JOSÉ OCTAVIO ISLAS-CARMONA
Universidad Autónoma del Carmen

FERNANDO IGNACIO GUTIÉRREZ-CORTÉS
Tecnológico de Monterrey

AMAIA ARRIBAS-URRUTIA
Universidad de Valladolid

Disinformation and political propaganda: An exploration of the risks of artificial intelligence

ABSTRACT

A significant shift is currently underway in the disinformation industry. We are transitioning from the era of disinformation fuelled by fake news and social media to disinformation on a larger scale generated through artificial intelligence (AI). Therefore, the objective of this text is to analyse this disinformation phenomenon, catalysed by social media and AI, from the media ecology perspective. This work is divided into two parts. In the first part of the text, we analyse the disinformation phenomenon, highlighting the involvement of certain governments. In the second part of the text, we focus on recognizing the effects that can arise from the use of AI within the extensive landscape of the disinformation industry.

KEYWORDS

internet
social media
government
misinformation
fake news
new media

INTRODUCTION

Media ecology can be understood as a complex metadiscipline dedicated to studying how, throughout history, technologies and media have transformed the ecologies and cultural environments of humans and societies. Its extensive repertoire of analysis spans from the invention of the phonetic alphabet (Logan 2004) to artificial intelligence (AI). Throughout history, we can observe how both positive and negative effects emerge from all media and technologies. As Postman (1998) warned, the negative repercussions of new media and technologies can sometimes significantly outweigh the benefits they introduce to societies. This is something we can even observe with AI.

As Lance Strate (2017) suggests, media ecology can be understood as a unique approach to studying environments as media. It can also be viewed as an approach to understanding the human condition. From this perspective, we can say that AI is a medium that creates an entire environment that, like any other medium, influences our perceptions, understandings, feelings and values.

In the current context, a significant shift is emerging in the disinformation industry, moving from the era of fake news and social media towards a phase where AI plays a crucial role. This phenomenon poses complex challenges for society and global security as it reflects the evolution of disinformation tactics. The transition to AI involves increased sophistication in creating and disseminating false content, as the technology can dynamically adapt, personalize messages and influence public perception more subtly and efficiently.

The ability of AI to manipulate public opinion and undermine trust in traditional sources of information increases the complexity of facing these challenges. Furthermore, the rapid evolution of these technologies underscores the urgent need for strategies and regulations that effectively address the intersection between AI-driven disinformation, cybersecurity and the preservation of democracy. To confront this shift, it is essential for society, government institutions (which, however, actively participate in the disinformation phenomenon by instrumentalizing fake news and manipulating social media, significantly contributing to information chaos) and technology companies to collaborate closely in developing preventive and corrective measures. Public awareness of information manipulation, education in media literacy and the promotion of ethics in AI development are crucial elements in building a resilient and misinformation-resistant digital environment.

AI, much like any other technological marvel, makes its debut as a captivating and mesmerizing toy, enchanting everyone. In the initial stages, only a select few show a genuine interest in delving into the essence of this technology, and only a handful possess the ability to comprehend how this groundbreaking innovation could reshape our environment. Often, people engage with it without fully contemplating its potential repercussions (Levinson 1977).

THE USE OF THE INTERNET AND SOCIAL MEDIA IN THE DISINFORMATION INDUSTRY

Many authors identify the internet and social media as the main culprits in the proliferation of fake news. Furthermore, it is often asserted that disinformation is a characteristic phenomenon of contemporary societies.

Fake news (Lazer et al. 2017) spread faster than true news due to their remarkable ability to evoke intense emotions in people and their rapid dissemination through social media. Aral et al. (2018) attribute this phenomenon to the influence exerted by some digital leaders on their followers, thanks to their charisma and persuasive abilities. Alandete (2020) argues that the internet and social media create a conducive environment for misinformation due to political polarization, audience fragmentation and poor digital literacy in societies. Amoros (2018) maintains that digital platforms facilitate the spread of false information due to their global nature, massive reach and ease of use.

