyond the direct manipulative potential of AI-generated content, another concern about deepfakes is not even deepfakes themselves but the way in which the proliferation of deepfakes creates dubiousness about information truth more generally. If people come to think that anything could be untrue, in other words, they might not trust that anything is true, leading to an erosion of information sources. The so-called “[liar’s dividend](#),” in which inauthentic information is so prevalent that deniability of reality becomes more plausible, becomes more of a problem as the existence of deepfakes undermines trust in genuine media and information.

Government officials across the world have taken the risks of generative AI seriously. The United States’ Cybersecurity and Infrastructure, for example, has cautioned that generative AI “may amplify existing risks to election infrastructure.” Wired magazine, which covers tech, called 2024 “[the Year of the Generative AI Election](#).” While the theoretical risk for harm exists, whether through text or, more likely, through images, have the experiences actually validated the concerns? The next section offers cautious optimism, suggesting that the use or rather misuse of generative AI has been limited despite the widespread availability of the technology.

Taking Stock of 2024 Elections

Warnings about the potential influence of generative AI and in particular deepfakes in the 2024 elections come from both the potential use case but also the rapid advancement of AI technology in the

last couple of years that has made the technology more accessible in previous major election cycles. Those warnings have not been entirely inflated.

Deepfakes have been used to target specific political figures in the 2024 elections in several different ways. AI-generated audio and video have been used to create fake recordings of politicians. In the Democratic primary in [New Hampshire](#), AI-generated audio was used in robocalls, with Biden's voice urging voters not to vote. Biden still won the New Hampshire primary handily despite the deepfake attempt to discourage voting.

In [Slovakia](#), a deepfake audio falsely attributed plans to rig an election to a political leader in advance of the 2023 parliamentary election. AI-generated images and videos have also been used to misrepresent politicians. AI-generated pictures [showed](#) Donald Trump with Black voters in ways intended to endear the constituency to Donald Trump. During the primary season in 2023, a political action committee associated with Florida Governor Ron DeSantis [used](#) AI-generated audio of Donald Trump to portray him attacking the Iowa Governor, intended to paint him as disrespectful of the caucus state. Other deepfakes have been used to create embarrassing or compromising content. For example, in the UK, an investigation [uncovered](#) 400 instances of digitally altered deep fakes showing 30 high-profile UK politicians in compromising situations.

In [Poland](#), the opposition party Civil Platform (Platforma Obywatelska) created a deepfake video imitating the voice of the Prime Minister. The content was based on leaked emails from the prime minister's chief of staff and alternated between showing real video clips of the Prime Minister speaking and AI-generated audio clips that read sections of the leaked emails.



» The auditory or visual nature of some types of generated content can have a stronger emotional impact on individuals compared to text-based misinformation. «



The apparent motivation was to contrast the Prime Minister's public statements about unity within the ruling coalition with private messages that acknowledged tensions in the government. Only after skepticism and criticism did the platform acknowledge that the content was AI-generated.

In the UK, the first day of the Labour Party conference in Liverpool saw the release of a deepfake audio clip showing

Keir Starmer verbally abusing and using profanity towards his staff members and another showing him criticizing the city of Liverpool. The clips [posted](#) on X by an account with less than 3,000 followers, receiving 1.4 million views. One deepfake detection company, Reality Defender, [indicated](#) that the audio was 75% likely manipulated and the British government's analysis also [confirmed](#) the inauthenticity of the content.

While voices across the political spectrum criticized the audio, it raised concerns about the threat of deepfakes to democracy and highlighted the challenges with debunking content.

The creation of deepfakes is certainly not limited to domestic politics. Russia has repeatedly been accused of creating deepfakes of the pro-Western Moldovan President Maia Sandu to ridicule the leader and undermine her credibility. CopyCop, a suspected Russian-aligned influence network, has used AI and inauthentic sites to create and distribute disinformation. Reports [suggest](#) little engagement or amplification on social media, however.

Although these examples point to ways that groups or individuals have used AI-generated deepfakes, the 2024 elections have also been notable for the absence of deepfakes. In Mexico's 2024 election, deepfakes did not feature at all and indeed, the election appears to have been conducted [without](#) major disruptive incidents, although outside actors had little probable cause.

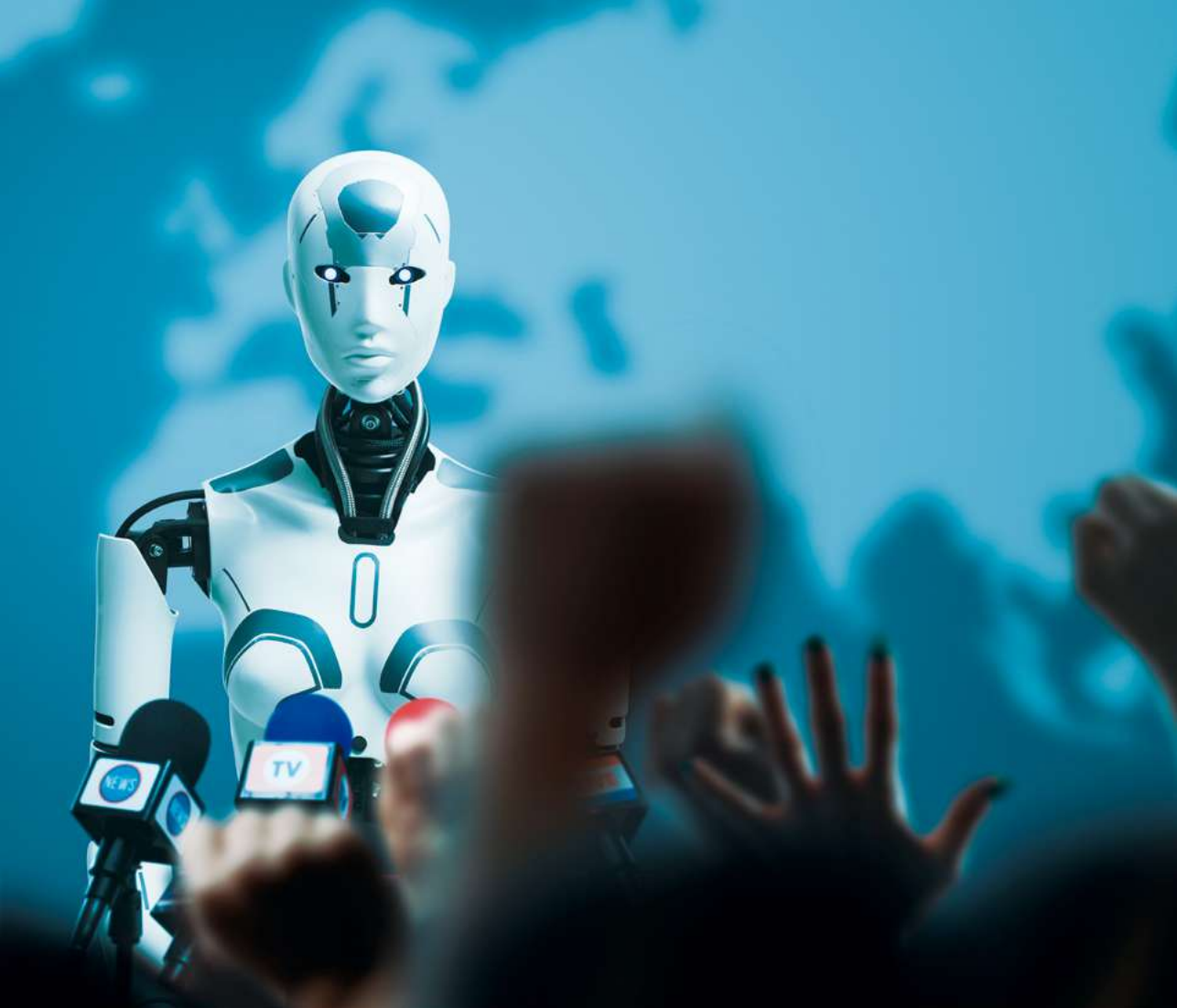
Despite the availability of deepfakes and tensions between Taiwan and China, its use was limited. China appeared to use AI-generated audio clips to [target](#) the Democratic Progressive Party presidential candidate Lai Ching-te. The AI-generated content included manipulated video content of the candidate, with audio appearing to be Lai talking about scandals that had not occurred and supporting a coalition with Kuomintang (KMT), the Chinese nationalist party that ruled until it was defeated by the Communist party of China on the mainland.

But more than deepfakes, China relied on other misinformation techniques. China pushed false or misleading stories on social media such as portraying the



» If people come to think that anything could be untrue, in other words, they might not trust that anything is true. «

United States as an unreliable ally that would abandon Taiwan, framing the election as a choice between “peace” (unifying with China) and “war” (continued independence), spreading false claims about U.S. biological labs in Taiwan,



promoting conspiracy theories about CIA interference in the elections, and promoting racist narratives against migrant workers.

China has become associated with techniques such as “[spamouflage](#)”, in which Chinese government-affiliated groups use accounts to actively promote particular narratives on social media. For example, these accounts aim to portray the United States in a negative light by highlighting urban decay, police brutality, and deteriorating infrastructure. These accounts are particularly active during events such as a natural disaster or election. In

April 2023, the US Department of Justice [charged](#) 40 employees of the Chinese Ministry of Public Security’s 912 Special Projects Working Group for their involvement in an influence campaign that appeared to be Spamouflage.

In addition to the continued spamouflage efforts, China not only distributed these messages but then relied on [Taiwanese proxies](#) to spread disinformation to make it more difficult to trace back to China. Beyond the immediate election outcomes, China appeared focused on eroding trust in Taiwan’s democracy and sovereignty over time.

But in no way were the deepfakes prevalent or did they appear to be effective. The same is true in the 2024 Indian election. Despite the concerns about widespread deepfake use, the actual number of verified AI-generated misinformation cases was relatively low. Out of 258 election-related fact-checks conducted by [Boom Live](#), only 12 involved AI-generated misinformation. Isolated cases included using deepfake technology to “resurrect” dead politicians for campaign purposes. Both major parties, BJP and Congress, both created and shared AI-generated content such as memes and satirical videos or AI-translated speeches or personalized AI robocalls. One study showed that of about 2000 viral WhatsApp messages, only 1% were generated by AI, a small footprint according to [Nature](#). The prevalence and impact of misused AI-generated content was limited, certainly less than initially feared.

Why Has AI-Generated Content Been Low Impact?

Scholars have shown proofs of concept for how malicious actors could use AI-generated misinformation at scale to disrupt democratic elections and yet there is a dearth of evidence to suggest that actors are either producing such content or having an impact on elections, which raises the question of why the predictions have been at odds with reality (Pawelec 2022).

One possibility for the limited effectiveness is that the technology is still nascent, particularly with respect to deepfake videos or audio. Users quickly identify and debunk the images or video in part because deepfakes are still discernible. The incorporation of video and audio sometimes creates mismatched synchronization in how words are projected,

which means attention to the lip movements can highlight inconsistencies. Relatedly, there are often subtle inconsistencies between the AI and the human likeness, much as were obvious with the Tom Cruise [deepfakes](#). Individuals viewing these images might experience a skeptical emotional response referred to as “[the uncanny valley](#)” when viewing a robotic image that is subtly not lifelike.

Another set of factors is that while the technology is nascent, individuals, political leaders, states, and social media platforms are actually [prepared](#) for these deepfakes, helping to neutralize the effect. Government officials, in some cases, have conducted simulations and tabletop exercises to respond to deepfakes. Some states have [passed](#) laws regulating the use of deepfakes in political campaigns, which may deter some potential bad actors. Algorithmic detection is [improving](#) in ways that allow social media platforms to flag and remove deepfakes, many of which have been banned by these platforms.

Another consideration, as the example of China in the 2024 Taiwanese election suggests, actors may simply see other forms of disinformation more practical or effective, such as appearing in online forums with particular perspectives or political valences to shape what people read.

Further, political persuasion is difficult. Research suggests that the effect of misinformation or disinformation (with the intention to mislead) tends to have limited impacts because people’s views are fairly entrenched. Indeed, studies of whether or how misinformation has affected political behavior often [produce](#) null findings because people tend not to change their minds, even when faced with the more viscerally powerful images (versus text).

» Democracies must foster a culture of critical thinking and media literacy among their citizens. «

Conclusion

Despite their potential for disruption, deepfakes and AI-generated content more generally have not yet been as prevalent or consequential during the 2024 election waves. Efforts to both identify and remove deepfakes have grown on the part of individuals, governments, and platforms. Impact on public opinion has been limited. The measures that society has adopted to guard against deepfakes have mitigated their impact but past success in mitigating the influence of deepfakes does not guarantee future immunity.

Nonetheless, as AI evolves, so will the efforts to manipulate democratic politics for electoral advantage. Public awareness and media literacy will continue to play an important role in reducing the impact. As the technology changes, so too should the campaigns by governments, non-profit organizations, and media outlets to highlight the type of critical thinking and skepticism needed to consume digital content to protect against manipulation and misinformation. News outlets will need to continue rigorous fact-checking to verify the authenticity of visual or audio content before publication to prevent the spread of deepfakes and prop up trust in the media. Regulatory and legal frameworks will need to stay current and should continue to evaluate the new technologies and ask what types of deepfakes are free speech and which ones should be prohibited.

This analysis produces an additional

cautionary note. As the Chinese online influence in the Taiwanese election suggests, actors seeking to influence public opinion will continue to use other means such as armies of online trolls or groups that engage in spamouflage, creating messages aiming to manipulate political opinion. Preoccupation with deepfakes may obscure those approaches and worse, divert resources and attention not just from those more established mechanisms that include phishing, ransomware, and other cyber threats. Security measures, funding, and research efforts might become disproportionately skewed towards combating deepfakes at the expense of broader cybersecurity initiatives, an opportunity cost in terms of time and effort invested in deepfake detection and prevention. Overemphasis on deepfakes may contribute to a self-fulfilling process, however, in the form of eroding trust in media and information by inculcating skepticism in the public and casting doubt on legitimate and authentic content.

Ongoing advancements in detection technologies, increased public awareness, and robust legal frameworks have proven effective in mitigating many of the threats posed by deepfakes. However, the dynamic nature of technology and the ever-evolving tactics of malicious actors require continuous vigilance and adaptation. Democracies must foster a culture of critical thinking and media literacy among their citizens while maintaining transparency and accountability in their institutions. By doing so, they can safeguard their integrity and continue to thrive in the digital age, turning potential vulnerabilities into strengths and ensuring that their foundational values endure against the challenges of modern technology. ♦

References

- 40 [Officers of China's National Police Charged in Transnational Repression Schemes Targeting U.S. Residents.](#) (2023, April 17). Office of Public Affairs.
- ADAM, D. (2024, June, 18). [Misinformation might sway elections — but not in the way that you think.](#) *Nature*.
- BICKERTON, J. (2023, March 24). [Deepfakes Could Destroy the 2024 Election.](#) *Newsweek*.
- CAHLAN, S. (2020, February 13). [How misinformation helped spark an attempted coup in Gabon.](#) *The Washington Post*.
- DE VYNCK, G. (2024, April 5). [The AI deepfake apocalypse is here. These are the ideas for fighting it.](#) *The Washington Post*.
- [Deepfake audio of Sir Keir Starmer released on first day of Labour conference. \(2023, October 9\).](#) *Sky News*.
- DEVINE, C., O'SULLIVAN, D., & LYNGAAS, S. (2024, February 1). [A fake recording of a candidate saying he'd rigged the election went viral. Experts say it's only the beginning.](#) *CNN*.
- ELLIOT, V. (2024, May 30). [2024 Is the Year of the Generative AI Election.](#) *Wired*.
- ELLISON, S., & WINGETT SANCHEZ, Y. (2024, May 8). [In Arizona, election workers trained with deepfakes to prepare for 2024.](#) *The Washington Post*.
- EWEL, K. (2023, December 28). [The Ultimate Election Year: All the Elections Around the World in 2024.](#) *Time*.
- FISHER, M. (2022, July 21). [How I Became the Fake Tom Cruise.](#) *Hollywood Reporter*.
- GARIMELLA, K. & CHAUCHARD, S. (2024, June 5). [How prevalent is AI misinformation? What our studies in India show so far.](#) *Nature*.
- GILLIS, A. (2024, February). [uncanny valley.](#) *Techtarget*.
- GOLDSTEIN, J., CHAO, J., GROSSMAN, S., STAMOS, A., & TOMZ, M. (2024, February 20). [How persuasive is AI-generated propaganda?](#) *PNAS Nexus*, 3(2).
- [Here's How Deepfakes, Like the One BJP Used, Twist the Truth.](#) (2020, February 20). *VICE*.
- HUNG, CH-L., FU, W.-CH., LIU, CH-C., & TSA, H-J. (2024). [AI Disinformation Attacks and Taiwan's Responses during the 2024 Presidential Election.](#) Thomson Foundation.
- INSIKT GROUP. (2024, June 24). [Russia-Linked CopyCop Expands to Cover US Elections, Target Political Leaders.](#) *Recorded Future*.
- ISENSTADT, A. (2023, July 17). [DeSantis PAC uses AI-generated Trump voice in ad attacking ex-president.](#) *Politico*.
- IYENGAR, R. (2024, January 23). [How China Exploited Taiwan's Election—and What It Could Do Next.](#) *FP*.
- JACKSON, K., SCHIFF, D., & BUENO, N. (2024, February 20). [The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?](#) *American Political Science Review, First View*, pp. 1-20.
- JACOB, N. (2024, June 3). [2024 Elections Report: Fake Polls, Cheap Voice Clones, Communal Claims Go Viral.](#) *Boom*.
- KREPS, S., et al. (2020, November). [All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation.](#) *Journal of Experimental Political Science*, 9(1), 1-14.
- KREPS, S., & KRINER, D. (2023, October). [How AI Threatens Democracy.](#) *Journal of Democracy*.
- KREPS, S., & KRINER, D. L. (2023). [The Potential Impact of Emerging Technologies on Democratic Representation: Evidence from a Field Experiment.](#) *New Media & Society*.

MARTIN, A. (2023, October 9). [UK opposition leader targeted by AI-generated fake audio smear.](#) *The Record*.

MARTINAU, K. (2023, April 20). [What is generative AI?](#) *IBM*.

MORGAN, L. (2024, July 2). [Deepfake pornography is being used to humiliate and silence powerful female politicians like Angela Rayner and Penny Mordaunt. Why doesn't the law protect them?](#) *Glamour*.

[Opposition criticised for using AI-generated deepfake voice of PM in Polish election ad](#) (2023, August 25). *Notes from Poland*.

PAWELEC, M. (2022, September). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, 1(2).

POLGÁR, J., & WEN, A. (2024, October 10). [Deceptive Audio or Visual Media \('Deepfakes'\) 2024 Legislation.](#) NCSL.

PRUNEDA, P., & SALAZAR UGARTE, P. (2024, May 29). [Elections in Mexico: Beyond "Deepfakes".](#) *Wilson Center*.

SAINATO, M. (2024, March 4). [AI-generated images of Trump with Black voters being spread by supporters.](#) *The Guardian*.

SWENSON, A., & WEISSERT, W. (2024, January 23). [New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary.](#) *AP*.

Tsu, T. (2024, February 15). [Chinese Influence Campaign Pushes Disunity Before U.S. Election, Study Says.](#) *The New York Times*.

Verma, P. (2023, December 17). [The rise of AI fake news is creating a "misinformation superspreader".](#) *The Washington Post*.

YASAR, K. (2024). [What is deepfake technology?](#) *Techtarget*.

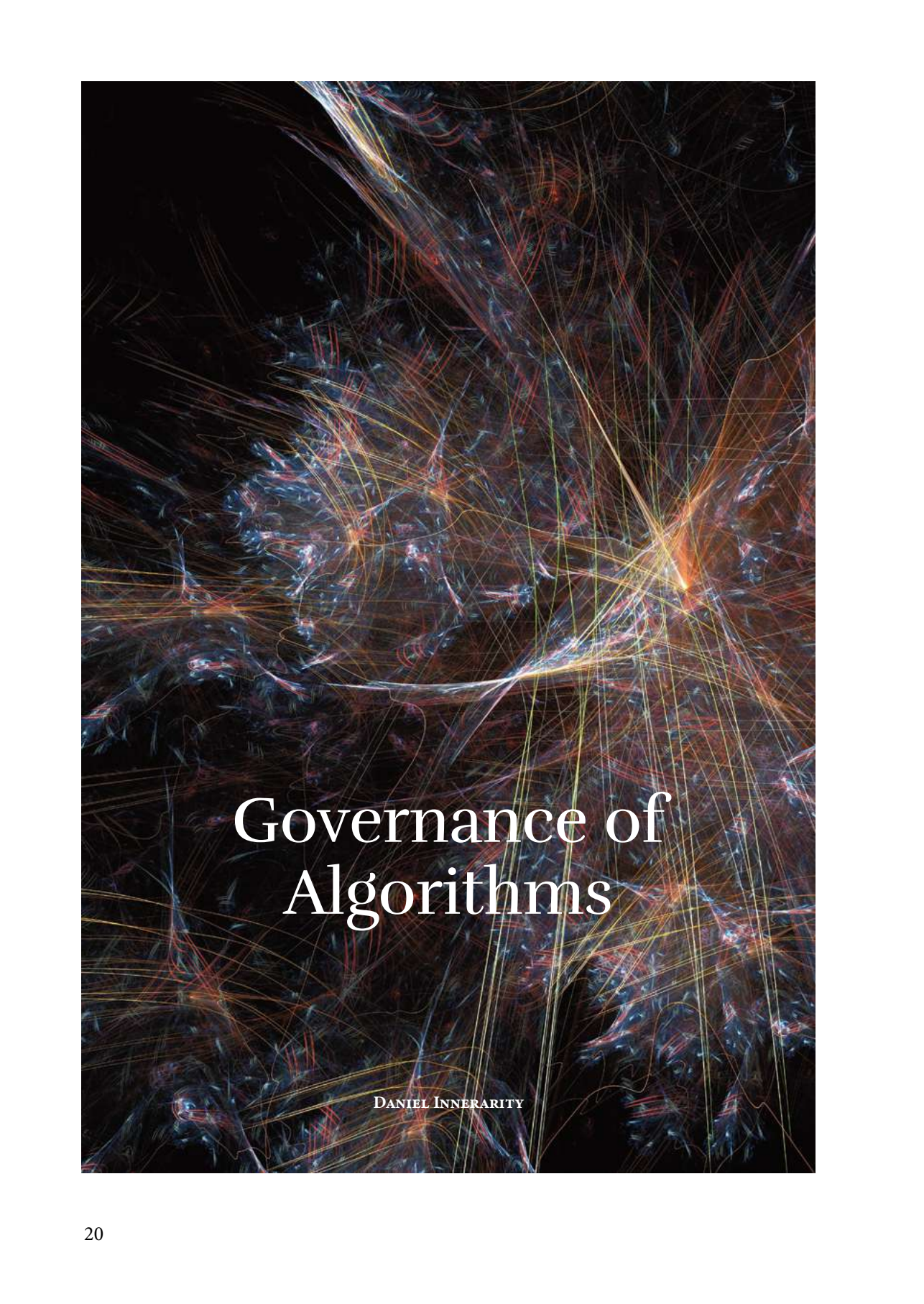


Sarah E. Kreps

US political scientist. Air Force veteran. Policy analyst on US foreign and defence policy. Professor of government at Cornell University and adjunct professor at West Point's Modern Warfare Institute. Author of books on drones, US military interventions and how the US funds its wars. Life member of the Council on Foreign Relations.

X: @sekreps

LI: sarah-kreps-51a3b7257

The background of the cover is a complex, abstract pattern of fine, glowing lines and fibers. These lines are primarily in shades of blue, red, and yellow, with some white highlights. They are set against a dark, almost black, background. The lines appear to be interconnected, forming a dense, web-like structure that fills the entire frame. The overall effect is one of dynamic energy and complexity.

Governance of Algorithms

DANIEL INNERARITY

One issue is how to govern algorithms, and another is whether algorithms will eventually govern us, to what extent, and with what legitimacy. To address this second issue, we need to examine the expectations of algorithmic governance and its limitations. As a result, it seems unlikely that algorithms can take over the entire political process with the efficiency they sometimes promise and with the legitimacy needed to justify such a new regime.

From bureaucracy to algorithmic governance

Once a political community reaches a certain level of complexity, the need to objectify and automate collective decisions naturally arises. When the number of actors and factors involved exceeds individual and centralized capacities, decision-making becomes more procedural and less charismatic.