Many of our conversations on social media could be influenced by AI, and users would hardly know it because autonomous programmes (better known as bots) impacting our interactions are becoming increasingly challenging to detect. Currently, instead of sending automatic messages that platforms could delete, bots are reprogrammed to amplify and disseminate messages generated by humans in the digital environment (Nonnecke et al. 2019).

However, misinformation is not a recent phenomenon. The palaeontology of information (Serrano 2007) allows us to understand that fake news and misinformation overflow media and cultural imaginaries. The use of lies, Serrano (2007: 216) points out, is not a privilege reserved for humans. In the animal kingdom, information implicit in states and signals is crucial in the dynamics of survival, evolution and species development. Chemical energies, for example, generate signals that can be perceived by the organs of smell and taste. Materials produced by organic catabolism, mainly sweat, urine and faeces, are sources of chemical signals, which, beyond our will, turn us into 'informants'.

In states of fertility, and beyond their will, females often secrete certain substances derived from their sex hormones, which, in that particular state – fertility – their organism generates. These signals can be perceived by males of their own species – the fertilizers – but also, of course, by other animals, some of them predators. Fertility represents a state of vulnerability, and not only for the female. Females of other species can mimic the fertility signals of vulnerable females to attract males from various species to surprise and devour them. In the animal kingdom, deception and trap are part of an extensive repertoire of survival mechanisms.

SOME GOVERNMENTS AS CENTRAL ACTORS IN THE DISINFORMATION INDUSTRY

Fake news has driven the development of a robust disinformation industry of global dimensions, which annually generates estimated profits in the billions of dollars.

In the lucrative disinformation industry, major players participate, including some governments, political parties, certain technology giants, corporations of all sizes and sectors, religious groups, extremist sects of various tendencies, individual actors and even cells of organized crime.

Some governments, particularly authoritarian ones, often employ the following procedures to diminish or neutralize the voices of their critics: website blocking, disinformation campaigns, technical attacks, arrests for opinions expressed online, enactment of new repressive laws, content removal and surveillance of the citizenry.

WEB BLOCKS

Deibert et al. (2008) found that some governments employ various techniques to control internet access, including network blocks. Bradshaw and Howard (2017) analysed the spread of false information online. Network blocks implemented by some governments aim to curb the dissemination of false information; however, they can also have a negative impact on freedom of expression. Aziz and Leung (2018) argue that network blocks are a form of censorship used to suppress freedom of expression and political dissent. Hillebrecht (2018) asserts that particularly authoritarian governments use network blocks to stifle freedom of expression and political dissent.

In 2022, for example, digital repression intensified in Iran. The death of Jina Mahsa Amini on 16 September 2022 sparked a significant number of protests against the government of Ebrahim Raisi. Mahsa Amini, a young Kurdistani woman, was 22 years old. Two days before her death, she was arrested by the Moral Police of the Iranian government. Her arrest was due to not covering her hair with the hijab, as required by Iranian law. Jina was tortured, fell into a coma, and died. Her death generated widespread public outrage and protests. The Iranian regime decided to intermittently restrict internet access and blocked WhatsApp and Instagram to prevent communication with the outside world. Additionally, two people were executed for alleged blasphemy after sharing their opinions on religious topics on Telegram.

DISINFORMATION CAMPAIGNS

Ansari et al. (2022) has focused on analysing the use of bots during elections. For example, in the referendum held in the United Kingdom on its membership in the European Union, bots were used to spread false information to stimulate votes in favour of the United Kingdom's exit from the European Union. Social media are effectively used to spread fake news (Soroush et al. 2018), and governments, political parties and other actors often use new social media to manipulate public opinion. The Senate Intelligence Committee (2019) have studied disinformation campaigns carried out by the government of Russia. The Senate Intelligence Committee concluded that the federal government, civil society, and the private sector, including social media and technology companies, each have a crucial role in deterring and defending against foreign influence operations targeting the United States. Gehlbach and Lorenz (2020) analysed the use of disinformation by the Russian government. The authors conclude that disinformation is a key tool for the control exercised by Putin's government over the population.

In 2023, Freedom House highlighted that, in Thailand, governed by King Maha Vajiralongkorn, the tenth monarch of the Chakri dynasty, specialized army personnel are dedicated to manipulating the narrative related to the monarchy on social media.

TECHNICAL ATTACKS

Allison and Geers (2013) studied cyberattacks and the development of a silent and devastating global war on the internet, where the players are not always recognizable. The mentioned researchers conclude that cyberattacks will become more frequent. The purpose is to affect the fundamental infrastructure of enemy nations, and groups of professional hackers, in the service of certain governments, participate in the development of operations such as the

NotPetya attack on Ukraine in 2017, which caused estimated damages in the billions of dollars; the Stuxnet attack on Iran in 2010, which damaged Iran's nuclear programme; the Flame attack on Syria in 2012, which used malware to spy on Syrian government officials; the WannaCry attack on 200,000 computers in 2017, which demanded a ransom for decryption; the SolarWinds attack on companies and government organizations in 2020, which allowed attackers to steal sensitive information, and, more recently, cyberattacks between Russia and Ukraine and between Israel and Hamas, which demonstrate that governments have the capability to use the internet and social media networks to carry out technical attacks against their adversaries, whether they be other governments or opponents. In 2022, in Kazakhstan, for example, during the presidential elections held in November, the government of President Kasim-Jomart Tokayev, who successfully secured re-election, conducted a series of cyberattacks on independent media outlets.

ARRESTS FOR ONLINE OPINIONS

Regarding reports on internet freedom by Freedom House, Cuba is a frequent reference due to arrests of citizens expressing online criticism against President Miguel Díaz-Canel's government. In Cuba, any manifestation of citizen dissent faces severe punishment.

According to Article 19, the Cuban government aims to limit the freedom of expression, assembly and demonstration. From January to June 2023, there were 41 attacks against journalists (thirteen arbitrary detentions, house arrests, eight internet service suspensions, seven threats, five harassment cases, three official summons, three fines, one exile). The Cuban government has criminalized independent journalistic work, as reported by Article 19. Influencers with a significant following who question the Cuban government, such as YouTuber Hilda Núñez Díaz, are forced into exile.

For the ninth consecutive year, Freedom House ranked China as the nation with the least internet freedom. In April 2023, the civic activist and blogger, Xu Zhiyong, was sentenced to fourteen years in prison. Xu Zhiyong and Ding Jiaxi were dedicated to promoting the citizen organization 'New Citizens Movement'.

In Myanmar (formerly Burma), internet freedom has practically dissolved. In 2021, General Min Aung Hlaing led a military coup in the Asian country. Aung Hlaing established a Civil Administration Council, of which he is the head. Despite challenging conditions, some citizens have expressed their support for the democratic resistance movement through the internet. However, to counter dissidents, the military and its informants infiltrated groups on Telegram. This allowed authorities to identify dissidents, arrest them and, in some cases, even make them disappear. In July 2022, the military executed the activist and writer, Kyaw Min Yu, known as Ko Jimmy.

ENACTMENT OF NEW REPRESSIVE LAWS

To control public opinion on the internet, authoritarian governments often craft repressive laws to justify the arrest and imprisonment of citizens, journalists and opponents. Bradshaw and Howard (2023) warn of a wave of new repressive laws being enacted worldwide to limit freedom on the internet and on social media.

In 2023, in Belarus, a court under the government of Alexander Lukashenko sentenced journalists, Maryna Zolatava and Liudmila Chekina, editor-in-chief

and director of TUT.by, respectively, to more than twelve years in prison. Both journalists were jailed in May 2021. TUT.by is the largest independent news outlet in Belarus, celebrating its 46th year in 2023. It is estimated that 35 journalists are currently imprisoned in Belarus.

Daniel Ortega, absolute ruler of Nicaragua since 10 January 2007, has recently forced some opponents to choose between remaining imprisoned and being sent into exile, also losing their citizenship. Catholic bishop, Rolando José Álvarez Lagos, whose citizenship was revoked for refusing to leave Nicaragua, received a 26-year prison sentence.