When an incompatibility is presented between any kind of standardized decisions and humanistic considerations, it's important to remember that these procedures were designed specifically to minimize human intervention in decision-making. Porter referred to the culture of quantification as the cult of impersonality, where the human element is minimized as much as possible: formalizable principles rather than subjective interpretations, unified standards instead of methodological chaos, and the rule of law over human power. In this new realm of objectivity, *mechanical objectivity* and disinterested science would reign, leaving out anything personal, idiosyncratic, or perspective-based; trust is no longer rooted in the integrity of truth-tellers or the prestige of exemplary institutions, but rather in highly standardized procedures (Porter, 1995). The most radical formula to express

it could be this: "Instead of freedom of will, machines would offer freedom from will" (Daston and Galison, 2010, p. 49). This hope for data and objectivity grows in a political and social culture marked by distrust, crises, and uncertainty; turning to a form of objectivity benefits both governors and the governed, protects decision-makers, and fosters confidence among those affected by their decisions.

The digital era has intensified this long-standing trend. Governing is already largely—and it will become even more so — an algorithmic act; a significant portion of government decisions is made by automated systems. This method of governance has been defined in various ways: "power is increasingly in the algorithms" (Lash, 2007, p. 71); "authority is increasingly expressed algorithmically" (Pasquale, 2015, p. 1).

The use of algorithms and automated decisions addresses the need to manage various forms of complexity, such as identifying the different perspectives and interests within an increasingly pluralistic society, as well as efficiently delivering public services. Algorithmic governance significantly improves management capabilities when handling large volumes of data and addressing complex problems. Thus, not only does the world appear to have become more unders-

tandable, but new possibilities for political intervention, increased efficiency, smarter regulation, and earlier anticipation of certain problems have also emerged. This promises a form of governance that would simplify the complexity of social phenomena to an acceptable level.

The rise of decision systems driven by algorithms and data means that machines not only support humans in their decision-making but can also replace them, either partially or entirely. The question raised by all this is to what extent and in what way the use of automated decision systems (ADS) is compatible with what we consider a political system of decision-making. Democracy is expected to fulfill the belief that it is a genuine form of self-governance by the people while also effectively addressing the problems faced by society.

The democratic expectations of algorithmic governance

Algorithms make a dual promise of objectivity and subjectivity, offering both ideological neutrality and, simultaneously, complete respect for our preferences. These two promises have very beneficial effects on democratic politics, as they enable a more objective assessment of public policies and a better understanding of social preferences. However, they also come with their limits and drawbacks.

The promise of objectivity

The promise of algorithmic decision-making is highly seductive; it is not merely about saving time and money, but about promoting objectivity. Algorithms are often seen as objective, with their evaluations consid-

ered fair, accurate, and free from subjectivity, errors, and power dynamics. Furthermore, this perceived objectivity lends them legitimacy as mediators of relevant knowledge. They are not only tools for decision-making but also stabilizers of trust, ensuring that “assessments are accurate and fair, without flaws, subjectivity, or distortions” (Gillespie, 2014, p. 79). The implementation of automated decision systems (ADS) is justified because they not only make decisions more efficiently but also reduce partisanship and enhance fairness. We would have tools that appear to fulfill the hope of bringing greater rationality to the decision-making process, counteracting the subjectivity, ideological biases, or other prejudices that often drive many human decisions.

This claim is not entirely new, nor is its criticism. Weber’s idea of bureaucratic authority had already praised the values of efficiency and objectivity, but he had also warned of their limits and that other types of authority could arise precisely because of the ideal of objectivity. In principle, all the pathological tendencies of traditional bureaucracies also apply to automated decisions. Ever since claims of objectivity were formulated, both in bureaucratic settings and the digital era, it has been consistently observed that such procedures fail to deliver on that promise, generating other types of distortions, being far from free of arbitrariness, and that algorithms often reflect, and even amplify, deeply rooted societal prejudices.

The promise of subjectivity

The second vector of democratization would stem from understanding the true will of the people, which a democratic government must serve. The chain of legitimization

would thereby be strengthened, as it would allow the real decisions of the people to serve as the foundation upon which the popular will is formed. In a world filled with sensors, algorithms, data, and intelligent objects, a kind of social sensorium is being shaped to personalize health, transportation, and energy. Thanks to data engineering, we are moving towards an increasingly granular understanding of individual interactions and systems that are more responsive to individual needs. By using micro-segmentation and granularity, we can shape a society finely tuned by algorithms, enabling us to understand citizens' desires with remarkable accuracy based on their everyday behaviors. The objectivity of algorithmic governance methods would be accompanied by greater subjectivity in its recipients, who would thereby see their individuality more thoroughly understood, respected, and fulfilled.

The comfortable paternalism of algorithmic societies lies in the fact that it gives people what they want, governs with proportionate incentives, and proceeds by inviting, suggesting, and guiding. Transferring this model to politics would not encounter major issues were it not for the fact that the price of these benefits is often the sacrifice of some aspect of personal freedom. Considering the discrepancy between the self-determination we claim to demand and the self-determination we are actually willing to exercise when comforts and benefits are at stake, the outcome is that the satisfaction of needs often comes at the cost of sacrificing spaces of freedom. It is true that many of our desires are satisfied in this way, but at the cost of a certain renunciation of reflecting on them; what we want takes precedence over what we want to want, and the minimal, implicit will of the consumer replaces the explicit political will.

» Political issues are those that can only be resolved through value judgments. «

The democratic limitations of algorithmic governance

Algorithmic governance is well suited to enhance certain aspects of the policy process, but it is of little use for others; it can correct human deficiencies and biases, identify preferences, and measure impacts. However, it is inadequate for dimensions of the political process that are not easily subject to computation and optimization—areas that are difficult to quantify and measure. This includes the genuinely democratic moments when the criteria and objectives that technology can later optimize are determined. The reason algorithms are politically limited stems from their instrumental nature. Algorithms are designed to achieve predetermined objectives, but they contribute little to determining those objectives, which is the responsibility of political will, democratic reflection, and deliberation. The role of politics is to determine the design of algorithmic optimization strategies and to consistently preserve the option to alter them, particularly in dynamic environments. In a democracy, everything must be open to moments of re-politicization, meaning there must be the possibility to question established objectives, priorities, and means. This is the purpose of politics and not of algorithms. Algorithmically optimized governance lacks the capacity to resolve genuine political conflicts or address the political dimensions of those conflicts, particularly when frameworks, ends, or values are involved. As Lucy Suchman noted in another context,

robots perform very well when the world has been organized as it was intended to be (Suchman, 2007).

This duality of ends and means, of political goals and algorithmic optimization strategies, can be illustrated by the student distribution system implemented for New York City schools and the ensuing debate regarding which values should be prioritized in that distribution (Krüger and Lischka, 2018). The system can prioritize the maximum satisfaction of individual preferences or a balanced social mix within schools. Both objectives have valid reasons supporting them; one option emphasizes individual desires, while the other promotes social cohesion. It's also up for debate what level of compromise or balance between the two values is most desirable and achievable if they're to be respected at the same time. To determine this, a political debate about values and the involvement of those affected is necessary—a discussion from which an algorithm cannot absolve us.

In this and similar cases, the issue is not merely about the implementation or transparency of the algorithms used, but about the value judgments involved in defining the objectives of education, which are diverse and sometimes conflicting, as one would expect in a pluralistic society. Political negotiation processes take priority over technical solutions, and technical solutions cannot replace the need for political negotiation. We are, therefore, addressing what we refer to as political issues.

Strictly speaking, political issues are those that can only be resolved through value judgments, while technical issues involve deciding on the implementation of intended objectives based on available knowledge. At times, it is also politically controversial what kind of optimization is considered satisfactory and which kinds



» The decision about what is computable cannot itself be computed. «

of knowledge are deemed relevant. It could even be argued that if optimization as a principle is desirable, the ideology of optimization—believing that the effective implementation of certain objectives can render political discussion about those objectives unnecessary—may serve as a strategy for depoliticization.

Algorithmic governance seeks to achieve objectives that have not been debated, and which it neither establishes nor questions. However, democratic politics



is not merely about processing information but interpreting it within a framework of guaranteed pluralism. It is not just about how best to achieve certain objectives but about how to decide upon them. The politics begins where the debate arises about what algorithms should satisfy, which values they should uphold, and what conception of fairness they should serve. This idea can be expressed by recalling John von Neumann's statement: we can build an instrument capable of doing everything that can be done, but we cannot build an instrument that tells us whether something can be done (Neumann, 1966, p. 51). In other words: the decision about what is computable cannot itself be computed.

» The great promise of algorithmic governance is that optimal results will lead us to forget the desired procedures. «

As in politics in general, when we talk about algorithmic governance, the notion of producing better decisions with the help of machines still requires a prior criterion for what constitutes a good decision. The tools responsible for optimizing decisions do not eliminate the need to discuss what constitutes a good decision. It is true that artificial intelligence aids in informing decisions and optimizing outcomes, but while some economists have tried to quantify and measure aggregate welfare, there is no predefined or uncontested notion of what constitutes a successful political outcome.

The great promise of algorithmic governance is that optimal results will lead us to forget the desired procedures. It is a type of governance that appears to prioritize effectiveness, even at the cost of excluding us from decision-making or reducing our role to a minimal, implicit, and individual presence, reflected in the form of requirements and preferences found in our digital footprints. If citizens are unable to oversee or influence algorithmic decisions, we can't truly call it self-government.

The inevitability of deciding

The great challenge of the digital era is to resist the allure of depoliticizing our societies and overcome the inertia of traditional governance methods. We must avoid being seduced by falsely apolitical or post-ideological rhetoric while also moving away from practices

» The ultimate legitimacy (of a society) doesn't stem from the correctness of its decisions, but from the decision-making power of the citizenry—regardless of how well or poorly that power is exercised «

that no longer align with new social realities. We are facing an attempt to conceptualize society in a depoliticized manner.

Contemporary societies require significant cognitive mobilization to address the problems they face, but the ultimate argument in favor of democracy is not epistemic, but decisional. Everything possible must be done to ensure that societies make the best decisions, but their ultimate legitimacy does not stem from the correctness of those decisions. Instead, it comes from the decision-making power of citizens, regardless of how well or poorly they use that power. Democracy tends to produce better decisions than alternative models, but its ultimate legitimacy comes not from the quality of those decisions, but from the popular authorization behind them. The need to make decisions is the core justification for democracy—a form of government where ordinary people have the final say over experts. There appears to be no technological device today that can entirely free us from the need to make decisions.

Artificial intelligence procedures cannot absolve us of that decision. Politics, exists where, despite all the sophistication of calculations, we are ultimately driven to make decisions that aren't backed by

overwhelming reasons or guided by infallible technologies. A humane world must be a negotiable world.♦

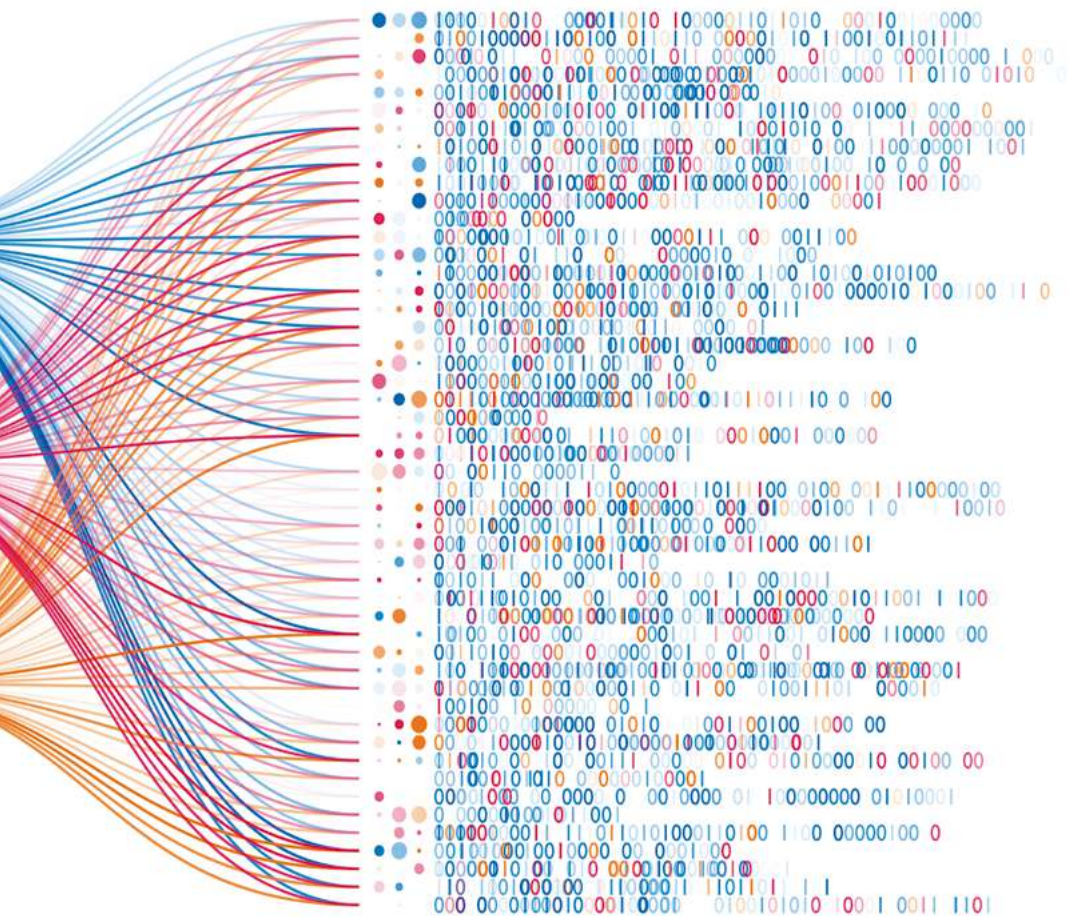
References

- DASTON, L., & GALISON, P. (2010). *Objectivity*. Princeton University Press.
- GILLESPIE, T. (2014). The Relevance of Algorithms. In T. GILLESPIE, P. J. BOCZKOWSKI & K. A. FOOT (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-193) Cambridge: The MIT Press.
- KRÜGER, J., & LISCHKA, K. (2018). Was zu

00000|100||10000||

00|000||100|0000||

00000|00000000|100



tunist, damit Maschinen den Menschen dienen. In R. MOHABBAT KAR, B. THAPA & P. PARYCEK (eds.), *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft* (pp. 440-470). Berlin: Fraunhofer-Institut für Offene Kommunikationssysteme fokus.

LASH, S. (2007). Power after hegemony. *Theory, Culture & Society*, 24(3), 55-78.

NEUMANN, J. VON (1966). *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press.

PASQUALE, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge: Harvard University Press.

PORTER, T. (1995). *Trust in numbers: The*

pursuit of objectivity in science and public life. Princeton University Press.

SUCHMAN, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge University Press.



Daniel Innerarity

Professor of Political Philosophy, Ikerbasque researcher at the University of the Basque Country, and holder of the Chair of Artificial Intelligence

and Democracy at the European Institute of Florence.

X: @daniinnerarity

LI: daniel-innerarity-406665112



Anticipatory governance with the help of technology

LYDIA GARRIDO LUZARDO

The transformative power and disruptive potential of AI requires ethical governance. Responsible anticipation and literate use of the future are key elements in policymaking. Parliamentary committees of the future offer recommendations for inclusive governance based on cooperation, transparency and collective intelligence.

The evolution of artificial intelligence (AI) has reached a turning point where its transformative and disruptive capabilities demand a profound assessment of how to govern it ethically and responsibly. Artificial general intelligence (AGI) refers to a type of AI capable of performing any human intellectual task. AGI poses challenges that would appear to call for different epistemological and methodological frameworks for its governance. These frameworks must not simply react to technological transformations once they have occurred, but rather explore the nature of the relationships that give rise to the invention and application of tools, taking into account the way dominant systems define and explore opportunities. To that end, a broader understanding of the attributes and relationships of anticipatory systems complexity is needed, one that integrates the epistemologies of collective intelligence with ethical values and anticipatory capacities, all rooted in a theory of anticipation that helps to clarify both why and how we imagine the future. Such a theory, the ‘discipline of anticipation,’ is what underpins efforts to enhance anticipatory capabilities and alter the conditions within which governance systems and practices function.

This article argues that effective anticipatory governance requires an approach based on anticipatory capabilities and ethical principles. Responsible anticipation is thus presented as both a capability and as an essential quality for a future oriented, proactive, responsive decision-making in order to create the conditions for an ethical development of AI for the common good of society. It is crucial to establish clear principles to orient AI evolution towards a safe and ethical agi and to adopt an approach that re-orientes the generative side of human agency towards an integration of futures imagined within an ethical frame that steers away from oppression, extractivism, and exploitation.

This text integrates elements of ethics, complexity, and *use of the future* into decision-making, applying them to the context of anticipatory governance for AI. It also draws on the experience of the Special Futures Committee of the Uruguay Parliament and the recent contributions of the Second World Summit of the Committees of the Future in Parliaments that took place in Montevideo, Uruguay, in September 2023.

The article is structured into three main sections. First, it analyzes the evolutionary nature of AI and its disruptive potential. Second, it examines the challenges that the use

of the future poses in anticipatory governance practices. Finally, it discusses the practical considerations of responsible anticipatory governance for AI, emphasizing the crucial role of Parliaments and other institutions in designing flexible, anticipatory, and adaptive governance frameworks.

The Evolutionary Nature of AI and its Disruptive Scope

There are numerous definitions of *artificial intelligence*. Some debates focus on how AI differs from human intelligence, but this article does not revolve on that aspect. AI's disruptive potential goes beyond such similarities or differences with humans, in other words, beyond an anthropomorphic and anthropocentric perspective. Here, we focus on a powerful tool, that is, in its current and potential capabilities, a source of opportunities for the good of humanity that, at the same time, may present serious threats.

The other crucial aspect of our approach is to consider the evolutionary nature of AI (precisely what so often generates differences and lack of consensus on a single definition), since one of its inherent characteristics is its permanent state of change. In December 2023, the OECD revised its definition of artificial intelligence systems:

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment¹.

Simply put, AI is not static; it is constantly evolving, expanding its capabilities and transforming its relationship with the surrounding technological and social dynamics. Such evolutionary process, characterized by increasing autonomy and disruptive potential, poses complex challenges for decision-makers and once again draws attention to the inadequacy of familiar governance systems that are premised on the proposition that the future is technocratically knowable. AI is intertwined with other emergent technologies such as the Internet of Things (IoT), autonomous systems, robotics, biotechnology, nanotechnology, as well as cognitive and neurocognitive sciences, which might further amplify its impact. An example of such intertwining can be found in precision medicine, where AI, along with biotechnology and nanotechnology, enables personalized treatments that, in turn, raise ethical and privacy concerns requiring new forms of governance.

From its origin, AI proved to be a powerful tool for solving specific problems. However, we are now on the brink of leaping into artificial general intelligence, a kind of AI capable of performing general tasks like humans and even solving complex problems without specific pre-programming. This transition prompts essential questions about how to govern these capabilities and what kind of futures we wish to build with them.

Understanding the evolutionary nature of AI is crucial for designing anticipatory frameworks that might provide a more practical foundation for the exercise of human agency and futureproofing governance.

ving the notion of inputs and rules that may or may not be set by a human. Instead, the OECD recognizes that the system is capable of learning new objectives by itself.

1 One of the major adjustments to this definition was remo-

Artificial Narrow, General, and Superintelligence

We draw on the distinction made by the Millenium Project, Futures Studies and Research² to define different AI types.

AI governance must start with a fundamental distinction between three types of AI: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial superintelligence (ASI). This distinction is critical because each type of AI entails different risks, opportunities, and challenges.

- ANI (*Artificial Narrow Intelligence*) is the AI we have today, designed to carry out specific tasks, such as facial recognition, autonomous driving, or recommendation systems in streaming platforms. ANI comprises tools with limited capacities that cannot function outside the parameters for which they were designed. The progress of artificial generative intelligence (such as ChatGPT and similar AI systems) already suggests a transition into the next type or stage of AI.
- AGI (*Artificial General Intelligence*) marks a qualitative leap in AI evolution. Unlike ANI, AGI might have the capacity to solve non-specific problems, function in a wide range of contexts with no need of constant human intervention, adapt to new situations, and learn autonomously. This would turn AGI into an autonomous agent capable of acting in ways not anticipated by

» Understanding the evolutionary nature of AI is crucial for designing anticipatory frameworks that might provide a more practical foundation for the exercise of human agency and futureproofing governance.«

the humans who deployed it. AGI can therefore improve its own code and evolve rapidly, raising concerns about its control and the unforeseen consequences of its development. Current advancements suggest AGI is near (experts estimate it could be five to twenty years before AI reaches this stage), and its impact will be profound, since it will enable AI systems to act as autonomous agents with abilities comparable to—or above—those of humans.

- ASI (*Artificial Superintelligence*) refers to AI evolution after AGI, an intelligence so advanced that it could establish its own objectives and act entirely independently from humans. Even though ASI remains a speculation, the possibility of its emergence from AGI requires anticipatory governance contemplating not only the current stages of AI development but also its possible future ramifications.

2 A think tank with over seventy country nodes of futurist professionals that originated as an advisory group for the first Millenium Goals. It currently conducts a Delphi study on Global Artificial General Intelligence Governance. <https://www.millennium-project.org/seeking-delphi-on-ai-implications-and-regulation-april-9-2024/>

In short, while current AI presents challenges concerning algorithm transparency, fairness, and personal privacy, AGI and ASI would amplify these challenges, introducing existential questions

about control, autonomy, and potential influence of AI on the structure of society and humanity (Glenn & Garrido, 2023).

Relational Complexity and the Need for a Paradigmatic Shift

The evolutionary nature of AI, particularly its transition toward AGI and ASI, requires a paradigmatic shift in how we approach its governance. It is imperative to move beyond a linear, simplistic approach and adopt a framework based on relational complexity, one that recognizes the multiple interdependencies between technology and other social and natural systems (economic, cultural, environmental, etc.). Governance cannot be limited to reactive regulation of current issues or to ex-post intervention. Instead, it must be anticipatory and offer alternatives through generative and preventive actions in the present, thereby creating socially consensual conditions that consider the qualitative changes in society and the role that evolving technologies might play. Therefore, participatory, multi-, and transdisciplinary models are needed—for instance, reticular models based on collective intelligence and social participation.

This anticipatory governance approach not only addresses risks—it also explores how to integrate novelty (stemming from uncertainty) into perceptions and choices in the present. This integration covers not only the imagined costs and benefits, opportunities and threats posed by AI but also ways for decision-makers to navigate complexity and live with the certainty of surprises. Anticipatory governance is at once modest—it does not pretend to be able to know the future—and resourceful—leveraging constant experimentation and emergence. On the one hand, it stresses respecting the creativity of the world

around us; on the other hand, it promotes experiments that, very importantly, test our hopes of shaping transformation in advance (whether by avoiding or fostering desired circumstances).

The Special Futures Committee established by the Parliament of Uruguay is a clear example of this shift. By adopting a more complex and interconnected approach, the committee works to integrate various perspectives and introduce aspects of the future, thus driving actions toward an innovative anticipatory governance ecosystem.

Anticipatory Governance Challenges and Responsible Anticipation

This document reflects on anticipatory governance for AI through the lens of *responsible anticipation* practices. The future itself does not exist (Miller, 2018), which raises an ontological tribulation that must be considered within the epistemic dimension of the problem. Given AI's evolutionary nature, it is vital for governance strategies not to focus solely on current issues, but to encompass possible stages of AI development. However, a critical question arises: How can we assume responsibility for futures that do not yet exist? The following section considers this issue and proposes an approach to address it.

Ethics and the Responsibility Imperative

Responsible anticipation seeks to establish an ethical stance beyond purely philosophical approaches and traditional foresight or futurology. In this context, ethical practice is situated in a conscious and reflective conjunction that allows bringing the future into the present for both perception and choice. From this point of departure, the

role of imagined futures is not reduced to choices or bets about tomorrow; rather, it includes a crucial reflective ethical practice that brings the future into the present through the anticipatory assumptions used in the different stages of the decision-making process. Certainly, this approach involves considering the role of different futures in the initial problem formulation, leveraging our capacity to reframe through collective and deliberative reflections, detecting and inventing alternatives, and thus reaching the selection of options for actions (Garrido, 2024). A futures literate approach is fundamental, as the next section will explain.