DELETE CONTENT

Some governments proceed to remove content on the internet and/or on social media if they deem that what has been posted online goes against their interests. Sometimes, in addition to deleting content, the authors of the posts have faced sanctions or imprisonment.

Ndlovu (2021) has analysed content removal on the internet and social media as a form of censorship in Africa. Bertoni and Collin (2017) has examined the situation in Latin America, where certain governments have imposed strict surveillance on content generated by journalists, citizens and opposition politicians posted on the internet.

In Turkey, in 2023, authorities ordered the removal of content from certain news articles that criticized President Recep Tayyip Erdoğan's administration. Additionally, strict restrictions were imposed on Meta and X regarding the posting of comments related to the elections.

SPYING ON THE CITIZENRY

Cyber espionage is also carried out by governments in the most developed nations. This has been expressed by Edward Snowden (2019), a former agent of the National Security Agency, who exposed US intelligence services for engaging in espionage practices targeted at their own citizens.

Also, the current President of Mexico, Andrés Manuel López Obrador, has expressed severe criticisms of his predecessors in the presidency for the use of the Israeli software Pegasus, developed by the NSO Company. During the government of President Enrique Peña Nieto (2012–18), Pegasus was used to spy on 15,000 people. However, in April 2023, an investigation by *New York Times* (Kitroeff and Bergman 2023) revealed that President López Obrador's own government had become the main user of Pegasus. The newspaper *Reforma* (Grupo Reforma 2023) highlighted that the Mexican army is the one that most uses the Pegasus software to spy on citizens.

THE CONTRIBUTION OF SOCIO-DIGITAL NETWORKS TO THE DISINFORMATION INDUSTRY

According to the Digital 2023 Global Overview Report, internet users were estimated at 5.16 billion, of which 4.76 billion people are users of digital platforms and socio-digital networks. Out of the total internet users, users of digital platforms and socio-digital networks represent 92.3%. The global internet penetration was established at 64.4%, and socio-digital network penetration at 59.4%. On average, socio-digital network users spend 2 hours and 31 minutes per day, and in a month, they use an average of 7.2 digital platforms and socio-digital networks.

The videocracy and teledirected societies that overwhelmed Sartori (1998) have faded away. Socio-digital networks have taken up the baton. Studies conducted by the Reuters Institute (2022) timely warned how digital platforms and socio-digital networks were beginning to extend their influence in societies, wresting from conventional news media – press, radio and television – the power to determine agenda setting (McCombs and Shaw 1972).

Effectively, the algorithms of platforms and social media are designed to stimulate user engagement (Kaiser 2019; Frankel and Kang 2021; Haugen 2023). Therefore, platforms prioritize controversial content. This type of content is more likely to generate intense emotional reactions from users, such as anger or indignation, which foster greater participation, longer time spent and increased involvement. All of this favours the spread of fake news and hate speech, which can be used to manipulate public opinion.

Del Vicario et al. (2016) investigated the phenomenon of 'echo chambers' on Facebook. The authors concluded that people tend to interact with others who share their opinions, which can lead to the spread of false information. Soroush et al. (2018) analysed the spread of fake news on Twitter. The authors concluded that fake news spreads faster than true news, and people who share fake news are more likely to be influenced by it. Allcott and Gentzkow (2017) analysed the impact of fake news on the US presidential elections in 2016. The authors concluded that fake news had a significant impact on the elections, contributing to increased support for Donald Trump. Benkler et al. (2018) analysed the use of socio-digital networks to spread fake news. The authors concluded that governments, political parties and other actors can use virtual networks to manipulate public opinion, promote their own interests or destabilize their adversaries. Howard et al. (2017) analysed the spread of fake news during the 2016 US presidential elections. The authors concluded that fake news spread massively through socio-digital networks and had a significant impact on the outcome of the presidential elections. The devastating effects against Hillary Clinton, the Democratic Party and the elections in the United States were marked by socio-digital networks through disinformation campaigns and fake news, overshadowing the secondary effects of actions taken by Russian hackers against the Democratic Party servers and Hillary Clinton's computers. Kaiser (2019), from the depths of Cambridge Analytica, revealed delicate details of the use of Big Data, micro-targeting techniques and algorithms deployed to persuade the American electorate and bring Donald Trump to the presidency of the United States.

Global Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation report allowed the identification of 81 governments that deploy their 'cyber troops' on social media platforms to disseminate propaganda and political disinformation. The term 'cyber troops' (Bradshaw et al. 2020) refers to government, military or political party teams dedicated to manipulating public opinion on social media primarily and represents a global phenomenon. Troll armies focus on inhibiting political activism and press freedom. Their actions have violated human rights, degraded the quality of political news and undermined the legitimacy of democratically elected governments. According to Bradshaw and Howard, the tasks commonly performed by 'cyber troops' (2017: 9) include commenting on social media posts, individual targeting, sponsorship of government-backed accounts, websites or applications, fake accounts and computational propaganda and content creation.

The goal of cyber troops is to impose narratives favourable to a specific political actor, minimizing and neutralizing criticisms or potential

questioning. Social media companies have taken some measures to combat the misuse of digital platforms by cyber troops, including the removal of thousands of fake accounts and pages. However, political ads and propaganda, in general, continue to circulate widely on social media. In recent years, private companies with extensive computer expertise have also proliferated, offering their services to develop campaigns aimed at manipulating online audiences.

THE CONTRIBUTIONS OF AI TO THE DISINFORMATION INDUSTRY AND POLITICAL PROPAGANDA

In disinformation campaigns, AI is already being used to disseminate false content. Government operators can employ generative AI to generate disinformation at scale. Generative AI is a technology that can be used to create fake texts, images and videos – deep fakes – which are practically indistinguishable from reality. Governments can use generative AI to create fake news or propaganda that they can use to manipulate public opinion.

Amit et al. (2019) conducted one of the first academic studies on deep fakes, which are fake videos or audios created by manipulating real images or sounds. The authors argue that deep fakes pose a threat to society, as they can be used to spread false information or damage the reputation of individuals.

One of the agencies specialized in AI-supported deep fake production is Synthesia. The CEO of this London-based company is Victor Riparbelli. The mentioned firm offers 150 'digital humans' for hire (Wired 2023). Synthesia's services were contracted in Venezuela by supporters of President Nicolás Maduro's government to create House of News in Spanish, a non-existent news channel. From this channel, an avatar named Darren, who acted as the programme's host, claimed that the information circulating in the international press about widespread poverty in Venezuela is exaggerated.

Authors such as Malik et al. (2022) have focused on reviewing existing methods for detecting deep fakes. The authors discuss the different types of deep fakes, as well as the challenges and opportunities they pose for detection.

In the United States, AI-manipulated content has been used to defame electoral opponents. Accounts affiliated with the campaigns of former President Donald Trump and Florida Governor Ron DeSantis, both seeking the Republican Party nomination for the 2024 presidential elections, shared videos in 2023 with AI-generated content to attack each other. In these smear campaigns, the teams of both candidates placed modified images alongside genuine photographs.

Deep fakes have been recurrently used in the United States. In February 2023, a manipulated video circulated showing President Biden allegedly making transphobic comments. The video quickly went viral on social media. However, Joe Biden did not make the alleged transphobic statements. It was a deep fake.

In April 2023, an AI-generated video circulated, presenting a catastrophic vision of what could happen if President Biden were to be re-elected in the upcoming presidential elections in the United States: China would invade Taiwan, 80,000 migrants would enter US territory in a single day, the imminent closure of 500 regional banks would occur and so on.

In May, a video circulated on Facebook showing President Biden placing an 'I Voted' sticker on his granddaughter's chest and giving her a kiss on the cheek. In the manipulated version of this video, the President of the United

States appears to be behaving inappropriately with the child. Additionally, a caption was added that labelled him as a paedophile.