Hans Jonas's pioneering work in the 1970s, still quite goal oriented as was so prevalent in the 20th Century, underscored the importance of incorporating future implications into contemporary ethical decisions. Jonas defied prevailing ethical standards by introducing a future-oriented framework that considers the impact of present actions on future generations and the environment. His concept of extended responsibility calls for an ethical reflection capable of addressing not only immediate consequences but also long-term effects. In his work, Jonas articulated the following *responsibility imperative*: "Act as if the effects of your action are compatible with the permanence of genuine human life on Earth." This imperative, deeply meaningful for our technological era, emphasizes the moral obligation to safeguard the dignity, autonomy, and integrity of present and future life.

Jonas not only revised the traditional ethical approach—he revitalized anticipatory thought in line with the Aristotelian concept of *final cause* as a practical and ethical guide for action. This perspective calls for a return to the fundamental principles of foresight and responsibility, both of which are crucial for preserving human

life in a context where natural and artificial systems are increasingly intertwined.

The notion of *responsible anticipation* proposed here encompasses an ethical mode of action considering the future consequences (consequentialist ethics), the moral obligations of our actions (deontological ethics), and the adoption of a proactive stance that promotes responsible care (virtue ethics). Each of these perspectives can be applied to both goal- and capability-based understandings of how humans anticipate, or 'use the future.' When the ends justify the means, there are clear ethical challenges—but so, too, when the means are the ends. Such a holistic approach is essential in fields such as health care, education, and, by extension, AI governance, where both the construction and vulnerability of the common good require an awareness of ethical imperatives

The moral structure of human beings displays our unique capability as reflective agents bearing the responsibility to make decisions that carry ethical implications. As Adela Cortina (2012) notes, human existence is inherently dramatic because of the constant need to make decisions and justify our actions. These dynamics of freedom, decision, and responsibility constitute the ethical axis of our actions, rendering us responsible not only for the immediate reach of our decisions but also for the futures we contribute to create.

Responsible anticipation answers the critical question of decision-makers: "What should I do now?" This query, central to Robert Rosen's (1985) approach, leads us to consider a paradigmatically different standpoint for decision-making. Rosen's conceptual framework of anticipatory systems provides new insights into how biological and social systems can make future-informed decisions, unlike reactive

systems, which merely respond to past stimuli.

At the cognitive level, anticipatory systems and assumptions generate information and knowledge to support decision-making. Explicit anticipation (Poli, 2010) is the conscious ability to incorporate or generate information about a subsequent moment with the intention of acting accordingly. Anticipatory assumptions are the concrete operative elements of this process. Paying attention to such assumptions allows for the exploration of the synergy between ethics, intention, and potential futures. These assumptions enrich both our theoretical understanding and the practical implementation of ethics. They provide a bridge between ethical theory and practice, offering a nuanced approach to responsible anticipation that integrates ethical deliberation with future-oriented thinking.

The Process of Using the Future in Decision-Making

Although the future itself does not yet exist, we use it every time we engage in anticipation. This is the abbreviated meaning of *using the future*, a concept encompassing the many purposes and forms of anticipation, including preparation, planning, and the exploration and creation of alternatives in the present (Miller, 2018, p. 10).

A literate use of the future requires an understanding of how we use that which does not yet exist (the future) to generate knowledge and inform decision-making. In this process, recognizing anticipatory systems and assumptions, as well as how they influence our perception of the present, becomes essential. At the same time, recognizing the contingent nature of the future—constantly shaped by a multitude of possi-

ble events and decisions—helps us better comprehend and relate to uncertainty and complexity.

This process involves not only imagining a range of potential futures—goals or scenarios ahead in time, i.e., substantivized futures—but also evaluating them in terms of desirability and viability. Moreover, it requires the recognition and selection of different anticipatory systems and assumptions (subjacent models) that take part in this perception and shape it. The models, therefore, must align with the purpose and the nature of the phenomena and problems.

In other words, this process is about understanding and dexterously using the systems and models that allow us to incorporate the not-yet-existent into our thinking, and this is achieved through reflexivity on the epistemic modes we employ, which can ultimately reshape how we see the present, and the opportunities and challenges we perceive (which may be biased, incomplete, or mistaken). As a result, decision-making is increasingly informed, nuanced, and aligned with ethical considerations of value, as there is greater dexterity in incorporating future into the analysis.

All of this may seem very abstract because it actually is. But we are referring to high cognition processes that enable anticipation (which is itself an action with practical implications). Furthermore, anticipation is inherently counterfactual since it impacts beforehand and can alter what is yet to happen—highlighting once more the need for new logics, methodologies, and skills.

Futures literacy is a crucial skill in this context, comparable to any other type of literacy (alphanumeric, computational, or emotional). Futures literacy enables individuals and organizations to introduce the future beyond the word future or the projections or extrapolations of the past (which is what is usually done). Instead, they can foster the

reflexivity and creativity needed to navigate uncertainty skillfully and responsibly, making sure their present actions are better informed.

In practice, this approach transforms policy design, allowing for the use of the future to become a powerful tool for anticipatory governance and responsible anticipation.

Following the work of Sripada (2016), we can distinguish two stages of decision-making processes: construction and selection. The construction stage involves creating meaningful options based on imagination and the exploration of future possibilities. This process is crucial to expand the set of possible alternatives, thus enriching decision-making. The selection stage, on the other hand, involves evaluating and appraising the options generated during the construction stage, ensuring that final decisions align with ethical principles and desirable futures.

Thus, responsible anticipation is not limited to mere prediction or responsibility for specific tasks. Rather, it implies a careful and reflective attitude throughout the whole decision-making process. From stating the problem and laying it out again, to achieving a deeper understanding—expressed in the ability to diversify alternatives and select the best options to transform beforehand—responsible anticipation enables decision-makers to act with a deep sense of ethical responsibility. This approach ensures that present decisions contribute to creating desirable futures while minimizing the inherent risks of uncertainty.

Applied Considerations for Anticipatory Governance for AI

Perhaps parliaments are the institutions that use the future with utmost intensity, and, therefore, with utmost responsibility to society

» This process is about understanding and dexterously using the systems and models that allow us to incorporate the not-yet-existent into our thinking. «

and humanity overall. Consequently, they play a critical role in ensuring a safe AI evolution: responsible AI, AI for the common good.

Through key functions such as accountability, supervision, representation, and legislation, parliaments have direct and concrete influence on the guidelines for AI development. That is why the Parliament of Uruguay created a Special Futures Committee, an innovative initiative that enables the government to traverse a learning curve for anticipatory governance. As a pluralistic setting for engaging with citizenry and other spheres and levels of government, it represents a great opportunity to spark an anticipatory governance ecosystem.

Regarding governance for AI, the imperative of a responsible anticipation practice is a condition *sine qua non*.

Recommendations for Anticipatory Governance for AI

The following recommendations were issued in the context of the Second World Summit of the Committees of the Future, held in Montevideo, Uruguay, in 2023.

1. *Devising an anticipatory governance framework for AI.* It is crucial to establish a global regulatory framework, coupled with international and regional guidelines. This framework should promote international

cooperation, ensure the ethical and responsible use of AI, and regulate its evolution to mitigate risks and maximize benefits.

2. *Promoting transparency and algorithm explicability.* Frameworks that require transparency in algorithm development are key to guarantee that decisions made by AI systems are comprehensible and auditable. Explicability is essential to avoid biases and foster public trust. Continuous auditing systems should be implemented to monitor AI behavior. Regarding the advancement of algorithm transparency and explicability, the European Union has established clear guidelines to ensure the responsibility and auditability of AI systems, setting a global standard in the field of technological regulation.

3. *Promoting inclusive and participatory governance.* The design of governance for AI should promote the inclusion of diverse stakeholders (governments, the private sector, civil society, and academia) and ensure that technology benefits society overall. Policies that guarantee fair access to emergent technologies should be prioritized to prevent technological gaps perpetuating inequality.

4. *Enhancing anticipatory capabilities.* Parliaments should develop anticipatory capabilities to manage AI evolution and prepare for disruptive changes. To this end, they should establish use-of-the-future specialized units and develop training programs for legislators regarding futures, AI, and complexity issues. Expanding such measures in a structured manner to other spheres of government and society is desirable, as this will create an anticipatory governance ecosystem.

5. *Using regulatory sandboxes.* The implementation of controlled experimentation environments (sandboxes) allows for iterative testing and adjustment of AI regulations. Sandboxes can enable a flexible adaptation to

» **Futures literacy, ethics of anticipation, and the development of anticipatory capabilities in decision-makers are key to ensure that AI evolves for the common good.** «

technological change, ensuring regulations evolve alongside technology.

6. *Adopting fundamental ethical principles.* Ethical principles such as transparency, fairness, privacy, and security, should be included in AI governance. These principles must be incorporated throughout the entire life cycle of AI systems, from design to implementation and use.

7. *Promoting AI education and literacy.* Developing educational and training programs on AI for legislators, citizens, and professionals of various sectors will promote a greater understanding of AI's risks and opportunities, preparing society to participate in governance processes.

8. *Fostering international cooperation on technological governance.* Fostering international cooperation and exchanging best practices between countries are key to address global AI challenges and promoting shared solutions applicable at the local and global level.

These recommendations seek to enhance the capabilities of parliaments and other government institutions, academia, developers, and civil society regarding anticipatory AI management. They also seek to promote anticipatory, participatory governance based on ethical principles that ensure the safe and beneficial use of this technology for the common good, encompassing both the social and the environmental spheres.

Toward a Responsible Anticipatory Governance

Anticipatory governance for AI is fundamental in an era of rapid and profound technological change. The development of agi and the potential advent of asi pose existential challenges that cannot be addressed with traditional governance approaches. Responsible anticipation, grounded in ethics and reflexivity, must guide the design of flexible and collaborative regulatory frameworks and allow for the management of risks and opportunities presented by AI's evolution.

The future of AI is yet to be written, and it hinges on the decisions we make today. Futures literacy, ethics of anticipation, and the development of anticipatory capabilities in decision-makers are key to ensure that AI evolves for the common good, guaranteeing its benefits reach society overall without compromising safety or dignity, in a beneficial and fair manner for all of humanity. ♦

References

- ARENDT, H. (2008) [1958]. *La condición humana*. Barcelona: Paidós.
- CORTINA, A. (2013). *¿Para qué sirve realmente la ética?* Madrid: Paidós.
- GARRIDO, L. (2024). Responsible Anticipation. Futures literacy capacities to enhance ethical stance in anticipatory governance decision-making. Learnings and applications in Parliaments. In T. FULLER et al. (ed). *Towards Principles for Responsible Futures*. Lincoln University, Taylor and Francis (in press).
- GLENN, J., & GARRIDO, L. (2023). *Parliaments and Artificial General Intelligence (AGI). An Anticipatory Governance*

Challenge. IDEA Internacional.

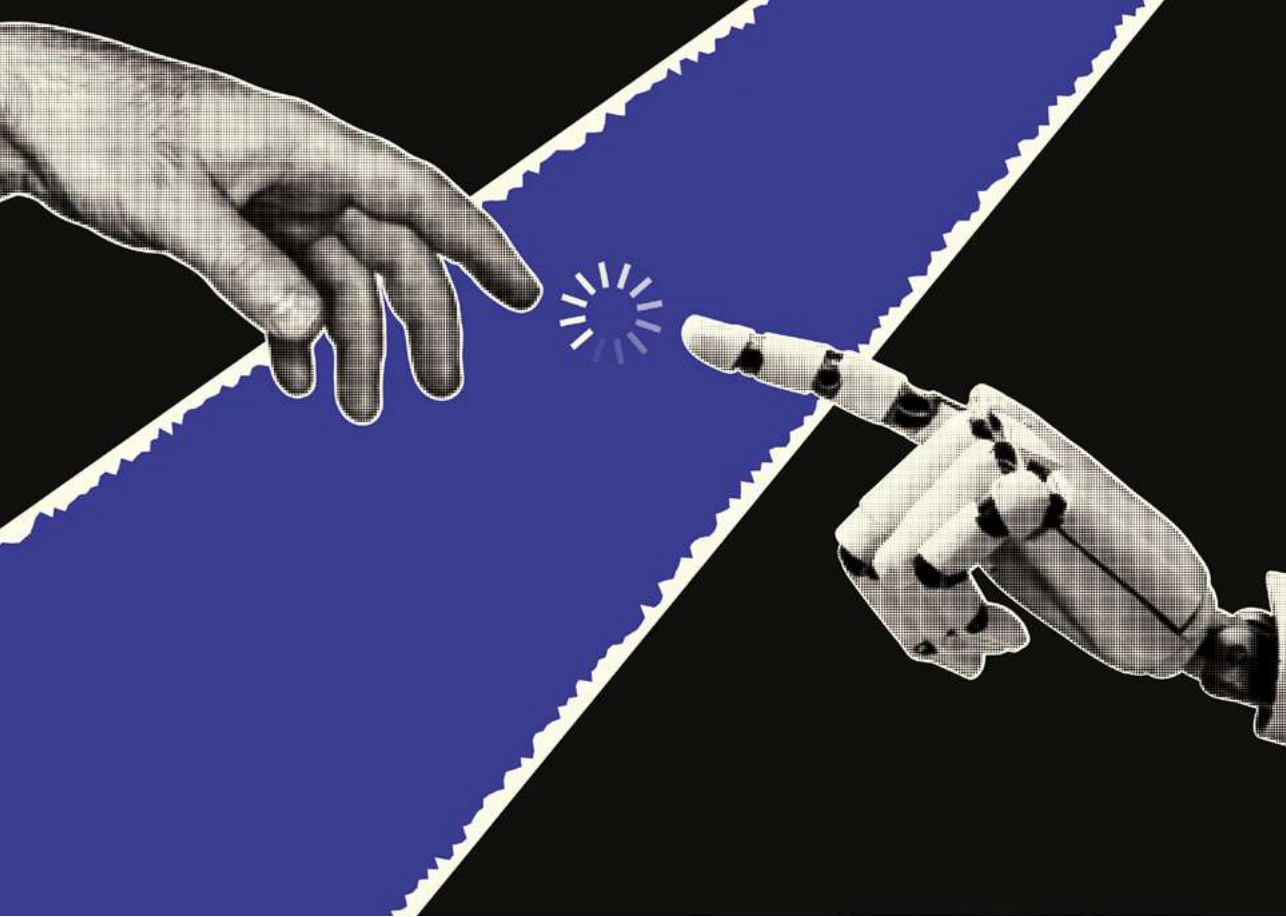
- JONAS, H. (2014) [1984]. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.
- MILLER, R. (2018). *Transforming the Future: Anticipation in the XXI Century*. París: Unesco. New York: Routledge.
- MILLER, R., & POLI, R. (2010). Anticipatory Systems and the Philosophical Foundations of Futures Studies. *Foresight*, 12(3), 3-6.
- POLI, R. (2010). An Introduction to the Ontology of Anticipation. *Futures*, 42(7), 769-776.
- ROSEN, R. (1985). *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. Oxford: Pergamon Press.
- RUSSELL, S., PERSET, K., & GROBELNIK, M. (2023). Updates to the OECD's definition of an AI system explained. *OECD.AI Policy Observatory*. <https://oecd.ai/en/work/ai-system-definition-update>
- SRIPADA, CH. (2016). Free Will and the Construction of Options. In M. SELIGMAN, P. RAILTON, R. BAUMEISTER & CH. SRIPADA (eds.), *Homo Prospectus*. New York: Oxford University Press.



Lydia Garrido Luzardo

Anthropologist and Futurist. PhD in Complex Thinking, M.A. in Integrative Research. Director of the UNESCO Chair in Sociocultural Anticipation and Resilience at the South American Institute for Resilience and Sustainability Studies (SARAS). Senior Advisor for the Special Futures Committee of the Parliament of Uruguay.

LI: lydia-garrido-luzardo-aa486831



Is it possible to regulate AI? Global experiences

XIMENA DOCARMO

The transition from artificial intelligence capable of interacting in natural language to the language and the advent of ChatGPT 3.5 highlights global disparities. These are also expressed in access to new technologies. The need to regulate presents challenges that need to be addressed. This article explains possible regulatory models.

The AI Control Dilemma

By the time ChatGPT 3.5 was launched at the end of 2022, some regions, such as the European Union, had already initiated public debates since 2021 about promoting ethical AI regulation, given the rapid advancements in this technology. In contrast, two years after the widespread public access to generative AI, many parts of the world still show a limited understanding and effort to regulate this technology.

The reasons for this may be multiple. To grasp this phenomenon, the reflections of David Collingridge are quite useful. In the 1980s, Collingridge introduced the control dilemma, which states that “attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow.”

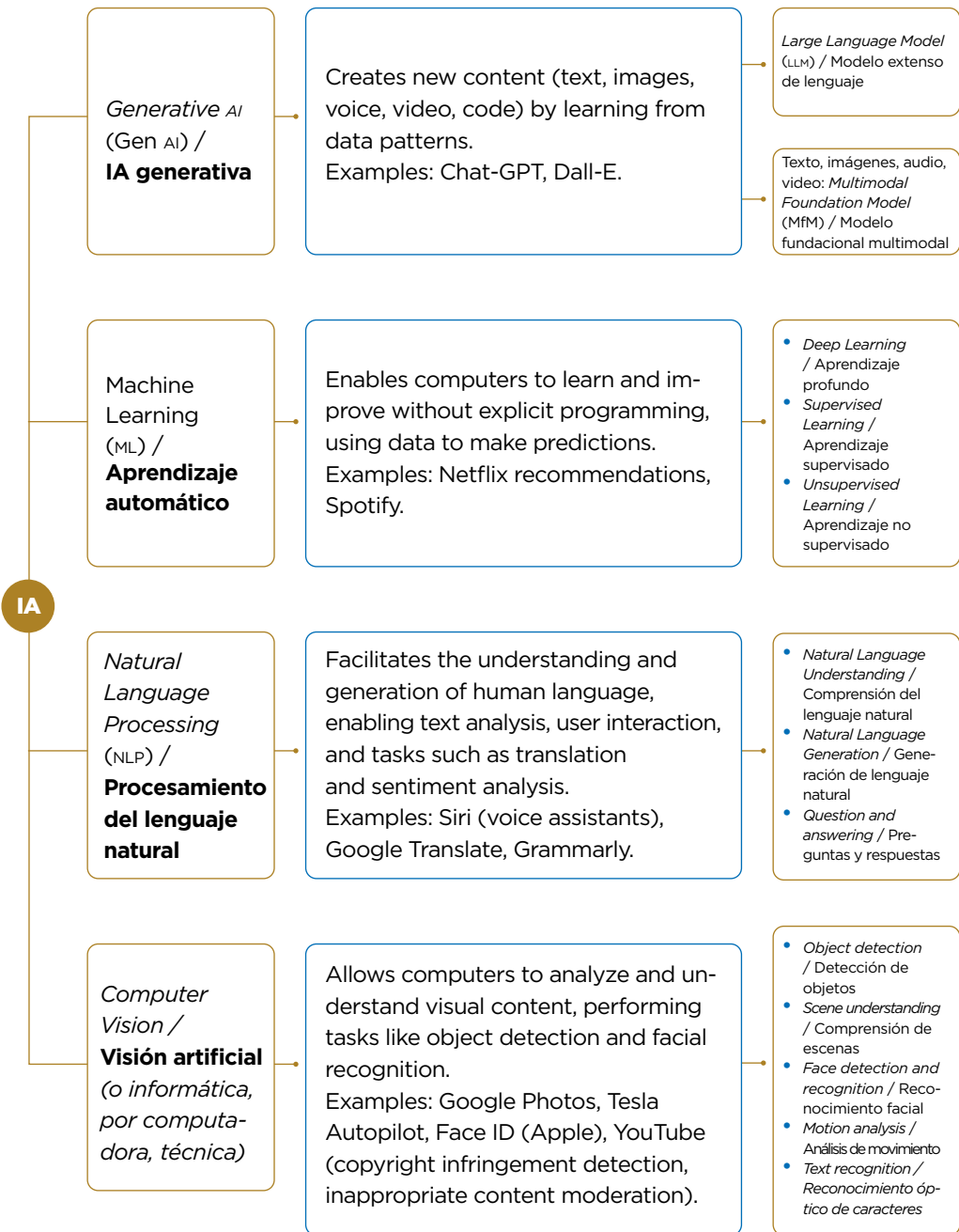
In this regard, Collingridge highlights that, to prevent unintended consequences of a technology, two conditions must be met: first, the harmful effects of the technology must be known, and second, it must be possible to modify the technology to avoid those effects. These are two conditions that, in the context

of AI, seem nearly impossible to foresee in their full scope, which may jeopardize our ability to manage its effects effectively and in a timely manner.

Why Regulate AI?

Long before ChatGPT 3.5 became a landmark in technological advancements in 2022, AI had already begun to be part of our daily lives. AI is not something of the future; it greets us every morning when we glance at the screen of our phone or other electronic devices (see Figure 1). From the moment we unlock our smartphones, we receive personalized suggestions about the weather, the perfect music to start the day, the fastest route to work, or even potential responses to a WhatsApp message. All of this is thanks to a branch of AI known as machine learning. As the day goes on, we may use Face ID to unlock our phones or Google Lens to translate a sign in another language or search for information from an image, interacting with another branch of AI: computer vision (also known as artificial vision, machine vision, or technical vision). This interaction may be less noticeable but still significant; for instance, when we watch videos on YouTube, another piece of this technology is activated

Figure 1. Specialized Branches of AI



Source: Own elaboration and examples. Adapted from *common understanding: simplified AI definitions from leading standards* (NSW Government, 2024).

to detect inappropriate content or copyright infringements in videos uploaded to the platform.

AI is not only present in our personal lives but also in our professional environments. Tools like Chatgpt, Gemini, and Canva have transformed the way we work. These platforms, based on *generative AI and natural language processing*, allow us to simplify complex tasks. From asking Alexa or Siri for help to writing an email in another language with Google Translate or proofreading texts with Grammarly, AI -powered programs have become almost imperceptible yet integral parts of everyday life for many people.

Although the primary limitation of AI lies in the need for internet access, this access is becoming increasingly widespread globally. According to the United Nations, by 2023, more than 65% of the global population was connected to the internet, and over 75% owned a mobile phone, a figure projected to rise to 78% within the next decade. With three out of four people around the world starting their day, in one way or another, by interacting with AI even before the arrival of ChatGPT 3.5 in 2022, it is logical to ask: what has changed? and why has the discussion on AI regulation intensified?

The control dilemma, explained earlier, helps guide answers to these questions. In the early stages of AI development, not enough was known about its consequences. Today, the effects of AI —especially generative AI— are becoming increasingly apparent, which makes regulation more necessary, even if it may eventually come too late and with high societal costs.

Although AI represents a significant scientific advancement, with the potential to close gaps in key sectors like education and healthcare and stimulate the economy through innovation, it also poses serious challenges to individual rights. AI models,

» The omnipresence of AI [...] means that users have only partial control over the use and privacy of their data. «

trained on information provided by humans, reflect the flaws and biases of our society. AI can amplify these biases, reinforcing discrimination against certain population groups, facilitating the misuse of personal data, infringing on freedom of expression, and spreading misinformation, among many other negative effects. Therefore, balancing the maximization of AI's benefits and the mitigation of its risks requires an ethical approach to its regulation that carefully weighs these potential impacts.

Whether consciously or unconsciously, people share various data about their preferences. However, the omnipresence of AI, combined with limited technological literacy in AI, means that users have only partial control over the use and privacy of their data. Data collected by private companies, such as purchase histories, online searches, or social media interactions, as well as data obtained by employers or even governments, can be used for different purposes. In authoritarian contexts, this capacity for surveillance and control can have worrying implications, exacerbating risks to individual rights and fundamental freedoms.

Given the imminent use of people's information, it is crucial for governments to address key issues regarding privacy, security, and the ethical use of both the data input into AI systems and the results processed by AI. Without strong regulatory frameworks, the risk of irresponsible use of AI increases significantly. Such regulation should include mechanisms that ensure the responsible and

sustainable use of the technology, as well as promote AI literacy to mitigate the risks.