Women involved in politics have also been targets of deep fakes. On 22 May 2019, a modified video was posted on the internet that portrayed Nancy Pelosi, then Speaker of the House in the United States, in an apparent state of drunkenness (Frenkel and Kang 2021). Donald Trump, then President of the United States, and Rudy Giuliani, former Mayor of New York, shared the mentioned video on Twitter, pretending to be surprised by Pelosi's supposed condition. However, the Speaker of the House does not consume a drop of alcohol. It was a deep fake. The video was removed from several digital platforms, but Mark Zuckerberg refused to take it down from Facebook.

The video against Pelosi marked the beginning of deep fakes that would be deployed against women involved in politics. Most non-consensual deep fakes tend to feature sexualized images of women in the public sphere. From the industry of blackmail and character defamation, dirty politics has evolved into the production of pornographic scenes, the majority of which are deep fakes.

AI has also been used to manipulate audio. In Nigeria, which scored 60 points in the Freedom House assessment, during the elections in February 2023, an AI-manipulated clip circulated on social media implicating an opposition presidential candidate in alleged operations to manipulate voting results.

Some governments have developed close relationships with certain networks of disinformation professionals who rent their services to the highest bidder to spread false and misleading content. Agencies dedicated to manipulating public opinion generate substantial income from such 'services' and, of course, avoid leaving traces of the conditions of their hiring or the 'fees' they receive for their 'work'. In some authoritarian governments, such tasks are carried out by the military.

In the large-scale image war, Russian agencies have deployed extensive disinformation campaigns about Russia's invasion of Ukraine. For example, the operation known as 'Doppelgänger' was denounced by the French government, which accused Russia of conducting a major disinformation campaign in which hostile news about Ukraine was published by alleged French media, which, of course, did not publish such information. Cyber Front Z, another Russian network, has used Telegram to harass critics of President Vladimir Putin and is engaged in promoting propaganda against Ukraine.

Several major disinformation companies operate in Israel. In 2023, an investigation by Forbidden Stories, *The Guardian* and *Haaretz* pointed out that the firm 'Team Jorge', whose CEO is Tal Hanan, has been involved in influencing dozens of elections worldwide. Team Jorge uses an online platform to automatically generate texts based on certain keywords. The resulting texts are immediately replicated across an extensive chain of fake accounts on social media (Andrzejeweski 2023). The firm has not only offered its services to the political class but has also been hired by certain companies interested in manipulating public opinion in their favour.

Another Israeli firm, Mind Force, was linked by Meta to a network of active accounts in Angola dedicated to supporting President João Manuel Gonçalves Lourenço's government (Freedom House 2023). The accounts operated by Mind Force harass critics and opponents of the President of Angola.

The six social impact lines of AI are due process, discrimination, association and assembly, access to information, freedom of speech and privacy. We will now discuss some of the main implications.

Due process

Due process is a fundamental legal principle that ensures individuals the right to a fair trial. AI-enabled surveillance tools can be used to collect a large amount of data about individuals, including their movements, communications and online activities. These data can be used to create profiles of citizens, which could be used to predict their behaviour or make decisions about them. This could lead to a significant number of injustices, as people would be judged based on their profiles rather than their actions.

Discrimination

Algorithmic systems are trained on data that reflect the biases that unfortunately exist in the real world. For example, if the training data for a facial recognition system are biased in favour of White men, the system will also be biased in favour of White men.

Algorithmic systems can perpetuate bias by classifying the population based on prejudices. This is because algorithmic systems can use features that are correlated with race, colour, religion or sexual orientation to classify people. For example, a loan system may consider the neighbourhood where a person lives as a determining factor in assessing their credit risk. This can lead to discrimination against people living in popular neighbourhoods, which are often inhabited by minorities.

Biases can have very negative repercussions in hiring systems, education and healthcare. For example, a hiring system may use a person's educational history as a feature to determine their eligibility for a job. This can lead to discrimination against individuals from disadvantaged backgrounds, who are often minorities. A facial recognition system may be more likely to misidentify people of colour as suspects. A loan system may be more likely to reject loan applications from Black individuals. An education system may be more likely to recommend lower-level courses to women. However, all of the above are consequences of existing biases.