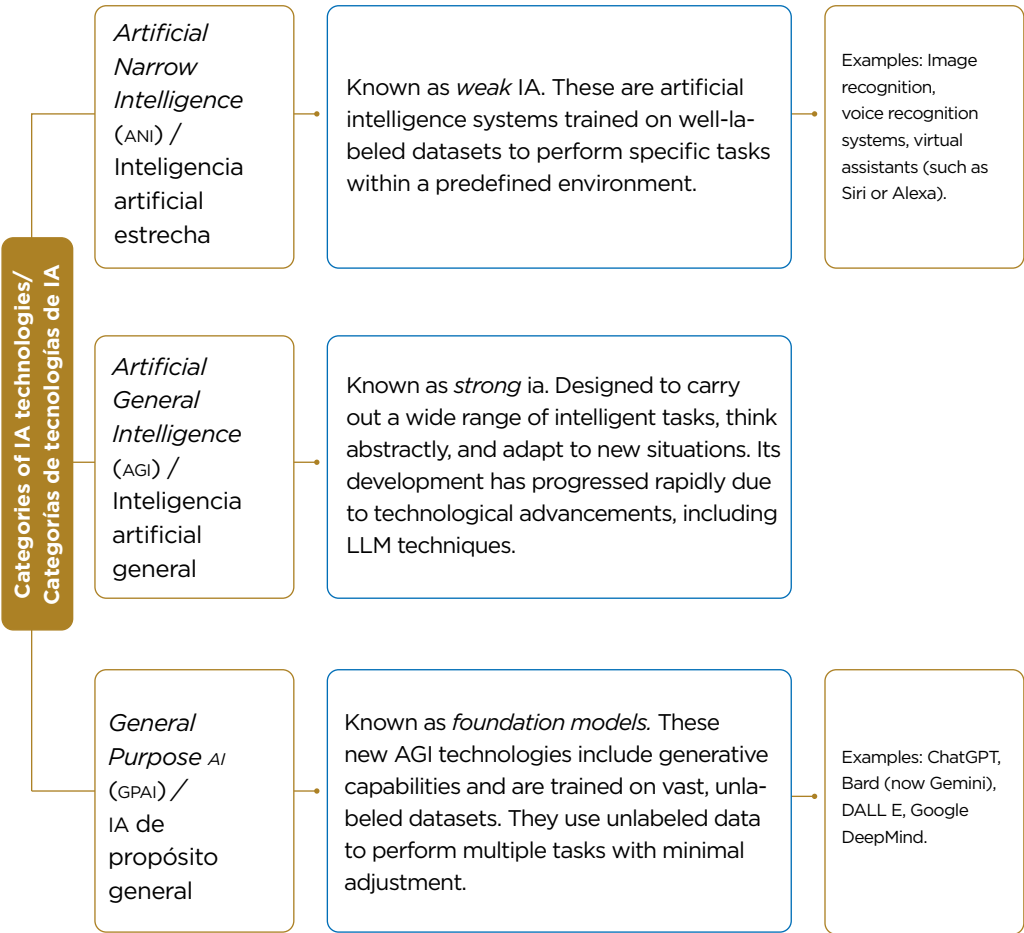
AI in the EU (2021-2030)

With the aim of ensuring that AI systems used in the European Union are “safe, transparent, traceable, non-discriminatory, and environmentally friendly,” the European Parliament formally adopted the first law regulating AI in March 2024 with 523 votes

in favor out of 705 seats, following three years of debates. The legislation, originally introduced by the European Commission in April 2021, proposed establishing the first regulatory framework for AI in the region.

Currently, there is no global consensus on a definition of AI, which means that each country or region regulating it will weigh different elements or categories when creating legislation for its use. In the case of the EU (see Figure 2), the AI Act focuses on General Purpose AI (GPAI). The so-called

Figure 2. Categories of AI Technologies



Source: Self-made. Adapted from the European Parliament (2023).

foundational models possess generative capabilities and are designed to perform a wide range of intelligent tasks. As part of Artificial General Intelligence (AGI), the existence of GPAI is made possible by Large Language Models (LLM) and their generative abilities. In contrast, Artificial Narrow Intelligence (ANI) is only capable of performing specific, predefined tasks. These three elements—GPAI, AGI, and ANI—are critical for understanding the EU's AI risk classification and the respective implementation schedule of these regulations.

The law classifies AI systems into four categories: unacceptable risk, high risk, limited risk, and minimal risk, taking into account specific actions and allowing flexibility to adapt to changes if certain uses evolve and present greater risks:

- *Unacceptable risk*: Systems designed to manipulate human behavior or social classification are completely banned. This includes those that use manipulative techniques, exploit the vulnerabilities of disadvantaged individuals, or implement social scoring.
- *High risk*: Systems that significantly impact safety or fundamental rights—such as those used in education, employment, critical infrastructure,

and the administration of justice—are subject to strict safety, transparency, and human oversight rules.

- *Limited risk*: These systems have less stringent obligations, such as ensuring users are aware they are interacting with AI. These include technologies like chatbots and deepfakes.
- *Minimal risk*: AI-enabled video games and spam filters are exempt from more rigorous regulations. However, this could change as generative AI advances.

The law restricts the use of real-time facial recognition in public spaces, with exceptions for cases such as searching for missing persons or preventing terrorist threats.

This regulation affects 27 Member States. Given the complexity of the European Union system, the AI Act, published in July 2024 and effective as of August the same year, outlines a long implementation process with a phased schedule for enforcing the different obligations (see Figure 3). While all Member States are required to report by November 2024 on the authorities responsible for implementing the legislation, the bans on certain unacceptable risk AI systems will come into force six months after the law's enactment, starting in February 2025. Similarly, regulations for GPAI models will be implemented within 12 months, and high-risk systems will begin to be regulated in 24 and 36 months. During this period, other key rules on governance, confidentiality, and sanctions will progressively come into effect.

From a temporal perspective, the EU AI Act follows a gradual approach that begins with AI literacy and prohibitions, followed by the introduction of rules on notifications, AI models, and sanctions. As it progresses, the grace period for high-risk AI systems concludes, operational provisions are applied, and specific regulations for high-

» There is no global consensus on a definition of AI, which means that each country or region regulating it will weigh different elements or categories when creating legislation for its use. «

Figure 3. Timeline for the Implementation of the ue AI Act

12.7.2024	The EU AI Act is published in the Official Journal of the EU on July 12, 2024.
1.8.2024	The EU AI Act comes into force on August 1, 2024 (Article 113).
2.2.2025	The rules on the purpose, scope, definitions, AI literacy, and prohibitions come into effect on February 2, 2025 (Article 113.a).
2.8.2025	The rules on notifications, GPAI models, certain enforcement matters, and sanctions come into effect on August 2, 2025 (Article 113.b).
2.8.2026	The EU AI Act is fully applied, the general grace period for high-risk AI systems ends, and most operational provisions come into effect on August 2, 2026 (Articles 111.2 and 113).
2.8.2027	Rules regarding high-risk AI systems under paragraph 1 of Article 6 come into effect on August 2, 2027 (Article 113.c).
2.8.2030	The grace period for high-risk AI systems intended for use by public authorities ends on August 2, 2030 (Article 111.2).

Source: Excerpt from EU AI Act enforcement timeline (2024) y White & Case (2024), extended English version.

risk AI are implemented. To ensure the effective implementation of the regulation, the establishment of the *EU AI Office* within the European Commission has been planned.

Developing AI Regulation Models

In contrast to the approach adopted by the European Union, other regions of the world with a high degree of AI development prioritize different components in their regulatory and technological adoption models. In the book *Digital Empires: The Global Battle to Regulate Technology*, Anu Bradford reflects on the contrasts between the European model, which focuses on establishing global regulatory standards, the *US model* that encourages the private sector,

and the *Chinese model* driven by the use of state resources.

She explains that in the US model, which is market-centric, the role of the government is limited, allowing large tech companies to lead governance. It fosters a favorable environment through incentives for innovation and exports its influence through services and technologies, consolidating its private power in the global economy.

Regarding AI regulation, the United States lacks comprehensive federal legislation, and its approach relies on laws and guidelines of limited scope. Key regulations include the National AI Initiative Act of 2020, which focuses on promoting research and development in this field, along with the establishment of the National Artificial Intelligence Initiative Office,



» *The Chinese model, driven by the state, seeks to establish the country as a technological superpower through the utilization of state resources.* «

tasked with implementing the national strategy. In October 2023, the White House issued an Executive Order on the Safe and Trustworthy Development and Use of AI, which sets guidelines for the development of federal standards, including elements of transparency in the results of safety testing. In the past year, various states have led initiatives concerning the regulation of high-

risk systems, algorithmic discrimination, and automated decision-making. The general trend is toward increased sectoral regulation at the state level; however, public discussions are expected to continue regarding the implementation of cohesive AI regulation and the establishment of a federal authority.

Meanwhile, the Chinese model, driven by the state, seeks to establish the country as a technological superpower through the utilization of state resources. This model manifests in surveillance, censorship, and propaganda, along with actions aimed at preserving political control. In turn, China exports its infrastructural power by developing 5G networks, data centers, and smart cities.

Regulatory-wise, in 2023, the first specific administrative regulation on generative

AI was published, called the *Provisional Measures for the Management of Generative AI Services*. This regulation does not categorize risks, but certain services, such as those with “public opinion attributes or social mobilization capacity,” are subject to stricter scrutiny, including security assessments and general application requirements like content moderation and labeling. Among the labeling norms, requirements include upholding socialist values and not generating prohibited content that incites subversion of state power or the overthrow of the socialist system, thereby endangering national security and interests. The responsibility for regulating generative AI primarily falls on the *cyberspace administration of China*.

This discussion includes other giants on the global stage. According to the World Economic Forum (WEF), the five largest economies in the world have made significant progress in developing AI ecosystems. In addition to the United States, China, and Germany (a member of the EU), Japan and India have also joined this list. Although both countries lack specific AI laws, they have adopted distinct approaches to address their regulation.

Japan has been a key player globally in launching the *Code of Conduct for Organizations Developing Advanced AI Systems* in the context of the G7 in 2024, an instrument that compiles 11 recommendations with a risk-based approach. Domestically, the country follows a soft law strategy, promoting AI governance through guidelines aimed at minimizing risks while prioritizing the promotion of innovation. In 2024, the AI Guidelines for Business Version 1.0 was also published, a non-binding guideline that seeks to promote voluntary efforts following a risk-based approach. However, a recent draft AI Bill could redirect the current strategy toward

» **International experience in AI regulation offers valuable lessons for Latin America, both in terms of promoting innovation and protecting individuals from the associated risks of this technology.** «

a hard law approach, including oversight of developers and the imposition of fines and penalties in case of non-compliance.

India, for its part, has established sector-specific frameworks, such as in finance and health, and its approach is guided by the *National AI Strategy* from 2018 and the *Operationalizing Principles for Responsible AI* from 2021, which prioritize training and incentives for the ethical design of AI. Although regulations are quite limited, the forthcoming India Digital Bill is expected to delineate and regulate high-risk AI systems.

Latin America: Is a Unified Regulation Possible?

International experience in AI regulation offers valuable lessons for Latin America, both in terms of promoting innovation and protecting individuals from the associated risks of this technology. However, the idea of a unified regulatory framework for the region may not be realistic or effective, as each country is progressing at its own pace in creating ecosystems for AI development.

Recent recommendations, such as the resolution published in 2024 by the United Nations titled *Harnessing the Opportunities of Safe and Reliable AI Systems*

for Sustainable Development and the OECD's Recommendation on AI originally adopted in 2019 and updated in 2021, highlight the need for clear governance, investment in technological infrastructure, and education in digital skills. For Latin American countries, however, these challenges will need to be approached from a more flexible perspective, considering the different realities and capabilities of each nation. Chile, Brazil, and Uruguay are leading in AI research and development according to the Latin American AI Index (ILIA -*Índice latinoamericano de inteligencia artificial*) 2024, but advancements are not uniform across the region.

In such a complex and rapidly evolving context, Latin America must balance the promotion of AI with the protection of fundamental rights, leveraging this technology for inclusive and sustainable development. The key will be to design regulatory frameworks that allow for responsible implementation while respecting the diversity and evolutionary pace of each country, without compromising individuals' rights or, ultimately, democracy. ♦

References

- COLLINGRIDGE, D. (1980). *The Social Control of Technology*. New York: St. Martin's Press.
- EU ARTIFICIAL INTELLIGENCE ACT. (2024). [EU Artificial Intelligence Act](#).
- EUROPEAN PARLIAMENT. (2023). [General-purpose artificial intelligence](#).
- ILIA, [ÍNDICE LATINOAMERICANO DE INTELIGENCIA ARTIFICIAL](#) (2024).
- NACIONES UNIDAS, ASUNTOS ECONÓMICOS. (2023). [Más del 75% de la población mundial tiene un teléfono celular y más del 65% usa el internet](#).
- NETFLIX. (2024). [Y ahora qué. El futuro según Bill Gates](#).
- Parlamento Europeo, [Ley de IA de la UE: primera normativa sobre inteligencia artificial](#) (2024)
- WHITE & CASE. (2024). [The global dash to regulate AI](#).
- WORLD ECONOMIC FORUM. (2024, June 2). [Así es como los capitalistas de riesgo invierten en la IA en cinco países](#).



Ximena Docarmo

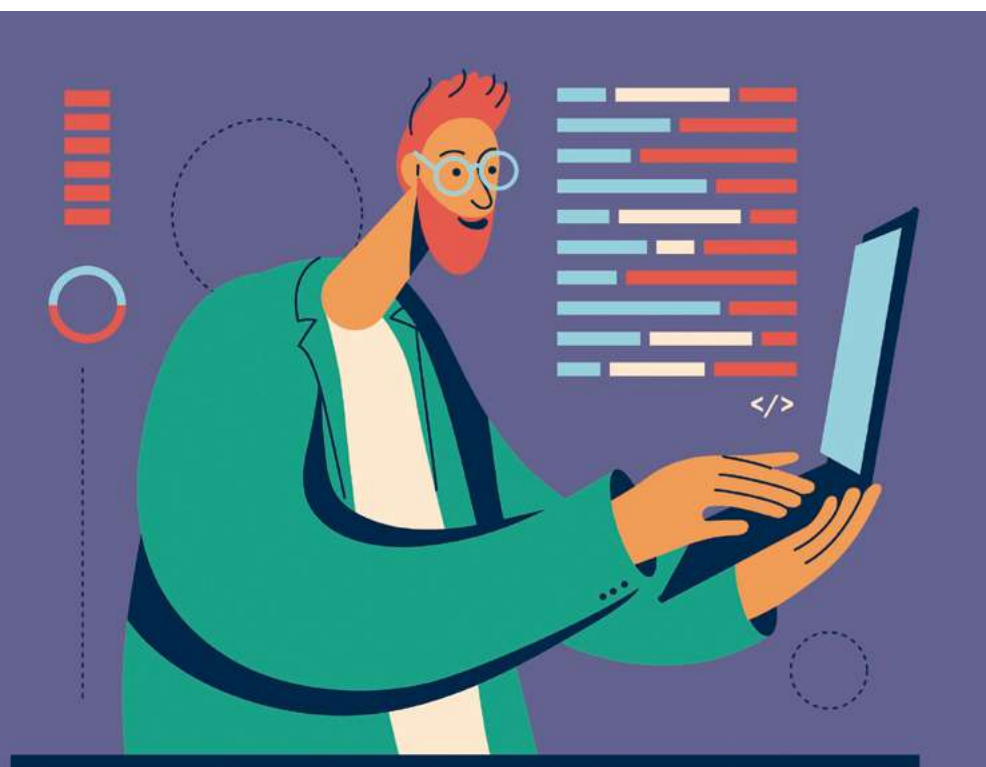
Master in Public Policy (Hertie School of Governance, Berlin). Fellow at the headquarters of the Konrad Adenauer Foundation, Berlín.

X: @xidocarmo

IG: @xidocarmo

Fb: xidocarmo

LI: xidocarmo



From mass networks to personalised voting

JESÚS DELGADO VALERY

The use of artificial intelligence transforms election campaigns with tools such as microtargeting. Campaigns can personalize their political communication and influence voters' decisions. This offers new opportunities and raises ethical challenges related to misinformation and the use of personal data, which impact on the quality of democracy.

In recent years, the use of artificial intelligence (AI) has become increasingly common, especially in the field of image generation. Political parties and electoral processes have been notably affected by this phenomenon. The widespread use of social media has been followed by the emergence of these new technologies, which are playing an increasingly sensitive role in the dissemination of political propaganda. This has raised countless debates, as new and more sophisticated approaches raise both ethical and technical questions.

At its core, AI-generated content allows for the creation of images based on user input. This enables campaigns to construct visual narratives representing potential future scenarios (showcasing the positive outcomes of certain policies or the negative consequences of others). However, from the outset, this practice introduces the issue of creating artificial images that do not reflect reality, potentially falling into the realm of fake news or biased narratives. A similar result occurs with AI technologies that can mimic a person's voice, facilitating the spread of falsified audio recordings.

In recent years, many campaigns have employed AI. On the surface, we see its use in generating videos that depict highly

optimistic alternative scenarios in the event of an electoral victory (or, conversely, devastating scenarios in case of a loss), as was done by the campaigns of Sergio Massa and Javier Milei in the 2023 Argentine elections. Similarly, AI-generated images of Donald Trump being arrested following his court conviction (featuring increasingly implausible arrest scenes aimed to generate viral content and memes).

Social media has leveled the playing field in terms of reaching the electorate, particularly for sectors that may have been marginalized by mainstream media. These platforms provide a straightforward and low-cost means of dissemination. However, their rapid reach also facilitates the widespread propagation of fake news with far greater ease, making fact-checking and refutation more complicated.

Although its use is relatively recent, numerous studies have already explored the impact of artificial intelligence on electoral processes, especially in content creation for campaigns. In this paper, we will attempt to shed light on the impact of artificial intelligence on network design and development, as well as its role in organizing social support—an area that has received less attention in research.

AI on Electoral Campaigns

Artificial intelligence refers to

[...] a discipline belonging to computer science, which proposes computational models of learning based on human biological neural networks. In this sense, several AI models have been proposed, which thanks to advances in computer technology have allowed the development of intelligent systems that facilitate the processing of a greater amount of data in a shorter time, speeding up decision making (Márquez Díaz, 2020).

The concept of artificial intelligence dates back to the second half of the 20th century. The famous *Turing test* posed the challenge of determining whether ordinary people could distinguish between interacting with a human or a chatbot in a written conversation (Turing, 1950). However, global enthusiasm for AI did not reach its current heights until recent years. Since 2023, AI-generated images capable of emulating human photographs have begun to emerge.

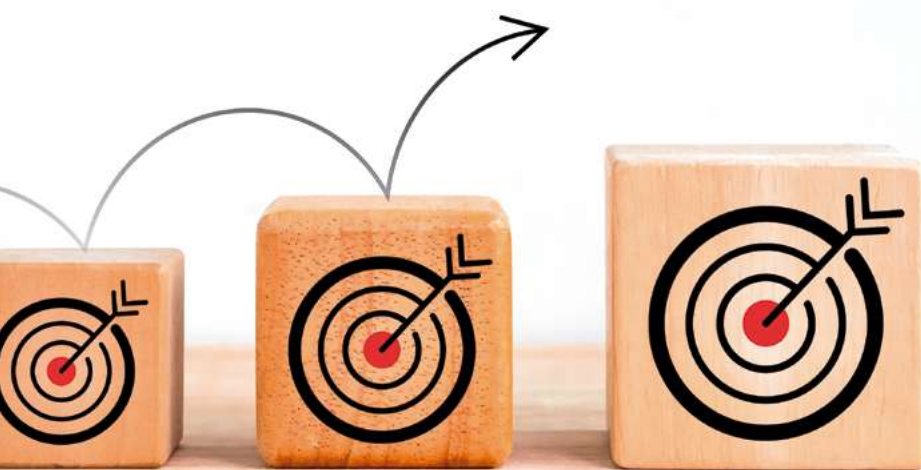
In the political arena, the impact of AI has been significant. Electoral campaigns are characterized by the need to persuade voters to support a particular candidate, and in recent years, to also demobilize or push voters to reject an opposing option when gaining their support seems unattainable. Although campaigning today is different from fifty or a hundred years ago, the ultimate goal remains the same.

Electoral campaigns in the 20th century were marked by their mass appeal and focus on general issues. While there are various theories about voter behavior, there is a consensus in the literature about the importance of social class in voters' decision-making, especially when this meant a determining factor.

» The electorate shifted from being a relatively uniform mass to numerous subgroups organized around much more specific ideals. «

Campaigns were largely conducted through mass rallies led by prominent political figures, while party cadres worked within their communities. The main communication channels were radio, television, and print media, all of which shared a “common characteristic: they were indiscriminate. In





other words, they were not tailored to specific profiles but aimed at persuading society as a whole” (Cebrián Beltrán, 2024).

In traditional or modern electoral campaigns, communication occurred in a controlled environment, with format limitations that curbed or reduced aggression and disinformation. The electorate generally maintained a moderate stance, which discouraged and even punished extremism. Moreover, there were fewer media outlets, and their impact and reach depended heavily on their credibility. As a result, they were highly cautious in fact-checking and verifying the information they disseminated (Rubio Núñez et al., 2024).

However, the decline of major political parties in the 20th century across Europe and Latin America coincided with a transformation in the electorate, which shifted from being a relatively uniform mass to numerous subgroups organized around much more specific ideals (environmentalists, advocates for women’s political participation, minority rights, regionalist parties, separatist or independence movements, etc.).

This new reality compelled political groups to make greater efforts to understand their electorate and to use new technologies, both to better comprehend them and to represent them.

The arrival of social media, of course, has

proved a revolutionary incentive, perhaps on par with or even surpassing the impact of television. While previous methods of communication have not been abandoned, social media has led to an unprecedented mass dissemination of information, allowing political messages to reach nearly every sector of society in much shorter timeframes. This has posed a challenge for majority sectors and even authoritarian regimes seeking to maintain control over the flow of information.

With current technologies, even in authoritarian contexts, an opposition candidate can design a coherent and competitive campaign that provides access to a wide range of resources (spots, jingles, interviews, and the dissemination of propaganda, among others)—something unimaginable in the authoritarian contexts of the 20th century.

This is evident in democracies. We have witnessed new ways of campaigning that significantly reduce the need for physical presence and large-scale rallies. For example, in Chile's 2021 presidential elections, Franco Parisi (from the *Partido de la Gente*) ran as a candidate despite residing in the United States. Amid the pandemic, which encouraged the use of various remote communication tools, Parisi campaigned from abroad and garnered 900,000 votes, equivalent to 12.8% of the total ballots cast (Serval, 2021).

Microtargeting: Understanding the Voter

With these new technologies, it is now possible to do something that previous ones either did not allow or limited: create personalized political messages directed at very specific segments of voters, known as microtargeting. This strategy relies on audience segmentation, using demographic, behavioral, and preference data to design messages that resonate with the characteristics

» **Microtargeting relies on audience segmentation, using demographic, behavioral, and preference data to design messages.** «

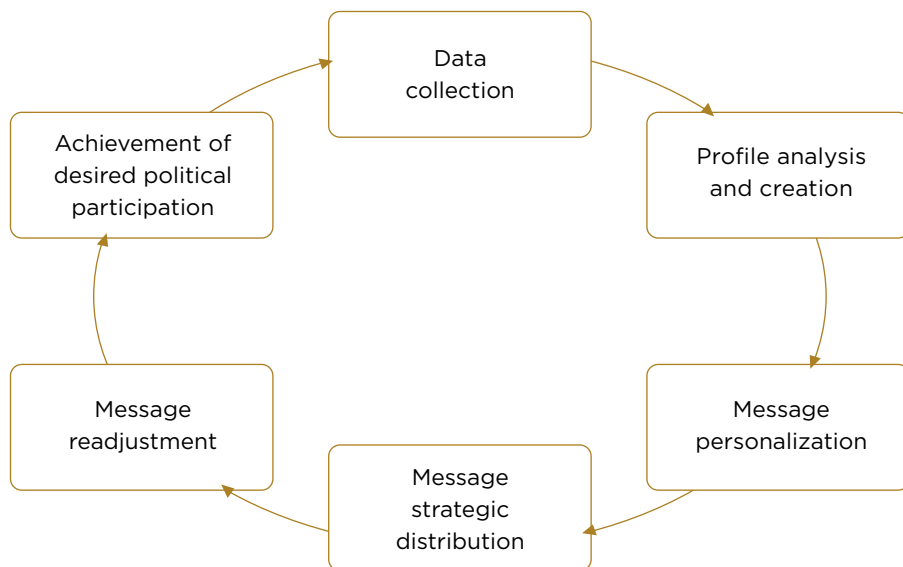
and needs of small groups of individuals, often referred to as *microsegments* or *clusters*. AI provides sophisticated and innovative mechanisms to facilitate this task.

The process works as follows: individuals provide their personal data when using apps, visiting websites, making online purchases, etc. The aggregation of all this information, along with its analysis and processing to correlate data and identify patterns and trends, is known as *big data*. This data serves as the foundation for algorithms, which are essentially “a sequence of commands that instruct a computer to convert an input into an output. For example, a list of individuals sorted by age. The computer takes the ages from the list (input) and produces the newly sorted list (output)” (FRA, 2018).

In the electoral realm, which is the focus of this article, the algorithm applied to big data aims to profile voters. In other words, the data is analyzed to understand voters' political stances, and predict and influence their electoral behavior. This process culminates in the formation of microsegments or clusters—groups of individuals who share similar profiles.

Once the different groups have been characterized, specific messages are crafted for each one. The content of these messages will depend on the objectives at hand. For example, if a group of potential voters is identified as being demotivated or disengaged, targeted messages can be designed to involve

The Cycle of Artificial Intelligence in Political Participation



Source: Cebrián Beltrán (2024).

them in the electoral process, raise awareness of the political situation, and convince them that their vote is crucial.

Conversely, if a segment is identified as likely to vote for the opposition, specific messages can be crafted to demobilize them. This might involve highlighting cases of corruption within the opposing party or statements from their leaders that conflict with the convictions of that particular segment.

Many of these messages do not appeal to voters' analytical capabilities but instead target their emotions. As such, it is increasingly common to see controversial topics (such as immigration, abortion, or religion) dominate political discussions to provoke strong emotional reactions and simplify the political debate. This trend has promoted the use of disinformation as a tool to solidify the electorate, as well as the use of generative AI to create and manipulate audiovisual content to incite outrage.

This technology is so fast-paced that it allows campaigns to assess its impact in real-time and make adjustments accordingly. While modern campaigns once relied on traditional opinion studies like polls or focus groups to gauge the impact of a message, proposal, or slogan, microtargeting and social media enable preliminary analyses of campaign strategies' effectiveness within minutes.

This new strategy enables political parties and candidates to address concerns of specific segments that may have previously gone unnoticed, despite their significant electoral potential, thereby increasing the efficiency of electoral campaigns by allowing real-time analysis.

Microtargeting, however, can also be misused to present voters with misleading or biased information, to spread fake news and disinformation, and to create echo chambers that limit debate. This undermines trust in institutions and weakens the foundations of democracy.

Impact on Electoral Outcomes

Measuring the impact of various AI-driven tools on electoral outcomes is still a challenging endeavor. Establishing a rigorous methodology for this purpose is, at least for the time being, impossible. However, mechanisms exist to analyze the communication environment and record the discussions taking place within specific societies. In this regard, it is possible to assess the success of AI-powered tools in establishing issues on the public agenda, as well as the positions citizens take concerning them.

It can also be asserted that certain topics or key ideas in some elections have significantly influenced the results. Furthermore, the use of technologies such as microtargeting has maximized the impact of these issues on the electorate.

Below are some emblematic cases where technology, particularly the use of big data to design specific campaigns, has affected electoral outcomes.

2008: The Obama Case

Long before AI became a topic of discussion, a paradigm shift marked a turning point in electoral campaigns: microtargeting. In the 2008 U.S. presidential elections, the campaign team of then-candidate Barack Obama succeeded in profiling every voter in the country, focusing on two key points: whether they were likely to vote and if they would vote for Obama. Based on this information, strategies were developed to influence their decisions, ultimately leading the candidate to the White House. This was one of the first large-scale experiences in voter profiling, albeit not with the level of sophistication we recognize today.

2016: Cambridge Analytica

Donald Trump's election was overshadowed by the Cambridge Analytica and Facebook scandal, which would become a case study in the use of personal data to craft ultra-segmented political messages.

Aleksandr Kogan, a professor at the University of Cambridge, designed a personality test for Facebook in 2013, through which he obtained data from 50 million people. The test was completed by 265,000 users of the social media, who, to participate, had to grant access to their friends' information without their consent.

Using the data from these individuals, Cambridge Analytica created psychological profiles and designed specific messages aimed at influencing their political preferences, even disseminating fake news (BBC World, 2018).

2023 in Argentina: Images and Videos



In the 2023 presidential elections in Argentina, the teams of candidates Sergio Massa and Javier Milei extensively utilized generative artificial intelligence to create promotional images and videos, as well as attacks on their opponents. The official candidate's team employed AI to produce posters and videos depicting Massa as a strong and charismatic leader, drawing inspiration from Soviet styles and pop culture. In response, Milei released images portraying a lion liberating Argentina and depicting Massa as a communist leader. The use of generative AI by Massa's campaign to illustrate a potential dystopian future in the event of Javier Milei's victory sparked controversy (Nicas and Cholakian Herrera, 2023).

» The emergence of big data and AI presents a challenge not only for parties and the electorate but also for institutions, the integrity of the communication space, and ultimately, the democratic system itself. »

2024 in India: Chatbots

In India, during the campaign for the general elections of 2024, a controversy erupted regarding deepfakes on social media when a user asked Google's AI tool, Gemini, about the alleged fascist nature of Prime Minister Narendra Modi and the Bharatiya Janata Party (BJP). The response indicated that Modi's government was "accused of implementing policies that some experts have characterized as fascist" (Dillon, 2024). Indian Minister of State for Electronics and IT, Rajeev Chandrasekhar, criticized this response, stating it violated the country's laws. This incident underscores the growing concern in India regarding disinformation and the use of AI in electoral contexts. Google promptly reacted, asserting that it was "working" to "improve the reliability" of the tool (Mukherjee, 2024).

Examples in Denmark and the United Kingdom

The *Synthetic Party* is a political party in Denmark led by an AI named Leader Lars, a chatbot accessible via Discord. Its goal is to engage citizens who typically do not vote and place technology at the center of political debate, promoting coexistence between AI and humans, as well as the regulation of AI accountability. The party, which defines itself as synthetic, develops its platform based on proposals from minor Danish parties dating back to 1970. Although led by an AI, the project is driven by the artist group Computer Lars and the technology center MindFuture, which seek to ensure the party's longevity and global expansion. They also propose the creation of a new Sustainable Development Goal (SDG) focused on the relationship between humans and robots (Vicente, 2023).

Another example emerged in the United Kingdom, in the lead-up to the 2024 elections, featuring the candidate of Smarter UK, AI Steve, an AI avatar representing legal candidate Steve Endacott in the Brighton Pavilion constituency. The aim was to rekindle interest among apathetic sectors of

the population in politics by allowing voters to directly influence decisions. If elected, Endacott would physically represent AI Steve in the British Parliament, acting according to the majority vote of the electorate. The project faced several ethical questions regarding its efficacy and raised various legal concerns. Nevertheless, AI Steve came in last in the constituency (a stronghold of the Green Party) with only 179 votes (Smith, 2024).

Conclusions

The changes in the way electoral campaigns are conducted over the past two decades have been staggering. This dynamic, combined with the decline of traditional parties and the emergence of disruptive, charismatic, and populist figures, has resulted in a 180-degree turn in political communication.

The advent of social media, along with the sophistication of methods for collecting, analyzing, and cross-referencing data on a massive scale, profiling the audience (or electorate), and crafting specific messages has ushered us into a new game—one that cannot be understood through old categories.

However, this new reality also brings forth new challenges, perhaps even greater than those of the past. In recent decades, we have experienced democratic fatigue, characterized by political disaffection, a crisis of representation, and a decline in citizens' adherence to democratic principles. The emergence of big data and AI presents a challenge not only for parties and the electorate but also for institutions, the integrity of the communication space, and ultimately, the democratic system itself.

On the other hand, new technologies have also allowed for a deeper understanding

of the electorate, complicating and problematizing it. This presents an opportunity for parties to consider the needs and concerns of voters, ensuring that these factors have a genuine impact on their programs and proposals.

For now, it appears we are still uncovering the effects of artificial intelligence on electoral campaigns, and it would be premature to draw definitive conclusions. Its use poses significant challenges and raises fundamental ethical debates, such as whether individuals are genuinely free to access plural, broad, and critical information, or whether they are increasingly confined within informational bubbles specifically designed to shape their choices. ♦

References

- BBC MUNDO. (2018, March 21). [5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera us\\$37.000 millones en un día.](#)
- CEBRIÁN BELTRÁN, S. (2024). [De la talla única al traje a medida: el microtargeting político para influir en las elecciones.](#) Paper at XXI Congreso de la Asociación de Constitucionalistas de España, round table «Garantías constitucionales de elecciones libres», Valladolid.
- DILLON, A. (2024, February 26). [India confronts Google over Gemini AI tool's 'fascist Modi' responses.](#) *The Guardian*.
- FRA. (2018). [#BigData: Discrimination in data-supported decision making.](#)
- ISSENBERG, S. (2012, December 19). [How Obama's Team Used Big Data to Rally Voters.](#) *MIT Technology Review*.
- MÁRQUEZ DÍAZ, J. (2020). [Inteligencia artificial y Big Data como soluciones frente a la covid-19.](#) *Revista de Bioética y Derecho*, n.º 50.

- MUKHERJEE, M. (2024, March 19). [AI deep-fakes, bad laws - and a big fat Indian election.](#) *Reuters Institute*.
- NICAS, J., & CHOLAKIAN HERRERA, L. (2023, November 15). [Las campañas electorales de Argentina recurren a la IA.](#) *The New York Times*.
- RUBIO NÚÑEZ, R., FRANCO ALVIM, F., & ANDRADE MONTEIRO, V. (2024). [Inteligencia artificial y campañas electorales algorítmicas.](#) Madrid: CEPIC.
- SERVEL. (2021). [Elección presidencial 2021.](#)
- SMITH, C. (2024, July 2). [Britain's first AI politician claims he will bring trust back to politics – so I put him to the test.](#) *The Conversation*.
- TURING, A. M. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236), 433-460.
- VICENTE, M. J. (2023). [Inteligencia artificial y política. Los casos de “Synthetic Party” & Tama.](#) In A. DAFONTE GÓMEZ y M. I. MÍGUEZ GONZÁLEZ (coords.), *El fenómeno de la desinformación: reflexiones, casos y propuestas* (pp. 603-617). ISBN 978-84-1170-538-7.



Jesús Delgado Valery

Executive Director Transparencia Electoral. DemoAmlat Coordinator. Bachelor's degree in International Studies from the Universidad Central de Venezuela. Master's candidate in Electoral Studies at the Universidad Nacional de San Martín, Argentina.

X: @JesusDValery

LI: jesús-delgado-valery-17651686



The fight against corruption from a new technological paradigm

DENISSE RODRÍGUEZ-OLIVARI

Artificial intelligence has emerged as an innovative tool in the fight against corruption. This new technology has countless unprecedented benefits in identifying irregularities, automate compliance checks and improve transparency in public and private organizations. We analyze the opportunities and challenges in the application of AI-based technologies to detect, prevent and fight corruption around the world, with a focus on Latin America.

Introduction

The fight against corruption remains one of the most pressing issues in Latin America. According to the American Society/Council of the Americas (as/coa) and Control Risks *Capacity to Combat Corruption Index*, 70% of experts regard corruption as a critical challenge in their countries, surpassed only by public insecurity and the post-pandemic economic situation. Furthermore, Transparency International's Corruption Perceptions Index indicates that most countries in the region are either stagnating or worsening regarding their anti-corruption efforts. Only Guyana and the Dominican Republic report significant progress, while Venezuela has dropped to an all-time low in the global rankings.

Amid this unfortunate scenario, artificial intelligence (AI) presents innovative solutions to this deeply complex problem. The European Commission's High-Level Expert Group on Artificial Intelligence defines AI as "systems that display intelligent behavior by analyzing their environment and taking actions—with some degree of autonomy—to achieve specific goals" (2019). AI technologies offer data analysis, anomaly

detection, corruption-risk prediction, and process automation. However, the risks of unethical use must not be ignored—for instance, personal data erosion, indiscriminate surveillance, algorithmic discrimination, and inappropriate implementation in electoral campaigns.

Despite all, the opportunities far outweigh the risks. The proliferation of AI as an anti-corruption technology—also known as AI-ACT—has captured the attention of public officials, activists, investigative journalists, and academics who specialize in the issue. AI-ACT encompasses socio-technical systems that enable the analysis of large volumes of data, reducing the discretionary power of public officials, and mediating interactions between citizens and governments (Mattoni, 2024).

AI's predictive and preventive capabilities provide agility and efficiency in anti-corruption efforts across Latin America, with Brazil leading the way in anti-corruption use of AI technology. Odilla (2023) documents over 30 initiatives driven by public officials (top-down) and civil society organizations (bottom-up) that promote monitoring, identifying, reporting, and predicting irregularities and corruption risks. In Latin America, Rosie,

Alice, and Monica are prominent examples of such initiatives. Alice (which analyzes public calls, contracts, and tender offers) and Monica (which monitors acquisitions) scan procurement processes to identify irregularities. The Brazilian Internal Revenue Service has also implemented AI to detect customs fraud, while various state agencies are expanding the use of AI in broader anti-corruption efforts. However, the adaptation to new forms of irregularities is slow, and there is limited expert audit capacity. Moreover, some aspects require refinement—namely, biased data outputs, responsible implementation, and perfection of failed algorithms.

AI in the Fight Against Corruption: Opportunities and Benefits

In the government sector, generative artificial intelligence and large language models (LLM) provide numerous opportunities for innovation in the fight against corruption. The following sections present some of their key applications.

Data Analysis and Pattern Recognition

AI facilitates the processing of large datasets to detect early signs of corruption through inconsistencies or duplicated information in financial transactions, tenders, contracts, or subsidies. AI-ACT can help predict areas or sectors at higher risk of fraud based on historical data and current trends. These capabilities help identify conflicts of interest and assess corruption risks to detect anomalies and corruption scenarios.

A notable case is that of proactive governance through Saler (Rapid Alert System), implemented by the General

Inspection of Services of the Generalitat Valenciana in Spain. Saler's main objective is to anticipate risks or weaknesses liable to harm public administration and arising from inertia or poor practices. It utilizes the vast digitized information handled by the Generalitat Valenciana, along with databases from registrars, notaries, and intellectual property entities, to analyze any administrative procedures of interest to compliance officials. The risks detected by Saler are linked to information security, tenders, selection committees, collusion, verifications, governance, ethics, compliance with the law, and human resources.

Saler relies on algorithms, for example, to monitor unemployment benefit recipients and detect fraudulent claims. This predictive tool evaluates individual's health and predicts the likelihood of their return to work. However, it is crucial to support these assignments with thorough analyses to prevent an overflow of false positives, which could harm citizens in genuine need of state aid.

Similar systems can analyze local government or ministry expenses to identify cost overruns, deficiencies, or poor practices. Public procurement accounts for 13% of the gross domestic product in OECD countries, and 8% in Latin America (Pérez, 2021). Despite this being an area vulnerable to corruption, resources for controlling and monitoring public spending are often scarce. The following case studies illustrate the use of AI in addressing this issue.

VigIA

Developed by the Tic Tank of Universidad del Rosario—a think tank focused on information technologies—and the Corporación Andina de Fomento (CAF) for Bogota's District Oversight Office, VigIA

is an AI-based system designed to oversee contracts with high risk of corruption and inefficiency issued by Bogota's Mayor's Office. The system leverages data from Colombia's Electronic Public Procurement System (Secop) through the National Public Procurement Agency (Colombia Compra Eficiente) to predict corruption risks in each contract using machine learning models. By expediting audit processes and detecting inefficiencies or irregularities, the system assigns risk scores, enabling oversight bodies to focus on contracts most vulnerable to corruption.

Love Serenade: AI for Social Control of Public Administration

Following the outbreak of the notorious Mensalão corruption scandal—known as “big monthly allowance” due to improper payments within the lower house of the Brazilian Parliament—data scientist Irio Musskopf, sociologist Eduardo Cuducos, and Businessman Felipe Cabral conceived this project in 2016. The initiative uses machine learning to analyze government data and flag suspicious public spending. Findings are posted on X via a bot named Rosie. Other tools, such as La Denunciante (The Whistleblower), Jarbas (for data visualization), and Toolbox join these efforts. In its early stages, Love Serenade identified 629 irregularities involving 216 out of 513 federal deputies.

Streamlining Procedures

AI facilitates the automation of routine tasks, reducing human error and increasing the speed of corruption risk detection. It primarily improves process efficiency and, consequently, anti-corruption efforts.

» AI facilitates the automation of routine tasks, reducing human error and increasing the speed of corruption risk detection. It primarily improves process efficiency. «

Aarvik (2019) highlights the case of the Kenyan government and the IBM Research Group, who teamed up to decrease the incentives for bribery in administrative procedures. It is worth remembering here that a good part of corrupt transactions is about greasing the wheels, that is, speeding up procedures that should not take so much time or resources in the first place. By making procedures more efficient, Kenya improved its position in the World Bank's Doing Business ranking, rising from 136th to 56th place¹. However, before engaging in this transformation, countries must have a high level of digitalization.

AI also offers opportunities to streamline reporting channels, another critical pillar of anti-corruption endeavors. AI algorithms can prioritize and categorize complaints, reducing the costs associated with officials processing each case individually. A study by Pierri and Lafuente (2022) from the Inter-American Development Bank provides evidence on the handling of citizen complaints in the New Talents in Government Control Program at the

¹ Following data irregularities in the 2018 and 2020 Doing Business report, World Bank management suspended the next edition and initiated a series of reviews and audits regarding the report and its methodology.

Comptroller General's Office (CGR) in Peru. In a sample of 5,000 occurrences, 40% were deemed unnecessary for CGR involvement. By applying prioritization and admission algorithms, the program improved the success rate of complaint handling by 36% and increased the effectiveness of warnings by 27%. According to these preliminary results, the program is an effective initiative to improve the internal processes of the CGR and contribute to the fight against corruption in Peru. Complaint systems enhancement involves the implementation of AI-driven recommendations based on previous steps or actions, thus optimizing resources and shortening the length of investigations.

AI in the Fight Against Corruption: Risks and Challenges

Despite growing social interest in AI's enormous potential and widespread agreement on its positive impact in combating corruption (Colonelli et al., 2020), there are direct and indirect risks that need consideration. Public figures from the field of technology, academia, government, and journalism have signed a letter highlighting AI's risks and advocating for mitigation strategies, prioritizing them globally alongside issues like post-pandemic management and nuclear warfare. Some of these risks are briefly discussed below.

System Manipulation by Corrupt Actors

Corrupt use of AI may occur when officials implement AI systems to obtain personal gains (Köbis et al., 2022). According to widely accepted definitions of corruption, public servants could exploit

the technological systems at their disposal to commit illicit acts and abuse their discretion. Since IA is a novel technology, opacity in its design, manipulation, and implementation can obscure understanding of the decision-making process, potentially eroding user trust. Such manipulation may not involve overtly corrupt use but rather exploitation of system vulnerabilities.

Risks of Mass Surveillance and Civil Rights Violations

While corruption remains a major concern for Latin Americans, insecurity is viewed as a burden in the region. Latin America has 9% of the world's population, yet it registers one-third of the global homicide rate. Mexico stands out in the region, with over 30,000 annual murders amid territorial in-fighting involving at least a dozen cartels.

It comes as no surprise that, during recent elections, former presidential candidate Marcelo Ebrard proposed a security plan—ANGEL (Advance Geolocation and Security Standards)—featuring face-recognition cameras and other devices to create an AI-based ecosystem across Mexican databases. ANGEL involved mass surveillance and biometric technologies in public spaces to introduce predictive surveillance, legislative efforts to implement cameras with face-recognition capabilities, vehicle geolocation, gait-based morphological criminal identification, drone use, and intelligent body cameras for the Mexican National Guard.

Notably, various studies have shown that mass surveillance and facial-recognition technologies are highly susceptible to misidentification leading to numerous cases of wrongful arrests of innocent people, especially those who are not white.



Over-Reliance on Technology

Over-reliance on technology can render institutions or governments vulnerable to cyberattacks and technical failures. The experience of Albania provides an example of the perils associated to excessive dependence on technology. As part of its process to join the European Union, Albania will be the first country to deploy AI, using ChatGPT—the most popular LLM model in the world, with one million users in its first week—to translate thousands of pages containing EU policies and laws into shqip and integrate them into Albanian legislation. This process will take place after an agreement with OpenAI, whose executive technology director, Mira Murati, is of Albanian origin. While this implementation will save the state apparatus time and resources, it calls for consideration of its ethical implications, especially regarding the legal void in terms of privacy, transparency, and an over-dependence on technology.

Algorithmic Biases in Identification

AI systems are trained on pre-existing data, oftentimes reflecting the exact biases (conscious or unconscious) occurring in the real world. This phenomenon, known as *algorithmic bias*, can lead AI-based anti-corruption tools to produce false positives, potentially reinforcing inequalities and discrimination. For instance, if historical data reveal a number of corruption cases involving people of specific ethnic origin, age, or occupation, an algorithm designed to detect future corruption cases may incorrectly flag individuals with these characteristics who have never been involved in illicit acts.

Studies in the United States show that facial recognition tools are less accurate for darker skin tones. These tools have been trained with existing data repositories where white men are overrepresented. Such inaccuracy is especially concerning when AI is deployed for policing. Training systems with biased data amounts to failure in

identification and, ultimately, to data misuse. To ensure efficiency and wide coverage of AI technology in security matters, it is crucial to involve experts who skillfully filter and analyze the data feeding AI systems.

Final Thoughts on the Potential of AI in the Fight Against Corruption and the Future of Public Governance in the Digital Age

While AI holds immense potential as an anti-corruption tool, addressing its ethical implications is essential to ensure its responsible and transparent use. Understanding how to leverage AI to analyze, predict, and automate data in anti-corruption initiatives, whether these originate in state agencies or civil society organizations, requires mechanisms that promote transparency, accountability, and bias mitigation to guarantee good governance for AI.

AI-ACT's challenges and opportunities may arise in anti-corruption agencies (top-down) or among civil society organizations (bottom-up). Agencies may perpetuate existing power asymmetries and inadvertently create dire consequences in their fight against corruption. On the other hand, civil society efforts face barriers to accessing open data, largely due to each government's level of digitalization. However, with the support of social media and under citizen scrutiny, AI-ACT initiatives offer great advantages in disseminating information about corruption cases and risks in real time.

Although anti-corruption AI initiatives expedite processes, they must never go without human oversight. The opacity of algorithm design and implementation remains a challenge for public trust,

especially in Latin America, where interpersonal and institutional trust is low. According to the Inter-American Development Bank, nine out of ten Latin Americans distrust others. This lack of trust encompasses the Judiciary in almost all countries in the region, to which we must add legislative gaps and the discretion held by political and economic elites. In such context, implementing effective anti-corruption mechanisms backed by citizens constitutes a challenge, whether or not they involve AI.

Technological innovation aside, sanctioning agencies must implement AI systems with utmost care to avoid exacerbating inequality and targeting individuals based on algorithmic bias. For instance, public officials from vulnerable or minority communities could face disproportionate investigation or sanctions, even as their actions are no more corrupt than those of other officials. Such bias not only perpetrates individual injustice but reinforces social stereotypes and diminishes the perceived fairness of anti-corruption efforts.

When citizens perceive that systems target specific groups instead of treating all officials equally, trust in the anti-corruption institutions dwindles. Furthermore, a lack of transparency about investigative decisions can contribute to mistrust. Some governments, including authoritarian regimes, use AI discretionarily—for instance, to target anti-racism activists in Miami and New York, feminist groups in the United States, feminist groups in Mexico City, pro-democracy groups in Hong Kong, journalists and political opponents in Egypt, and even to systematically monitor, profile, and persecute ethnic minorities such as Uighurs in Xinjiang.

Algorithmic transparency is a key strategy for risk mitigation. Explaining

how algorithms work and what decisions they make allows audits and monitoring, thus minimizing the perpetuation of bias. Diversifying training sources also helps to feed algorithms more accurately while protecting user privacy.

The goal is to increase accuracy and trust in corruption risk predictions without undermining the expertise of human auditors. This is a significant task in Latin America, where there are large digital infrastructure gaps—only 57% of citizens have mobile internet access, with disparities between South America (77%) and Central America (37%), and extreme differences between countries (e.g. Brazil at 77% and Haiti at 6%). Closing infrastructure and digitalization gaps between and within countries is crucial to ensure fair implementation and deployment of AI architecture across regions in the fight against corruption. Such implementation should consider rural and urban disparities, the need for neutral systems that minimize bias, and the inequalities that AI training may bring about.

Governments, companies, and civil society must commit to harness AI's transformative potential through constant vigilance, new regulatory frameworks, and ethical standards. Let us remember that, while the corrupt actions of an individual may reach a limited number of people, corrupting an algorithm could impact thousands in an instant. ♦

References

- AARVIK, P. (2019). [Artificial Intelligence – a promising anti-corruption tool in development settings?](#) U4 Anti-Corruption Resource Centre.
- ADAM, I., & FAZEKAS, M. (2021). [Are emerging technologies helping win the fight against corruption? A review of the](#)

[state of evidence](#). *Information Economics and Policy*, 57, p. 100950.

- COLONNELLI, E., GALLEGU, J. A., & PREM, M. (2020, December 26). [What Predicts Corruption?](#)

- DÁVILA PÉREZ, J. (2021). [Impacto y beneficios de las reformas en los sistemas de contratación pública en América Latina y el Caribe](#). *Red Interamericana de Compras Gubernamentales*.

- KÖBIS, N. (2023). [Bribes for Bias: Can AI be corrupted?](#) *Transparency International Blog*.

- KÖBIS, N., STARKE, CH., & EDWARD-GILL, J. (2022). [The corruption risks of artificial intelligence](#). Transparency International Working Paper.

- ODILLA, F. (2023). [Bots against corruption: Exploring the benefits and limitations of AI-based anti-corruption technology](#). *Crime Law Soc Change*, 80(4), 1-44.

- PIERRI, G., & LAFUENTE, M. (2022). [Human Talent Management and Corruption Control: The Effect of the New Talents in Government Control Program on the Detection of Corruption in Peru](#). IADB Discussion Paper, IDB-DP-952.



Denisse Rodríguez-Olivari

PhD in Political Science (Humboldt-Universität, Berlin). Master's degree in International Development (University of Manchester). Bachelor's degree in Political Science and Government (Pontificia Universidad Católica del Perú). Research Associate at the University of Glasgow, Adam Smith Business School. Expert in anti-corruption and integrity.

X: @DenisseROlivari

IG: @denisser

LI: denisserodriguezolivari



Artificial Intelligence and Public Safety in Latin American Democracies

GUSTAVO FLORES-MACÍAS Y BÁRBARA HERNÁNDEZ

Artificial intelligence is revolutionizing various areas of public policies. Its application in the fight against crime permeates contemporary political discourse and appears in the public security proposals of presidential candidates. Ensuring that AI is used responsibly and ethically is essential to maximize its benefits and mitigate the risks it poses to civil liberties in Latin American democracies.

Recent breakthroughs in artificial intelligence (AI) are revolutionizing a number of policy domains. Public safety is no exception, with AI and machine learning applied to policing and law enforcement more generally. The public safety uses of AI-based technology range from systems that identify violators of traffic rules, to those that predict where future crime is likely to occur, to those that help prevent online fraud to forensic DNA testing.

Several uses of AI for public safety, defined as “the application of algorithms to large sets of data to either assist human policing or replace it,” have received considerable attention in the United States and other advanced industrialized countries. Among the most prominent uses are face recognition systems—which capture individuals’ unique facial features and sort through potentially millions of possibilities to establish a suspect’s identity—and license plate reading systems—which can capture, identify, and match license plate numbers to vehicles and their owners and find their whereabouts.

Some of these applications, including the recognition of vast numbers of images or patterns of behavior to predict illegal activity or the assignment of crime

propensity scores to individuals based on risk factors, remain controversial due to ethical and privacy considerations. Although its adoption shows great potential to address crime, AI also presents important challenges for liberal democratic contexts in which the protection of civil liberties is paramount.

Despite these concerns, many Latin American countries have quickly turned to adopting AI technologies for law enforcement purposes. Proponents point to speed, the ability to analyze vast amounts of data that would be prohibitive for humans, and the reduction of human errors and bias as important benefits that warrant the widespread adoption of AI for public safety.

Mindful that AI-based crime-fighting technologies remain understudied in Latin America compared to the United States, this text presents a primer on the current state of AI-based crime-fighting technologies in the region’s democracies. In the following sections, we turn to examples of AI used to enhance public safety in Latin America. Through these examples, we then illustrate their intended purposes, as well as the opportunities and challenges that they represent.

The Promise of AI for Public Safety in Latin America

AI technology to address crime holds great promise in Latin America. With only around 8% of the world's population, the region accounts for about 30% of global homicides. Additionally, while there is considerable variation in terms of rates of violent crime across countries—from Chile, with a rate of about 4.5 homicides per 100,000 people, to Ecuador, with a rate of about 44 per 100,000—national averages can mask considerable higher rates within countries.

In this context, AI-based technology can help law enforcement agencies carry out several tasks more efficiently and effectively. In particular, AI can assist with the collection and analysis of massive amounts of crime data and predict potential hotspots for criminal activity, allowing law enforcement to deploy resources more effectively. By identifying patterns in time, location, and behavior, AI can help resource-constrained law enforcement agencies preemptively act to deter crime.

AI can also enable the region's police to carry out real-time monitoring to maintain public safety. AI tools can help monitor surveillance cameras, social media, or other open sources of information to detect suspicious activity and alert authorities as events unfold. Similarly, AI can be leveraged for facial recognition and biometrics toward the identification of suspects or missing persons more efficiently, particularly in crowded or chaotic environments.

On the forensics side, AI tools can play an important role in evidence management and analysis. They are valuable to organize and sift through vast amounts of digital evidence, improving the efficiency of investigations and judicial proceedings.

In doing so, AI has the potential to help address human biases and guide decision-making based on objective data to lead to more transparent and just outcomes.

AI Use Examples in Latin America

Because of the region's pressing concern with violent crime, Latin American governments have looked to incorporate technological advancements in AI towards public safety. While the region's adoption of AI for public safety has not been linear—rather, there has been considerable trial and error—several examples of its adoption are encouraging. As the number of cases increases, and as governments and publics become accustomed to the technology, its adoption is likely to become more generalized.

Governments at all levels have incorporated AI for public safety in the region. In Colombia, the National Police published a national strategy to provide local governments with resources to purchase unmanned drones and surveillance cameras to prevent and detect crime. To connect the vast amounts of criminal data it held and the various points of information it had to make decisions, the government hired an external software from Amazon Web Services to organize and store data from several sources. Apart from analyzing results quickly, the software also leveraged a series of tools including Nuvu's XCrime app to aggregate and analyze information as well as predict potential crimes.

In Chile, the National Insurance Association developed AI software to identify and report stolen vehicles. The software turns cellphones into license plate readers, which police departments can easily use. It automatically identifies stolen vehicles by scanning license plates and

comparing them to a database in seconds. The program was initially rolled out in 60 municipalities and expanded to 345 in 2022.

Efforts have taken place at the state and local levels as well. In the state of Mexico, the United Nations Office of Drug and Crime (UNODC) partnered with the country's statistics agency INEGI to develop a program that leverages deep neural networks to identify key words and phrases from previous 911 calls suspected of domestic violence or crimes against women. At the municipal level, local governments have also made efforts to incorporate AI—especially AI-assisted surveillance programs—to address crime.

The government of Benito Juárez, one of Mexico City's 16 boroughs, adopted a public safety strategy in 2018 to address crime through AI-based public surveillance tools. The municipality increased the number of police units and created a wide surveillance network comprised of the government's cameras and private cameras given to private citizens as part of public safety kits. Footage is linked to a Control and Command Center (C2) which leverages facial recognition and license plate technology to identify and track suspects once a report is filed. The program also leveraged citizen participation by giving them access to Blindar BJ - an app where citizens can report crimes and track police unit locations. In 2021, the program was expanded to the neighboring borough of Álvaro Obregón.

In 2018, the municipal government of Tlajomulco de Zúñiga, in the state of Jalisco, also sought to improve public safety by integrating AI-based technology into surveillance systems, including monitoring public areas with video surveillance cameras and offering citizens publicly-available information on crime incidence. The effort involves 564 cameras for face recognition and license plate readers to monitor

» Latin American governments have looked to incorporate technological advancements in AI towards public safety [...] there has been considerable trial and error –several examples of its adoption are encouraging «

behavior, identify patterns, and report suspicious activity.

In Brazil, the government of the city of São Paulo inaugurated in 2022 the use of an AI-based facial recognition system to monitor Line 3 of the subway system through 14,000 cameras in 18 stations. The city is also expanding its biometrics-based surveillance beyond the subway through the Smart Sampa program which will link ~20,000 cameras equipped with facial recognition to a monitoring center and to other agencies' databases for tasks ranging from identifying subjects to locating ambulances. Although the project has received criticism over potential biases, the city's mayor announced the center's launch in April and expects to fully implement the video surveillance system by the end of 2024.

In Uruguay, as Montevideo started experiencing higher crime rates, the city's police department adopted in 2016 a predictive policing tool named PredPol. By leveraging historical crime data, the software generated crime predictions maps to inform police efforts to deter crime or respond quickly to emerging threats. PredPol was adapted to Montevideo from the tool first used by the Los Angeles Police

Department in California to generate crime predictions based on crime type, location, and date/time. At a cost of US\$ 140 million annually and with a proprietary algorithm to which governments are not privy, this software helped to streamline police deployment and preemptively address crime. However, the government ended the program after finding no differences between zones that used the predictive software and those that did not.

Challenges for the adoption of AI for public safety

While these examples show the promise of AI-based technology for policing and the enthusiasm it has generated among governments as a potential aide in addressing rising levels of crime, the use of AI in law enforcement poses several important challenges. Facing public pressure to deliver public safety results, governments have moved quickly to adopt AI technologies toward this end, but they often do so without proper legislation in place to protect privacy and civil liberties, promote transparent procurement processes, and ensure the sustainability of AI. In this section, we discuss some of the main challenges facing governments in the adoption of AI technology for public safety, actions taken to address these shortcomings, and opportunities.

Regulatory gaps

AI Regulation is crucial to prevent its misuse and anticipate unintended consequences. While most Latin American countries lack AI legislation in general, and for public safety purposes in particular, some are in the process of generating regulatory frameworks (e. g., Argentina and Brazil)

» AI can be used to monitor citizens' lawful activity, as is the case in authoritarian regimes, such as China.«

based on multilateral regulatory regimes such as the European Union's AI Act or with support of the Inter American Development Bank. However, as regulators grapple with safeguarding rights without stifling economic growth, debates on how to approach AI regulation continue. For example, Brazil has two regulatory legislative proposals: the first places fewer restrictions on AI, as it aims to create "a decentralized system and restricts government intervention," while the second includes principles modeled after the European Union's AI Act. While AI legislation is less common, some countries such as Chile, Brazil, and Colombia have published national AI strategies that provide guidelines to advance AI adoption.

Additionally, international guidelines are also influencing the creation of AI frameworks in the region. For example, Brazil aligned its national AI strategy with several OECD principles. Further, several Latin American countries recently signed the Santiago Declaration that encourages more active participation in AI deliberations and shows a commitment to creating governance frameworks tailored to Latin America's own needs.

Privacy and Civil Liberties

Although AI can be an effective tool to address crime, it also carries important risks for human rights. For example, one common use is to monitor public spaces

to identify potentially harmful behavioral patterns and prevent crimes. Although this can be beneficial, surveilling public spaces to predict crime has deeper implications, as this constant monitoring can turn everyone into a potential suspect even when no crime has been committed. In other words, guilt is anticipated and estimated instead of giving people the benefit of innocence until proven guilty.

In this context, an important concern is the potential use of AI to undermine the protections afforded to citizens under liberal democracy. In particular, AI can be used to monitor citizens' lawful activity, as is the case in authoritarian regimes, such as China. Facial recognition, license plate readers, or cell phone location transmitters can be used to track individuals, even when they have not committed a crime.

A few examples from the region are illustrative. In Argentina, the police mistakenly logged the wrong variables to identify a suspect and instead jailed an innocent man for six days. In Mexico City's Blindar Benito Juarez Program, some users noted the lack of a privacy disclosure notice when downloading the government's app, raising concerns over users' privacy. In Ecuador, there are reports that the government's intelligence agency relies on the ECU-911 crime surveillance system imported from China to spy on journalists and politicians for political advantage. These incidents show how AI technology can undermine civil liberties if governments put it to the wrong use.

Procurement and Corruption

Government procurement processes in Latin America are often opaque, and the acquisition of AI-based technology is no

exception. Governments in the region have been quick to adopt AI for public safety because of the rapid growth of organized crime, but transparency in the acquisition processes has lagged. A report by AccessNow found that some foreign firms that supplied services in Latin America denied selling surveillance tools, rephrased their purpose, or deflected accountability to end users.

In the case of Mexico City's Blindar BJ and Alvaro Obregon, the lack of transparency over funds' allocation has raised eyebrows. As reference, the government invested MX\$385 million (about US\$19.6 million) in a span of three years, but it has faced allegations of inflated equipment valuation. While market research estimated that each camera cost around MX\$2,700 (US\$159), the borough government budgeted them for MX\$35,000 (US\$2,000).

Technology Adoption and Maintenance

Apart from requiring more robust regulatory frameworks, Latin American countries also tend to lack sufficient infrastructure to ensure technologies function properly. In Bogotá, Colombia, the city government adopted a system to identify and predict crime based on AI. With a significant investment of \$11M, the system raised concerns about human rights and insufficient police to act on the information generated, but in particular it was criticized because an estimated 22% of its cameras were not working properly. Further, differences in the software used across AI platforms limited information sharing and efficiency—as was the case with the softwares for the public bus rapid transit system Transmilenio and for the central command center for the city's surveillance cameras, C4.

Similarly, in Mexico City, the Blindar Benito Juárez program faced important implementation challenges. In particular, neighbors complained that cameras were not recording, making it difficult to present evidence when crimes took place. These challenges are likely to shape perceptions of the usefulness of AI for law enforcement if left unaddressed.

“Black-Box” Decision Making

The algorithms employed in the AI software are typically not made known to government agencies. In most cases police departments relying on the technology will have little knowledge about how input data are weighted or the extent to which biases can be present. In the case of Montevideo’s adoption of AI, high costs (US\$140 million per year), dependence on a fixed training data set, and biases in the reporting of historical crime data were important concerns. Since the algorithm used historical police data to feed the algorithm, critics noted the model could direct attention to already policed areas, creating a biased feedback loop. Further, given the software’s proprietary nature, authorities did not have access to the software’s algorithms. Similarly, with Brazil’s Smart Sampa project, human rights organizations have highlighted its potential to incarcerate Black and low income individuals more frequently than other individuals.

AI in the public eye

As Latin America’s use of AI-technology for law enforcement increases, so does its presence in the public domain. In particular, AI-based technology to fight crime is

entering contemporary mainstream political discourse, including presidential candidates’ public safety proposals. For example, during Panama’s 2024 presidential election, several candidates proposed leveraging AI technology to prevent crime. President José Raúl Mulino proposed additional training and procurement of new AI technology to prevent crime during his campaign, and former president Martín Torrijos, advocated during his campaign for the use of “... technology and surveillance and monitoring systems with artificial intelligence” coupled with increased police force units.

Similarly, candidates in Mexico’s presidential and mayoral elections have also campaigned on the use of AI to curb crime, ranging from installing the highest number of surveillance cameras to expanding public safety programs that have leveraged AI technologies such as surveillance cameras and real-time police unit tracking. As the right-of-center candidate in the 2024 presidential election put it, “we are all in on the use of technology and AI to address crime.” Although these efforts seem promising, critics are concerned that candidates place more emphasis on implementing or expanding programs without also considering the policies required to regulate these technologies and ensure their responsible use.

Despite Latin American countries’ controversial history of government surveillance on civil society and political opponents, there does not seem to be a strong opposition to the technology. In fact, recent polls show citizens may have a positive outlook towards AI as well. In 2023, Ipsos, a global market research firm, surveyed over 22,000 individuals across 31 countries, including four in Latin America: Mexico, Colombia, Chile, and Argentina, asking about public perceptions of AI.



Globally, around 67% of respondents strongly or somewhat agreed that they had a good understanding of AI. Comparatively, Latin American averages were generally higher (Mexico - 75%, Colombia 73%, Chile 70%, and Argentina 67%). Similarly, while 54% of respondents worldwide strongly or somewhat agreed that AI products and services provide more benefits than drawbacks, the share of respondents sharing this view was higher in Latin America: 73% in Mexico, 65% in Colombia, 59% in Chile, and 57% in Argentina. Notably, attitudes toward AI among Latin American countries generally were more favorable than those in European counterparts.

Conclusion

The adoption of AI-based technologies for public safety in Latin America's democracies offers significant promise

in addressing the region's high rates of violent crime and improving law enforcement efficiency. From predictive policing to facial recognition and real-time surveillance, AI tools have the potential to transform crime-fighting efforts by enhancing data analysis, resource deployment, and forensic investigations in a region where drug trafficking and organized crime more generally are on the rise. However, the rapid integration of these technologies also presents substantial challenges for liberal democracies, including regulatory gaps, privacy concerns, and issues of transparency in government procurement. The risk of civil liberties being undermined, coupled with the lack of proper infrastructure and maintenance, underscores the need for a more cautious and regulated approach to AI implementation.

As AI becomes more prominent in political discourse and public safety

strategies, governments must prioritize establishing robust legal frameworks that balance innovation with human rights protections. Ensuring that AI is used responsibly and ethically will be key to maximizing its benefits while mitigating the risks it poses to civil liberties in Latin American democracies. ♦

References

- A otro nivel. Plan de gobierno 2024-2029 Martín Torrijos. (2024).
- ACCESS NOW. (2021, August 10). Made Abroad, Deployed at Home.
- ACCESS NOW. (2023, April). Remote biometric surveillance in Latin America. Are companies respecting human rights App “Blindar Benito Juárez” te permite ver cámaras de seguridad; así funciona. (2021, April 8). *El Herald de México*.
- ASOCIACIÓN DE ASEGURADORES DE CHILE. (2022, June 22). Asociación de Aseguradores firma convenio con municipios para detectar y denunciar autos robados. Benito Juárez presenta su centro digital de vigilancia. (2021, April 7). *El Universal*.
- COSME, M. (2021, November 17). Lía Limón solicita 5 mil 784 mdp de presupuesto 2022 para Álvaro Obregón. *El Sol de México*.
- CUMBRE MINISTERIAL Y DE ALTAS AUTORIDADES DE AMÉRICA LATINA Y EL CARIBE. (2023, October 24). Declaración de Santiago.
- DÍAZ, O. (2023, November 13). Miguel Hidalgo, Álvaro Obregón y Azcapotzalco piden aumento de presupuesto para el 2024. *El Universal*.
- EGA. (2024). Artificial Intelligence. Latin America’s Regulatory and Policy Environment.
- FREITAS, H. (2024, July 4). Smart Sampa: Pre-feitura inicia programa de câmeras com reconhecimento facial em São Paulo. *O Globo*.
- GAUDÍN, A. (2017, May 5). Uruguay Tries Preventative Policing with a High-tech Twist.
- GOBIERNO DE JALISCO. (2021). Plan Municipal de Desarrollo y Gobernanza. Tlajomulco de Zúñiga.
- GOBIERNO DE LA CIUDAD DE MÉXICO. (2023, 28 de marzo). Síntesis informativa. Alcaldía. IA al servicio de la seguridad pública: el éxito de la Policía Nacional de Colombia prediciendo crímenes en AWS. (2023). AWS.
- IPSOS. (2023, July). Global Views on A.I. 2023. How people across the world feel about artificial intelligence and expect it will impact their life.
- JOH, E. (2018). Artificial Intelligence and Policing: First Questions. *Seattle University Law Review*, 41, 1139-1144.
- KESSEL, J. (2019, 26 de abril). In a Secret Bunker in the Andes, a Wall That Was Really a Window. *The New York Times*.
- LUM, K. (2016, 10 de octubre). Predictive Policing Reinforces Police Bias. hrdag.
- MANJARRÉS, J., & NEWTON, CH. (2024, February 21). InSight Crime’s 2023 Homicide Round-Up. *InSight Crime*.
- MARI, A. (2023, July 13). Facial recognition surveillance in São Paulo could worsen racismo.
- MASCELLINO, A. (2022, December 9). Brazil deploys iss facial recognition to secure São Paulo metro. *biometricupdate.com*.
- MINISTERIO DEL INTERIOR, MINISTERIO DE DEFENSA NACIONAL, CONSEJERÍA PRESIDENCIAL PARA LA SEGURIDAD NACIONAL. (2020). Política Marco de Convivencia y Seguridad Ciudadana 2019-2022. Dirección de Antinarcóticos.
- MUNICIPIO DE TENO. (2024). Municipio de Teno instala lectores de patentes para detectar autos robados.
- NAUNDORF, K. (2023, September 15). Un escándalo en Buenos Aires revela los peligros del reconocimiento facial. *Wired*.

NUVU. (2023). [Policía Nacional de Colombia](#).
 OECD.AI Policy Observatory. (2021). [National AI policies & strategies](#).
 OECD.AI Policy Observatory. (2022, September 6). [Brazilian ai Strategy](#).
[Plan de Gobierno del candidato a la presidencia José Raúl Mulino](#). (2024).
 RANGEL, L. (2024, March 7). [Tecnología en propuestas de seguridad de candidatas: uso de datos biométricos y videovigilancia pone en riesgo derechos humanos](#). *El Sabueso*.
[Sin transparencia el funcionamiento de las aplicaciones de seguridad de la BJ](#). (2021, February 25). *DDM Benito Juárez*.
[¿Sistema de videovigilancia de la Ciudad ¿en cuidados intensivos?](#) (2022). Concejo de Bogotá.
 TENORIO, G. (2018, May 6). [Santiago Ta- boada hace público su Plan de Gobierno 2018-2021 para la alcaldía de Benito Juárez](#). *Periódico Leo*.
[Tlajomulco presenta modelo de vídeo vigilancia](#). (2018). *Tlajo*.
 UNODC MÉXICO. (2023, January 30). [Inteligencia artificial para detectar y prevenir la violencia contra las mujeres](#).
 VALENCIA GÓMEZ, M. (2021, January 26). [El reto de anticipar delitos con tecnología en Bogotá](#). *El Espectador*.
 VÁSQUEZ CRUZ, E. (2023, September 27). [Inteligencia Artificial para mejorar la seguridad pública, una tendencia mundial](#). *Alcaldes de México*.



Bárbara Hernández

B.A. in International Relations and Psychology from Stanford University. Independent researcher.



Gustavo Flores-Macías

Professor of comparative government and public policy at Cornell University, USA. Research affiliate of the Cornell Tech

Policy Institute.

X: @Gustavo_F_M

Ll: gustavo-flores-macias

A portrait of Lars Zimmermann, a man with short, light-colored hair, wearing a dark sweater. He is looking directly at the camera with a neutral expression. The background is a blurred indoor setting with bright, diagonal light streaks on the right side.

Lars Zimmermann:
“Technology
won’t relieve us of
democracy’s tasks”

Today's AI still relies on people to work alongside it, control and guide it. What specific political responsibilities come with ensuring good governance in light of this new technological advancement? We spoke with an expert in public administration modernization and innovation about how AI is transforming the state.

Lars Zimmermann is a co-founder and board member of GovTech Campus Deutschland e.V., the world's first innovation, development, and learning hub dedicated to modernizing government and administration. The German government is promoting the Campus as a collaborative platform aimed at developing digital innovations and technologies for federal, state, and local administrations in partnership with the tech community. The goal is to make these advancements available for reuse.

Before founding the GovTech Campus, Zimmermann worked as a technology and transformation consultant and was the founder and board spokesperson of the Stiftung Neue Verantwortung. He has been engaged for many years in state modernization and administrative reform, developing numerous initiatives and projects in this field. Since early 2024, he has been working as a research associate at the Konrad Adenauer Foundation.

—With the rise of AI, are we on the verge of a societal leap comparable to the arrival of the steam engine?

—Honestly, we don't know yet. I'm always a bit cautious about declaring a major revolution. However, I suspect that AI will have a similarly profound impact on society. We are

on the verge of a development leap that would have been unthinkable 15 years ago. Just four years ago, none of us were even discussing, let alone working with, large language models like ChatGPT.

There has been a significant disruptive breakthrough in AI that has, for the first time, reached the general population at various levels. If we extrapolate from this advancement, we must conclude that AI, much like the steam engine or other groundbreaking technologies, could represent a significant leap forward in development. Of course, this comes with risks, but above all, there are also numerous opportunities. In my opinion, this evolution has arrived at the right moment.

—Why did it arrive at the right moment?

—For example, globally, especially in industrialized countries, we face the challenge of a growing demographic gap. If we develop and utilize artificial intelligence effectively, we can help address these shortfalls in many professions. A second example is the enhancement of computing capabilities. With artificial intelligence and the strengthening of IT infrastructures, we can now process, manage, and contextualize data—capabilities that were not available to us ten years ago. These advances bring

significant benefits, particularly in health research and safety matters.

So, I believe that while we are in a time of significant challenges, we also have tremendous opportunities because of AI. We are witnessing major technological advancements that can help us combat diseases, develop medicines, and address workforce shortages.

—What time frame are we discussing now?

—It's a good question that no one can answer with certainty today. Perhaps we can frame it this way: short-term innovative leaps in AI are often significantly overestimated, while those in the medium and long term are greatly underestimated. If we look back, four years ago, nobody knew what a large language model was. I am confident that we will witness significant leaps in development over the next 15 to 20 years.

—And in the short term?

—In the short term, for example, large language models can lead to efficiency gains. These models are already quite powerful and capable of handling many tasks. However, they cannot, for instance, replace advisors in a federal ministry. Today's AI still needs people to work alongside it and to oversee and guide its operations. Clearly, this could change when technology and computing capabilities advance to the point that AI can take on certain tasks without human assistance.

Naturally, these changes will also impact us. AI could assume tasks currently performed by humans, making the corresponding jobs

superfluous. Despite this, I do not believe that AI will result in mass unemployment. On the contrary, I see AI as a means to address gaps created by labor shortages. It is not about reducing resources but rather about filling those gaps.

—AI is often viewed as a threat and is frequently linked to job losses, fake news, and deepfakes. So, what are the opportunities?

—For one thing, AI can significantly enhance administrative processes. It offers administrative staff a tool that allows them to perform their tasks more efficiently and effectively. A good example of this is large language models, which enable texts to be written and synthesized quickly.

Suppose that ministerial advisors need to create templates. Often, a minister will ask for a long text to be condensed into a concise page. In the past, this used to take a considerable amount of time. With today's large language models, this process can now be easily automated. What once took an hour can now be completed in just a few seconds. This saves a significant amount of time and boosts efficiency.

Another example is this very interview, which we conducted in German but will later be printed in Spanish. In the past, translation would have taken a lot of time and money. Today, thanks to large language models like ChatGPT, texts can be translated quickly and efficiently into any language. This considerably simplifies and accelerates the entire process.

—However, we cannot replace the translator we currently employ. We still need someone to review the solutions generated by the language models.

—I don't believe it's wrong for a human to continue overseeing translations. However, it is likely that translation technology will advance to the point where human oversight may no longer be necessary. The error rate could

» IA, much like the steam engine or other groundbreaking technologies, could represent a significant leap forward in development. «

then be comparable to that of a simultaneous translator, who is also not error-free. This development is likely to occur relatively quickly. However, I do not believe that translators will be out of work. When it comes to analyzing and interpreting speech, as well as assessing nuances that indicate self-confidence or insecurity in the voice, AI is not yet ready. These skills require human perception, so it will take time before AI reaches the same level in this area. However, it is very likely that it will get there eventually.

—Does this not imply that primarily routine tasks requiring average qualifications will be replaced, thus eliminating many jobs, while a shortage of highly qualified specialists remains?

—I'm not sure about that yet. I believe mid-level positions are more at risk than very low-level ones. The more demanding a job is—meaning mid-level or higher skilled—the more likely it is to be replaced by AI. This also impacts many oversight roles in administration that are currently handled by more qualified specialists but could be automated in the future. For example, positions held by academics in the insurance industry could also be at risk from AI. So, it's not just basic jobs, but also those that require a higher level of qualification. In the next phase, AI may not make any job obsolete, but it will alter the traditional requirement profiles for those positions. Those who do not adapt to this change will find it challenging.

—We would like to hear your thoughts on how the development of AI will impact Latin America.

—Ultimately, the development of AI is still evolving. It is less about how proficient different countries are at developing AI and more about the capabilities of companies in this field. Currently, only a handful of companies are driving AI on a global scale, and most of them

are based in the United States.

The impact of AI on various social systems remains unclear. Opportunities and risks are present in industrialized countries as well as in other parts of the world. Take Africa, for example: many African countries have made greater progress than Germany in payment systems and digital support for microenterprises. This demonstrates that even less developed economies can benefit from technological advancements.

If economies in South or Central America continue to develop, there is a real risk that AI-supported innovations will also lead to increased labor efficiency and, consequently, job cuts. This means that governments in these economies must carefully examine the impact of AI to avoid being left behind and to make the most of the opportunities presented by this technology.

Governments [in economies like those in South or Central America] need to closely examine the impact of AI to avoid being left behind.

—Currently the largest companies in the AI sector come from the United States. The so-called network effect is particularly strong in this field, where the market leader gains significant advantages and leaves little room for competitors.

—There's a well-known phrase: "In the field of artificial intelligence, there is no third place anymore." You're either the world market leader or the one trying to catch up. But this

» Governments [in economies like those in South or Central America] need to closely examine the impact of AI to avoid being left behind. «

risk applies to all countries. In my view, this is not a typical issue of development politics.

It's important to understand that these challenges cannot be tackled by a single country alone. Especially in regions like South America, it would be crucial for countries to collaborate at the supranational and regional levels to effectively address the issue of AI. However, good governance remains essential. Poorly governed countries will continue to struggle to leverage technological innovations effectively. High-quality governance rooted in fundamental democratic values will be even more crucial in the face of rapid technological progress. AI cannot take on the task of good governance; this responsibility must stay in the hands of citizens and will likely become more important.

—What specific tasks should politics address to ensure good governance in light of these new technological advances?

—First and foremost, I believe politicians need to cultivate a strong curiosity about AI. Countries that thrive in times of innovation are often those that embrace new technologies with openness and curiosity. It's important not to fear or demonize technology. Openness to new developments is crucial.

Second, cooperation is essential. No country, not even the United States, can tackle AI alone and with the necessary depth. Even large nations like the United States and China face significant challenges. For smaller countries like Germany and France, cross-border collaboration is even more crucial. Therefore, a strong commitment to cooperation and working together on these issues is even more important.

The third step is also very important: creating the framework conditions that enable innovation.

The fourth step is the actual application of AI. It's important that countries and their citizens do not demonize AI but actively

» I don't believe that AI will automatically make us better democrats or more informed voters. «



» I am confident that we will witness significant leaps in development over the next 15 to 20 years.«

engage with it. Those who use AI also have an opinion. An example from the past: the Germans became world leaders in the automotive industry because the domestic market was strong, and Germans themselves bought cars in large numbers. In the field of

AI, no country has yet established itself as a leader in its application, whether in education or healthcare. Therefore, all regions still have similar starting points. In fact, countries that are not as saturated and have fewer existing structures may even have an advantage, as they could be more open to new developments.

The fifth step focuses on training. Education is crucial for all areas, including AI. The development and application of AI and other technologies will not succeed without top-tier training centers and educational systems. It is essential to have strong educational institutions that can prepare the next generation for new technologies and integrate them into the education system.

—What positive narratives could political parties promote about AI?

—There are many positive stories that can be told about AI. For example, it can help address our demographic challenges by making public services more efficient. It could help reduce the need for staff, which in the long run would lower administrative costs. It could also lead to better management of taxpayers' money and potentially even lowering taxes.

Another advantage is the boost in innovation. AI enables the processing of large amounts of data and makes efficient use of data centers. This can lead to groundbreaking (disruptive, revolutionary) research in the healthcare sector. I am convinced that I will live to see cancer no longer be a death sentence. With advancements in computing power and health data processing, we could develop treatments that defeat diseases once considered incurable twenty or thirty years ago.

Consider the example of the war in Ukraine. Without AI capabilities, Ukraine would not be able to use its drones as efficiently and effectively for defense, despite limited access to traditional weaponry. This AI capability, combined with drones, is crucial to the country's defense efforts.

» **Today's AI still needs people to work alongside it and to oversee and guide its operations.** «

These examples show that AI can bring breakthroughs across a range of political domains. Essentially, we can identify two broad areas: AI enhances efficiency and reduces costs, and it drives innovation and solutions in various fields.

—Will AI also lead to new forms of participation in a representative democracy?

—Before I answer, I'd like to mention that you're speaking with someone who strongly supports the traditional form of democracy. I am convinced that technology will not relieve us of democracy's essential tasks. What do I mean by this? I don't believe that AI will automatically make us better democrats or more informed voters. In my view, democracy is a system that functions through people and institutions. People will always be the central actors in a democracy, making decisions through majority vote. That's why I consider myself conservative, in the best sense of the word. I don't believe that AI-driven democratic systems will save democracy.

—In the U.S. election campaign, we saw that voters are becoming increasingly transparent to strategists through the analysis of large data sets. What impact does this have on democracy?

—That is not necessarily a problem. Strong democrats make strong democracies. You allude to the fact that parties now understand their voters better and can place targeted ads—but this is part of an ongoing development. If we look back at economic history, the invention of the printing press allowed handbills to be printed and distributed in mass for the first

time, and I'm certain the first political handbill appeared not long after.

The principle remains the same: technologies enable us to spread information faster and in real time. Is it risky for politicians to know more about the population? Not necessarily. It becomes dangerous when states have access to everything, with the power to consolidate this scattered information into a comprehensive data block and use it to shape policy. Here again, good governance will ultimately be the decisive factor.

AI could be indirectly blamed for this, because without it the state would not be able to collect this information to the same extent. But you have to be able to resist innovation. Democracy has survived the printing press and the invention of television, and it will also survive AI.

—What can people do to prepare?

—For one, it's important to stay curious and open-minded. As long you engage with new developments, you'll stay better informed and more self-confident. The biggest danger is becoming complacent and letting others constantly tell you what to think.

It's crucial to question things. For example, why do you see ads for a car on Instagram after talking about it with your friends – Is the smartphone listening? Can it recognize those sentence fragments? Asking these kinds of questions helps you stay critical of new technologies and make informed decisions.

It's important to develop a critical and constructive attitude. There's no value in immediately rejecting everything new or accepting it without question. Instead, try to form a balanced opinion and critically analyze what's new.

It's also useful to try things out. When you understand how something works, you can make a more informed decision about whether to use it or not. Germany has gained innovation leadership in many areas because people were willing to experiment with new

ideas. The same approach should apply to AI—people should explore and understand different applications.

For example, we're already using AI today, perhaps without even realizing it. My bank uses a simple AI system to detect suspicious transactions in my account. This type of AI protects us and makes many everyday tasks easier.

To sum up, openness to new ideas, critical thinking, and practical experience are essential to find a path in the AI era and recognizing its opportunities and risks it presents to each individual. In doing so, everyone can help move society and the country forward on this path.♦

German-Spanish translation by Manfred Steffen



Jonathan Neu

Deputy Head of Office at Regional Program Party Dialogue and Democracy in Latin America of the Konrad Adenauer

Foundation, based in Montevideo, Uruguay. Studied mathematics and history at the universities of Leipzig and Salamanca. He specialized in Intellectual history.

X: @NeuJonathan



Manfred Steffen

Master in Environmental Sciences, Universidad de la República, Uruguay. Printing Engineer by the FHD Stuttgart, Germany. Editor of Diálogo Político

and project coordinator of the Regional Program Political Parties and Democracy in Latin America of the Konrad Adenauer Foundation.

X: @ManfredSteffenN



Freedom and Truth Captured by Artificial Intelligence

MIGUEL PASTORINO

Artificial intelligence is a new reality with which we coexist, and it is also transforming who we are, how we perceive ourselves, and how we live—much like a new environment. Its impact is radically reshaping everything from education to medicine, from the economy and politics to work and interpersonal relationships.

The so-called artificial intelligence (AI) is not merely a technological leap; it has generated an anthropological shift, as it redefines human life, our ways of thinking and living, of knowing, learning, connecting, forms of power, and rights' limits. It reshapes our conception of freedom and truth. It stands as one of the most significant philosophical issues of our time. Because it's not just a tool, but it's beginning to merge with our environment and with who we are, transforming our daily lives, leaving no area untouched—from education to medicine, from the economy and politics to work and human connections.

As we have shaped machines, those same machines have shaped us. The world built by humans—the world of machines or sociotechnical systems—has, in turn, redesigned us, influencing our abilities and shaping our own values and beliefs (Savulescu and Lara, 2021).

AI is neither morally nor philosophically neutral. It is not an instrument, but a new reality with which we coexist, one that also transforms who we are, how we perceive ourselves, and how we live—a *new environment*. We find ourselves in

a dynamic of mutual interaction, of co-participation in operations.

The technological environment has its own dynamics and logic. It has not replaced the natural environment but has completely transformed and reconfigured it. "Technology is no longer a means but the Medium, the environment in which we live. It is a Medium because, through its self-regulation and automation, it behaves as something independent and constantly evolving, capable of surrounding us and continuously creating new contexts for human existence... New technologies interact with the environment around them, responding to stimuli and altering their behavior independently..." (Varela, 2022).

The fact that technology opens new spaces for action necessarily calls for a discussion about those actions, which are never neutral. This compels us to reflect on their social, economic, and political consequences, but first, we must ask ourselves what technology truly is and its impact on the human condition.

What Kind of Intelligence Are We Talking About?

Human intelligence cannot be reduced to functions that are metaphorically compared to those we attribute to AI. Confusing functions with the uniqueness of human intelligence is a common form of reductionism. We have grown accustomed to the borrowed use of concepts from cognitive sciences being applied to computer systems, often without the necessary conceptual clarifications. The terms we use shape our mental representations: we hear about synaptic chips, *artificial neural networks*, *neural processors*, etc.

“The principle of computational intelligence modeled after our own is flawed because the two hardly share any meaningful similarities” (Sadin, 2020), except when we fall into the trap of reducing intelligence to a set of functions, and reducing reality to binary codes, thus excluding the countless dimensions our human subjectivity can experience—dimensions that cannot be captured by mathematical models. “What we are faced with is a truncated, restricted, and biased understanding of the intelligence process, which is inseparable from its tension with a multisensory, unsystematizable grasp of the environment” (Sadin, 2020).

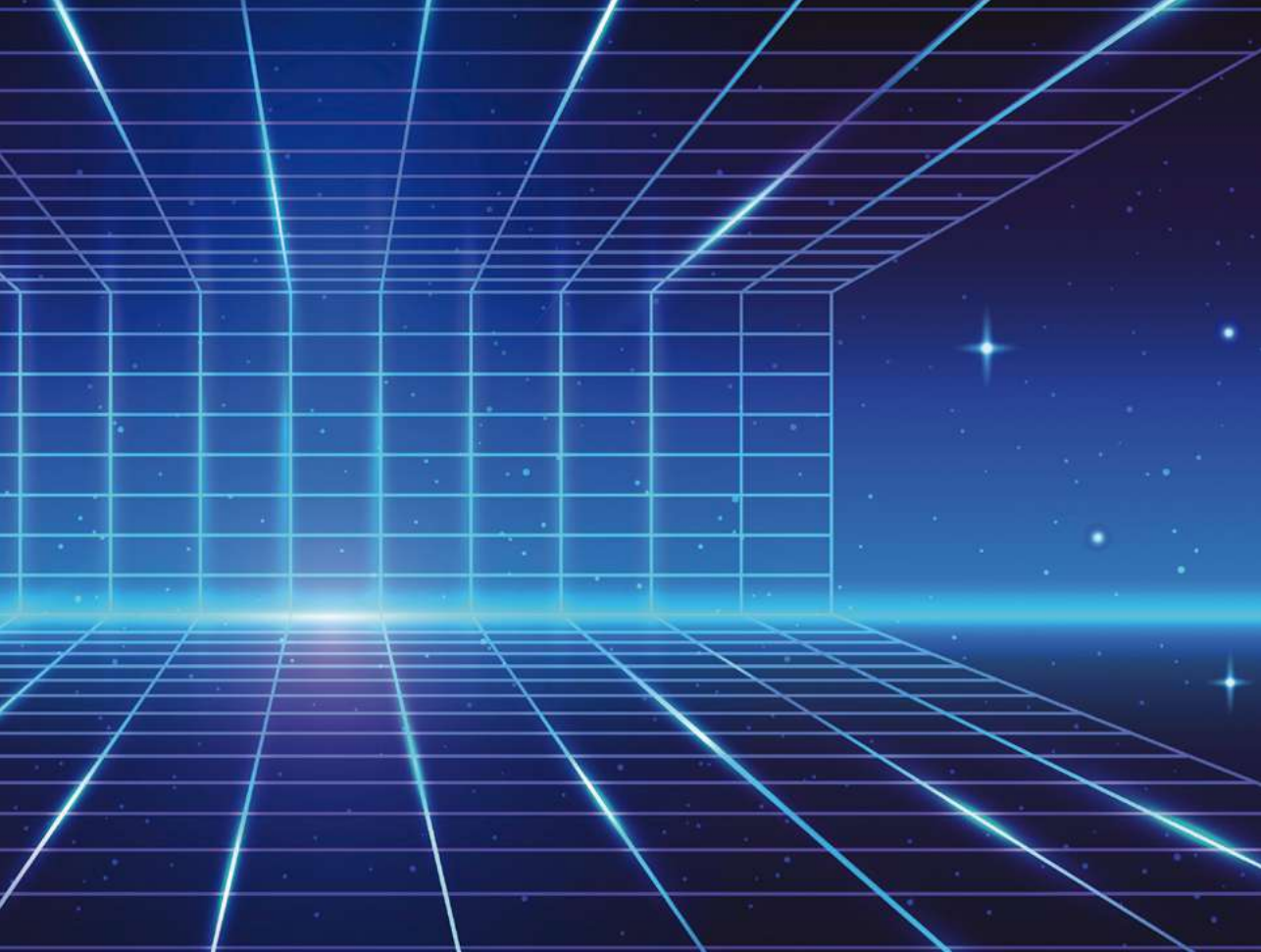
Big data emerged as a form of absolute knowledge, where hidden correlations between things are revealed, but we have neglected to ask ourselves about the meaning behind things, the ‘why,’ the ultimate reason behind events, and the purpose of life.

“Everything becomes calculable, predictable, and controllable. A whole new era of knowledge is proclaimed. In reality, it is a rather primitive form of knowledge. Data



» AI is neither morally nor philosophically neutral. It is not an instrument, but a new reality with which we coexist, one that also transforms who we are, how we perceive ourselves, and how we live—a new environment. «

mining uncovers correlations. According to Hegel's logic, correlation represents the lowest form of knowledge” (Han, 2021), because with correlations we do not know why things happen, we simply know that they happen.



Despite the impressive advances in generative artificial intelligence and the new transformations in science and technology, the truth is that we are not talking about intelligence in the human sense. While AI can perform, through machine learning, a range of functions that humans do—such as calculation, mathematical procedures, information selection, pattern recognition, and reproducing what it has learned—at a speed and with an amount of information impossible for any human being, this does not mean it thinks in a human way. AI lacks consciousness and subjectivity, even though it can simulate emotions and interact with humans by learning and reacting to the information it receives. The issue arises when we reduce intelligence to the ability

to calculate and process information. Machines do not produce wisdom because they lack subjectivity and self-awareness, even though they may simulate it and lead us to believe otherwise, and impress us.

AI is neither artificial nor intelligent. Rather, it exists in a tangible form, made up of natural resources, fuel, labor, infrastructure, logistics, histories, and classifications. AI systems are not autonomous, rational, or capable of discerning anything without extensive and computationally intensive training, relying on massive datasets or predefined rules and rewards. In fact, AI as we know it is entirely dependent on a much broader set of political and social

structures. And because of the capital required to build AI on a large scale, and the ways of seeing that it optimizes, AI systems are ultimately designed to serve existing dominant interests (Crawford, 2022).

Submission to the Artificial Oracle

Under the pretext of making the best decisions in all areas of life—finance, transportation, healthcare, sports, justice, and more—human affairs are increasingly being resolved from the lofty heights of artificial *superintelligence*, where larger quantities of data are processed. We are witnessing a growing reliance on artificial oracles, acting as gurus or spiritual directors, imposing daily routines as if they possessed superior and unquestionable knowledge. This process often starts at a basic level, such as coaching, where an app guides emotional life, nutrition, or relationships, prescribing how to think and act. But it can escalate to more prescriptive levels, where AI decides one's career future or determines eligibility for a bank loan. Sadin points to an even more radical stage we've reached—a coercive level where AI will ultimately decide on expenditures, cutbacks, and even administer justice.

Humanity is rapidly equipping itself with an apparatus that renders it increasingly dispensable—surrendering its right to make decisions with full awareness and responsibility over matters that directly concern it. A new anthropological and ontological framework is taking shape, in which the human figure submits to the equations of its own artifacts, with the primary objective of serving private interests and establishing a societal order based on



predominantly utilitarian criteria (Sadin, 2020).

It is essential to ask ourselves questions and engage in critical reflection on these matters. What challenges does AI pose to political philosophy? How do we address biases in AI programming when hiring, evaluating employees, or pursuing criminal justice, knowing that AI can hallucinate, make mistakes, and also discriminate? What will be the political effects of robotics in terms of justice and equality? What impacts does AI have on democracy, particularly concerning voter manipulation? How is it transforming



» The real issue lies in how much we are willing to surrender our freedom and which skills we are prepared to forfeit for convenience. «

journalism and news generation? How does it affect human relationships, learning, and mental health? What should be the degree of citizen participation in the regulation of AI? What implications does AI have for animals and agricultural production? What effects could it have

on climate and the environment? What would digital rights look like for data protection and ensuring respect for human dignity?

Moreover, we often confuse predictions with the future, as if a new form of superstition gives us certainty about a controllable or knowable future.

Artificial intelligence learns from the past. The future it calculates is not a future in the true sense of the word; it is blind to events. However, thought possesses an event-like quality. It brings something entirely different into the world... AI

merely selects from pre-existing options, ultimately between one and zero. It does not venture beyond what is already given into uncharted territory (Han, 2021).

Cognitive Sedentarism

What would happen if we asked someone to exercise for us, relieving us of the effort involved in such activities? The obvious answer: we would lose the opportunity to improve our physical condition and health, becoming physically atrophied, with all the consequences that need no further explanation. Even in the rhetoric of the gym, no one finds it excessive to speak of sacrifice, effort, dedication, and pushing oneself until it hurts; the more time and effort we invest, the better the results: No pain, no gain. Cultivating oneself as a person in all possible dimensions is an imperative present in every time and culture. Generally, everyone wants to be better than they are and to develop in various aspects of their lives. None of this feels strange to us. However, we live with a paradox regarding the care and development of our capabilities because the same does not apply to intellectual cultivation. What if the criteria we use for physical training were applied to intellectual life? Can you imagine a teacher today discussing sacrifice, effort, dedication...? Parents and colleagues would look at them with bewilderment, as if they were a dinosaur. Why is that?

It seems we are witnessing an atrophy of thought, a promotion of a culture that favors shortcuts and minimal intellectual effort. If someone can save us time in thinking, reading, writing, comparing, calculating, synthesizing, or analyzing, we thank them as if they are doing us a great favor. And now, thanks to generative artificial intelligence (GAI), we can avoid engaging in academic work that develops

essential intellectual skills, leading to brains that will become atrophied in fundamental capacities for clear thinking. It is not that using GAI collaboratively for study and work is without merit; the real issue lies in how much we are willing to surrender our freedom and which skills we are prepared to forfeit for convenience. The substantial risk is that we stop teaching the value of effort and concentration—the ability to sit focused on something challenging for hours with the purpose of solving it. How can we develop tenacity and resilience if we instantly abandon tasks for someone or something else to resolve, sparing us the effort?

Losing the ability to calculate, to maintain attention, or to engage in sustained, deliberate effort to solve a difficult problem is part of a phenomenon we refer to as cognitive sedentarism (Sigman and Bilinkis, 2023).

The best way to combat cognitive sedentarism is to convey our own passion for knowledge and the benefits of developing intellectual skills that enable us to think for ourselves with greater depth, without renouncing our freedom to choose who we want to be and where we want to go.

What Are We Willing to Lose?

In today's automated systems, computers often take on intellectual tasks—observing, perceiving, analyzing, evaluating, and even making decisions—those until recently were considered strictly human domains. The person operating the computer plays the role of a technology employee who inputs data, monitors responses, and looks for errors. Instead of opening new frontiers

» There remains a certain naivety in believing that everything can be solved with an increasing amount of data, as if the answers to human dramas depended solely on information management rather than on deep reflection about who we are and what we truly want to achieve for the future generations to come. «

of thought and action for human collaborators, the software narrows our perspective. We trade subtle, specialized talents for more routine and less distinctive ones (Carr, 2014).

Day after day, we risk becoming unable to write an email, create a shopping list, navigate our own city, devise a business strategy, or compose a message, speech, or essay. With great enthusiasm and comfort, we surrender to the ever-helpful invitations: “What can I do for you?” We feel simultaneously pampered and served by technology, while we elevate it to a superior instance that will do almost everything for us and will know how to do it better.

Can we imagine the effects on individual and collective psyches of being in a position where we expect everything— as if we were lounging on our sofa— from systems that resemble infinitely superior butlers? This environment fosters the atrophy of both our impulse toward outward engagement and our intellectual faculties... (Sadin, 2024).

According to Sadin, we are in an era where everything seeks to satisfy well-

defined objectives in real time, leaving no room for spontaneity or activities deemed useless or inefficient.

In the workplace, often the challenge matters more than the final result; the process and meaning of what we do provide us with a sense of fulfillment. Thus, in the professional world, the key to self-worth lies in the significance of our work and the knowledge that we are making a meaningful impact.

We appreciate the things we have created—our own works—simply because they are ours and we understand the effort they required. Perhaps in a few years, only a small minority will have access to those challenges that give life meaning. If that is the case, it could represent one of the greatest impacts of AI on the workforce (Sigman and Bilinkis, 2023).

Truth Reduced to Data

AI performs data management functions that far exceed our capacity and speed, but it does not replace other human abilities related to how we connect with one another or the meaning of life— issues that cannot be resolved through data, statistics, or patterns. Reducing knowledge to mere information fosters a naive optimism about AI’s various possibilities regarding human life.

Regardless of the direction AI development takes, we cannot delegate responsibility or wisdom to it. There remains a certain naivety in believing that everything can be solved with an increasing amount of data, as if the answers to human dramas depended solely on information management rather than on deep reflection about who we are and what we truly want

to achieve for the future generations to come. It is evident that we cannot evade technoscientific progress, and it is desirable that we think responsibly about how to accompany these processes. It would be irresponsible to fall into a determinism that suggests we should simply ride the wave without reflection, as if nothing depended on us other than accepting a future already programmed by uncontrollable forces.

The future is shaped by our present decisions, and it is commendable that political actors are thinking ahead in a responsible manner while listening to experts from various disciplines. The governance of technology will increasingly become an unavoidable issue on the political agenda. The abduction of truth through its reduction to mere data transforms AI into a sacred power, a reliable source for judging reality.

“Digital technology stands as an authority capable of determining reality more reliably than ourselves, as well as revealing dimensions hidden from our consciousness” (Sadin, 2020). Machines are anthropomorphized as if they possess the best discernment, leaving us with nothing to do but obey and relieve ourselves of the burden of thinking. We save time and mental effort while surrendering our freedom without resistance and accepting this new truth without question.

While we can work collaboratively and leverage the possibilities of technology, the greatest challenge lies in thoughtfully considering what we are willing to renounce of our human condition for convenience and what our non-negotiable minimums are.

Human thought is more than calculation and problem-solving. It clarifies and illuminates the world, bringing forth an entirely different reality. The intelligence

of machines poses the primary danger that human thought may begin to resemble it and become mechanical itself (Han, 2021).

Cyber Leviathan and Technocratic Power

In his work *Ciberleviatán* (2019), José María Lasalle presents the crossroads facing humanity: the choice between losing freedom for greater security or, through responsible political action, establishing a genuine pact that ensures citizens’ freedom, protects data, and sets new digital rights.

We find ourselves submerged in a swarm of humans “lacking critical capacity and devoted to consuming technological applications within an overwhelming flow of information that grows exponentially” (Lasalle, 2019).

According to this Spanish philosopher, humanist liberalism primarily aims to limit power, and it now confronts the seductive allure of technological power that seeks to be omnipresent and omniscient, without resistance. We are witnessing a new reconfiguration of power:

Today, the data generated by the internet and the mathematical algorithms that discriminate and organize it for our consumption form a binary of control and domination that technology imposes on humanity. To the extent that humans are acquiring the characteristics of digitally assisted beings, largely due to their inability to decide for themselves (Lasalle, 2019).

The fascination with the unlimited power of technology, viewed as inevitable and unavoidable, which promises greater control and certainty in decision-making, gradually erodes trust in the fragility and

spontaneity of the human factor. Thus, the freedom that is so valued and defended begins to be seen as a problem for progress, leading humans to accept that their freedom should be assisted by a superior, almost divine intelligence: artificial intelligence. Some authors are beginning to see in this technocratic sociocultural shift a promise to protect humans from their dangerous spontaneity, suggesting that it might be better to program ourselves according to what is deemed best by utilitarian criteria.

We are losing freedoms under the illusion that we can access new developmental possibilities, as if becoming ostensibly more free requires us to renounce fundamental liberties. Plus, we are doing this passively and with a certain naturalness and fascination.

Thus, we encounter a convergence of the technical, economic, and political realms, where power becomes disproportionately centralized over a growing number of activities, including health, education, and labor. According to Lasalle, “algorithmic despotism is returning humanity to a new minority status that unravels the liberal tradition of knowledge that fostered the Enlightenment.” ♦

Bibliography for further reading

- CARR, N. (2014). *Atrapados. Cómo las máquinas se apoderan de nuestras vidas*. Barcelona: Taurus.
- COECKELBERGH, M. (2021). *Ética de la inteligencia artificial*. Madrid: Cátedra.
- COECKELBERGH, M. (2023). *La filosofía política de la inteligencia artificial*. Madrid: Cátedra.
- CRAWFORD, K. (2022). *Atlas de la inteligencia artificial. Poder, política y costos planetarios*. Buenos Aires: FCE.
- HAN, B.-CH. (2021). *No-cosas: quiebres del mundo de hoy*. Buenos Aires: Penguin Random House.

LASALLE, J. M. (2019). *Cyberleviatán: El colapso de la democracia liberal frente a la revolución digital*. Madrid: Arpa.

SADIN, E. (2020). *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*. Buenos Aires: Caja Negra.

SADIN, E. (2022). *La era del individuo tirano. El fin del mundo común*. Buenos Aires: Caja Negra.

SADIN, E. (2024). *La vida espectral. Pensar la era del metaverso y las inteligencias artificiales generativas*. Buenos Aires: Caja Negra.

SAVULESCU, J., & LARA, F. (2021). *Más que humanos. Biotecnología, inteligencia artificial y ética de la mejora*. Madrid: Tecnos.

SIGMAN, M., & BILINKIS, S. (2023). *Artificial. La nueva inteligencia y el contorno de lo humano*. Barcelona: Penguin Random House.

VARELA, L. (2022). *Espejos: filosofía y nuevas tecnologías*. Barcelona: Herder.



Miguel Pastorino

PhD in Philosophy. Master's in Bioethics. Master's in Communication Management. Professor in the Department of Humanities and Communi-

cation at the Universidad Católica del Uruguay and Director of its Centro Core.

X: @MiguelPastorino

IG: @pastorinomiguel

LI: miguelpastorino



Personalising fakes: towards the disinformation apocalypse?

CHRISTOPH NEHRING

This article explores the effects and implications of AI-generated disinformation. It examines its forms, in particular deepfakes, and its impact on recent and upcoming elections. It also offers practical ideas for identifying and combating AI-driven disinformation, with particular attention to the role of influencers, journalists and other media professionals, and the unique challenges they face.

— —

Introduction

Artificial intelligence (AI) is rapidly transforming the global information landscape, creating new opportunities and unprecedented risks. Generating, spreading and boosting dis- and misinformation are prominent examples. Despite widespread fear and confusion, empirical knowledge on AI disinformation, its forms, impact and effects remain scarce, which in turn contributes to uncertainty, fear, mistrust and the demand for balanced, high-quality information.

Disinformation, deepfakes and manipulation

As early as 2023 disinformation experts spoke about the potential of generative AI as a “weapon of mass deception”, boosting and supercharging dis- and misinformation. Even though such doomsday scenarios have not yet materialized, AI possesses several qualities that have a strong effect on the production and distribution of dis- and misinformation. AI can make disinformation:

— *faster* (both in generating content as

well as automatically distributing such content)

— *cheaper* (e.g. by automating production and distribution, reducing human and financial resources needed)

— *more persuasive* (e.g. by using super real deepfakes)

— *more customized* (e.g. by using AI software for data analysis, identifying more effective messages and challenges to reach certain target audiences)

— *more far reaching* (e.g. by using AI bots and automation for distribution of disinformation or simply because AI tools are available for all normal social media users).

Experiments conducted by hackers and journalists have shown, for example, that the costs of using ChatGPT to power a fully automated fake news web site run entirely and only by ChatGPT dropped from 400 US\$ in 2023 to 105 US\$ in 2024.

Forms of AI disinformation

The *generative artificial intelligence* (genAI) revolution affects software designed to create every kind of content, text, images, videos and audio. Thus, known forms of AI disinformation include:

a. *Fake news web sites*: Even though difficult to detect, several thousand such sites whose contents (text, images, videos) are entirely created by ChatGPT or other chatbots have already been identified. Some, such as “electionwatch” or “TheDCWeekly” focus on organized disinformation about US politics and the 2024 US Presidential election, while others are commercial web sites simply rewriting and republishing old news for profit.

b. *AI-Images*: AI created images have started to flood social media platforms, messenger services and web portals. Some show either persons (most often politicians) in situations that never happens (e.g. Donald Trump dancing with underage girls) or depict events that never happened (e.g. a terrorist attack on the US Pentagon). While professional fake news web sites and outlets most likely backed by state actors also use AI images to publish along with fake articles, the vast majority of such AI generated images, however, are created and spread by “normal” social media and forum users. Such images are especially wide spread during the conflict in the Middle East, mostly emphasizing war damage and victims in Gaza. In some instances, such AI images that were spread on social media found their way to large stock databases (e.g. Adobe Stock), where they were sold for commercial use. Ukrainian users, on the other hand, are increasingly using AI generated images to portray support for the Ukrainian Army in their struggle against the Russian war of aggression. This trend demonstrates the effects of the “democratization” of genAI tools and their misuse.

c. *Deepfakes*: So-called “deepfakes” (deriving from “deep learning” and “fakes”) are AI produced or manipulated video and audio content. There are various different types of deepfakes, differing in either their

application (e.g. face swapping for deep porn or fraud and scamming calls) or the intention behind them. Deepfakes that are produced for the purpose of political disinformation have appeared in many different contexts, e.g. the Russian war against Ukraine and particularly during election campaigns all around the world (see below). Most often, they are used to produce fake discrediting evidence for scandalous statements or positions, participation in illegal or otherwise discrediting events or pornography. Their victims are most often publicly exposed person, e.g. celebrities, politicians, CEOs, influencers and journalists. Deepfakes have various qualities that have led to a high level of public fear and confusion: a) The impressive quality of such fakes; b) Their ability to convince and persuade audiences; c) The lack of reliable detection software and methods and d) the insecurity and inability of audiences to recognize and deal with deepfakes. During the following parts of this essay, its focus will be on deepfakes as one of the most imminent and dangerous examples of AI disinformation.

Experts and state investigators have found empirical proof for the existence of all these forms of AI disinformation. Yet, due to the so-called “detection challenge” of AI-content, it remains difficult to correctly assess and determine the quantity and actual amount of AI disinformation. Up to date, there is no 100% accurate detection method for AI generated content, no automatic upload filters or take down service etc. This means, while we can observe how the quality and the quantity of AI generated disinformation is rapidly increasing, its true quantity remains difficult to assess.

Deepfakes and elections in 2023 and 2024

AI disinformation and particularly deepfakes have become a weapon used to influence political campaigns and elections during the past two years. In most instances, deepfake technology was used to produce video or audio content featuring politicians, candidates, but also journalists and other popular voices in negative, discrediting scenarios. Some are reputational attacks on individuals trying to undermine their credibility, image and public reputation, others are part of negative political campaigns, trying to discredit political opinions, decisions or events. All of them, however, try to influence voter behaviour by deliberately spreading artificially created false, untrue or decontextualized information.

In other cases, deepfakes are used for official political campaigning. Such deepfakes differ in as much as they are a) attributable to an “official” source (e.g. a candidate, party, institution or organization), b) (often) labelled as AI generated content and c) not necessarily contain false information. During the elections for the European Parliament in June 2024, several far-right and right-wing parties (e.g. in France and Italy) used deepfake-technology to advance their messages and narratives via memes, images or AI-generated songs. In Pakistan, former PM Imran Khan and his team used deepfake technology to make him appear in campaign videos despite being imprisoned; in *India*, *Indonesia* and *the Philippines*, parties and campaigning teams created deepfakes of dead politicians or popular public figures for election campaigning. During the presidential elections in *Argentina*, both candidates and their teams heavily used all forms of generative AI (images,

» AI disinformation and particularly deepfakes have become a weapon used to influence political campaigns and elections. «

videos, text) for their campaigns. This also included malicious deepfake videos of both candidates that crossed the line between campaigning and disinformation by deliberately spreading aggressive lies. In *Mexico*, then presidential candidate and former mayor of Mexico City, Claudia Sheinbaum, featured in a deepfake video allegedly promoting a crude financial scheme, thus undermining her political credibility. Every country and election in 2024 saw political deepfakes that were meant to discredit political candidates and/or promote certain (mostly aggressive) narratives. By far the most use of AI-fakes was apparent during the US Presidential Election: Both sides (official party channels as well as supporters) published AI-generated images to get their messages across; however, more dangerous forms of deepfakes, e.g. AI-generated “robocalls” using the voice of President Joe Biden calling upon voters not to participate, AI-fakes depicting Taylor Swift cheering for Donald Trump or AI-generated non-existent content allegedly featuring in JD Vance’s book, spread on social media.

While deepfakes were part of every election in 2024, all empirical evidence points towards the fact that — unlike apocalyptic expectations — they did not

have a significant impact on the outcome of elections. So far, only in two cases did deepfakes that occurred during the last 48 hours before election day show a decisive influence on the election: In Slovakia, an audio-deepfake of one of the candidates allegedly discussing how to buy minority votes on the phone, seems to have had a direct effect on the outcomes, even though it did not “swing” the final results in favour of another candidate. During presidential election in Turkey, on the other hand, a deep-porn video of one of the candidates led to the withdrawal of said candidate from the elections. This obviously influenced the outcome of the election, yet as all polls saw the running president clearly winning the election anyway, this deepfake may have influenced the result, but not “swing” the election.

Journalism and Influencer: the Global Information Space

Generative AI has the potential to completely change the global information space. This includes all forms of political communication, content creation and presentation (including journalism and influencer).

AI and Journalism

GenAI has a strong effect on content creation as well as content presentation when it comes to journalism. Yet, there is an apparent “AI-gap”: Whereas traditional, quality media struggles to come up with concrete answers, boundaries and regulation concerning the ethical use of AI in journalism, low-quality media, tabloids, state-sponsored propaganda

outlets and fraudsters are already using AI for their purposes. The Russian foreign propaganda outlet “RT”, for example, already uses “deepfake personas”, i.e. non-existent, completely AI-generated automated avatars (which they call “digital presenters”) for their Spanish language programme. Several Chinese and state other news channels have been known for quite some time to use AI for these purposes as well.

And while traditional quality media all over the world mostly refrain from using genAI for creation of “core news”, i.e. from



» Russia is considered to be one of the most active actors of Foreign Information Manipulation and Interference. «

creating information itself, other actors are less prone to do so. News organisations have identified thousands of websites that rely on AI (most often ChatGPT) to run fully automated “news websites”. These websites either republish and rewrite old content for

advertising revenues or to spread outright political disinformation. Channel 1, a new news station established in Los Angeles in 2024, on the other hand, is the first reported news outlet that claims to be a serious media actor, but runs its programmes entirely with genAI, i.e. both for content creation as well as content presentation. Another important issue of genAI in journalism is the question of social media platforms regulate, label and publish AI generated content. While it will be soon become mandatory for platforms to label and specify AI content in North America and the EU, there are no such

unified norms for other parts of the world. Most social platforms themselves claim in their community standards and terms of use that AI-content and AI-profiles need to be clearly marked and registered. Yet, just like in the past, the extent to which these rules are being enforced varies significantly.

AI and Influencers

In the world of influencers, similar developments seem to be taking place and content generation and presentation are heavily affected by genAI. Virtual influencers, i.e. entirely AI-created and AI-driven avatars that pose as influencers on social media platforms have attracted millions of followers already (i.e. in China, Brazil, the USA or India). This is also true concerning the spread of mis- and disinformation, conspiracy theories etc. Throughout the world, influencers are gaining more and more significance as a target group and tool for professional disinformation actors, but also as professional creators and spreaders of disinformation.

Some TikTok-Influencers, for example, have turned AI-created videos about ever new conspiracy theories into their business model and are discussing in closed chat rooms about how to utilize genAI to increase revenues. In other cases, Russian embassies all over Africa have been proven to pay local influencers to spread disinformation. AI-influencer, on the other hand, have so far not been caught engaging in political campaigning and disinformation, yet they obviously bear a high risk potential in that regard.

AI manipulations and deepfakes also affect media professionals and influencers in other ways: Both groups (journalists and influencers) are regularly the victims of deepfake discreditation attacks using AI. Deepfake videos depicting journalists

promoting financial fraud schemes or dubious products without their consent have already become a regular occurrence in the US and all over Europe. Influencers, on the other hand, often face the risk of falling victim to deepfakes that target their reputation (and thus their business model). The most common scenario here is images





of female influencers being used for deep-porn. Such attacks may also be part of targeted reputation attacks for political purposes, as e.g. deepfakes of Taylor Swift following her mobilization during the US election in 2024 demonstrate.

» Artificial intelligence is fundamentally altering the landscape of disinformation, election interference, and information manipulation.«

“Over-Pollution” or: Drowning in a Sea of AI-Content

Yet another dimension of genAI in the global information space is the very real possibility of its “over pollution” with AI-content, automated AI-bots etc. Pessimistic scenarios propose the possibility of 90% of all online content being generated by AI in 2026, whereas automated online behaviour (e.g. bots and programmes) according to some studies already make up for the majority of all online activities. If genAI makes up for the majority of online content, presentation and activities, this will seriously affect political and other news, information and societies. Hence, “over pollution” might be one of the most serious long-term risks of genAI.

AI in Russian Foreign Information Manipulation and Interference

Russia is considered to be one of the most active actors of “Foreign Information Manipulation and Interference – FIMI”. The coordinated and covert spread of false, misleading and manipulated information to influence societies, events and elections is one important tool for these activities. Russian disinformation targets nearly every election on the planet and uses a large variety of complex tools and instruments. Russian embassies and consulates, Russian media, PR companies, paid journalists and influencers, anonymous web portals and local proxies are the most important actors of Russian disinformation. Its tactics span from simple propaganda information, paying influencers and journalists to complex information operations that include fakeing traditional quality media outlets and publishing covert disinformation. Narratives and messages of Russian disinformation usually center

❖ **AI-driven strategies can be used to micro-target voters, tailor messages, and enhance campaign efficiency** ❖

around certain key topics (e.g. anti-West, anti-Ukraine, anti-LGBTQ) which are reworked into customized messages for local audiences all around the world. In the Global South, for instance, such narrative usually focus on discrediting the Global West (e.g. Colonialism, social tensions, economic and social injustice etc.).

Use of AI

In its disinformation operations Russian FIMI-actors have shown their willingness to exploit the full and unrestrained potential of genAI for their purposes. The Spanish language programme of the Russian foreign propaganda broadcaster “RT” now includes two “digital presenters”, i.e. AI-avatars; in the US, several fake news websites that published fully automated negative articles about the US presidential election written with genAI were traced back to Russia; during an elaborate global disinformation campaign called “Doppelganger”, which centers around faked websites of the world’s most famous traditional media, Russian actors were caught using ChatGPT to generate and translate social media posts and comments; and in the ongoing war against Ukraine, Russian actors have time and again deployed deepfake videos (e.g. a fake video of President Zelenskyy calling for surrender or a fake video of Ukrainian intelligence chiefs allegedly admitting

their hand in an Islamist terror attack in Moscow) for foreign and domestic disinformation.

Conclusion

Artificial intelligence is fundamentally altering the landscape of disinformation, election interference, and information manipulation. Already today, all of these malicious activities have seen the integration of AI and no election now occurs anymore without some level of AI-generated disinformation. AI enables the production and dissemination of such content at unprecedented speeds, lower costs, and with increasing ease, making disinformation campaigns not only more accessible but also more automated, customizable, persuasive and large scale.

Despite these advancements, the feared “information apocalypse” has yet to materialize. No single election has been decisively influenced by AI-driven disinformation, even though there have been notable cases where deepfakes played a role. For example, in recent elections in Turkey and Slovakia, while deepfakes gained attention and raised concerns, they did not ultimately “swing the results” altogether in favor of any candidate or party.

Meanwhile, AI is not only a tool for disinformation but also a growing force in political campaigning. AI-driven strategies can be used to micro-target voters, tailor messages, and enhance campaign efficiency. As this trend grows, so too does the range of risks associated with AI manipulations, particularly deepfakes. Beyond elections, deepfakes contribute to cybermobbing, fraud and scams, and cybersecurity breaches, with influencers especially vulnerable to falling victim to such malicious uses of AI technology (e.g. cyberbullying with deepfake pornography). ♦

Bibliography

- BONTCHEVA, K. (ed.) (2024). [Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities.](#)
- FERRARA, E. (2024). [GenAI Against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models.](#)
- GEHRINGER, F. A., NEHRING, CH., & LABUZ, M. (2024, May 10). [The influence of Deep Fakes on Elections: Legitimate Concern or Mere Alarmism?](#) KAS Monitor 2024.
- HABGOOD-COOTE, J. (2023). [Deepfakes and the epistemic apocalypse.](#) *Synthese*, v. 201 103/2023.
- LABUZ, M. & NEHRING, CH. (2024, April 26). [On the way to deep fake democracy? Deep fakes in election campaigns in 2023.](#) *Eur Polit Sci.*
- MARCHAL, N., & XU, R. (2024, August 2). [Mapping the misuse of generative AI.](#) *GoogleDeepmind.*
- MUÑOZ, M. (2024). The AI Election Year: How to Counter the Impact of Artificial Intelligence. DGAP Memo, v. 1.
- SCHICK, N. (2020). *Deep Fakes and the Infocalypse.* Ottawa.



Christoph Nehring

Researcher, analyst and journalist. Guest lecturer and analyst at the media programme of the Konrad Adenauer Foundation, author for *Tagesspiegel*, *Deutsche Welle*, *NZZ*, *Spiegel* and many others. Passionate about AI and disinformation. He has been researching disinformation, manipulation and secret services for more than ten years.
LI: christopher-nehring-423b06257

DIÁLOGO POLÍTICO is a platform for democratic dialogue among political opinionmakers about topics of relevance in Latin America, based on the values of freedom, solidarity, and justice. It connects the region with the great geostrategic debates in the world. It is a window for publicizing projects of the Konrad Adenauer Foundation in Latin America.

DIÁLOGO POLÍTICO is part of the Regional Political parties and Democracy in Latin America Program (kas Partidos). Its objective is to reduce political polarization through constructive and informed pluralist debate oriented to the common good, to strengthen the political center from its Social-Christian, Liberal and Conservative roots.

6 **Foreword**
Artificial Democracy

58 **The fight against**
corruption from a new
technological paradigm
Denisse Rodríguez-Olivari

8 **Artificial Intelligence and**
elections: premature threats?
Sarah Kreps

66 **Artificial Intelligence and**
Public Safety in Latin
American Democracies
Gustavo Flores Macías y
Bárbara Hernández

20 **Governance of algorithms**
Daniel Innerarity

76 **Lars Zimmermann:**
“Technology won’t relieve us
of democracy’s tasks”
Interview

28 **Anticipatory governance**
with the help of technology
Lydia Garrido Luzardo

84 **Freedom and Truth Captured**
by Artificial Intelligence
Miguel Pastorino

38 **Is it possible to regulate AI?**
Global experiences
Ximena Docarmo

94 **Personalising fakes:**
towards the disinformation
apocalypse?
Christoph Nehring

48 **From mass networks**
to personalised voting
Jesús Delgado Valery