Association and assembly

AI systems with capabilities such as facial recognition can pose a serious threat to the rights of association and assembly. For example, an authoritarian government can use facial recognition systems to identify protesters demonstrating against its actions, decisions or policies. Once the protesters have been identified, state forces could arrest them or retaliate against them.

The Chinese government has used facial recognition systems to identify and arrest pro-democracy protesters. In Hong Kong, the government has used facial recognition systems to suppress protests.

The use of facial recognition systems through AI extends beyond the repressive imagination of governments. Companies could also use facial recognition systems to identify employees who, for example, participate in a gathering or join a union.

Access to information

The recent report from the Reuters Institute (2022) warns of a concerning phenomenon. People have been losing interest in the news. If people lose interest in the news, they will also lose interest in public affairs, and ultimately, they will lose interest in democracy.

The loss of interest in news events may be considered one of the collateral damages arising from the degradation of political communication. The majority of the population, as noted by the Reuters Institute, believes that the media is subject to undue political influence. Only a small group of people believes in the possibility that most news outlets prioritize the well-being of society over their own commercial interests. This is more concerning among young people, who are not interested in the news. If they prefer to stay informed about what is happening in the world, they prefer to do so through social media.

Freedom of speech

Automated systems can be used to censor political, social and religious speech. This is possible by detecting keywords or phrases that may be considered 'sensitive'. For example, an automated system can be used to block online comments containing insults or threats.

Additionally, AI-enabled surveillance tools can be directed to suppress dissent by identifying individuals expressing critical opinions of the government or established social order. These individuals may face harassment, persecution or even imprisonment. This can lead to a more repressive society, where people are afraid to express themselves freely.

AI surveillance can also promote self-censorship. This is because people may be afraid to express their opinions if they believe they are being monitored. For example, an individual may avoid discussing politics in public if they think the government is watching them. These concepts can apply to a variety of environments, such as social media, conventional media and political protests. A government may use automated systems to censor online comments questioning its actions.

Privacy

AI-enabled surveillance tools can impact people's privacy. Privacy is a fundamental right that ensures individuals have the right to control their own data. However, AI-enabled surveillance tools can be used to collect data about individuals without their knowledge or consent. This can lead to a sense of constant surveillance, which can have a negative impact on people's mental health and quality of life.

CONCLUSION

Social media play a catalysing role in the spread of information, serving not only as dissemination channels but also as spaces where false information rapidly amplifies, reaching massive audiences in real time. The AI's ability to generate persuasive content and adapt to individual preferences amplifies the threat, with its capacity to analyse behavioural patterns and preferences, taking disinformation to highly sophisticated levels, increasing its effectiveness and complicating its detection.

Media literacy is crucial to educate society in identifying misinformation, promoting critical awareness and fostering discernment skills to counteract the harmful effects of AI-driven disinformation. Ethics plays a crucial role in building responsible AI systems and preventing their exploitation for disinformative purposes. The implementation of constant monitoring and evaluation of tactics used in AI-based disinformation is necessary to respond

agilely and proactively, involving both security experts and authorities. The need for developing legal frameworks for AI use, as recently approved by the European Union, to address misinformation and preserve individuals' privacy is demonstrated.

This new paradigm of disinformation poses a significant challenge to cybersecurity because AI introduces additional vulnerabilities, adapting dynamically to bypass traditional detection measures. Therefore, global strategies of a transnational nature with a cooperative approach among countries and international organizations are needed. However, as demonstrated in the text, government interests have prevailed over citizens' rights.

Ethical technological innovation plays a crucial role in addressing the challenge of AI-driven disinformation. This approach involves the development and implementation of technologies that are not only advanced in their capabilities but also ethically responsible. This implies designing algorithms and systems that promote transparency, fairness and accountability, thus minimizing the risk of information manipulation for disinformative purposes. Moreover, ethical technological innovation is also related to the creation of effective tools for detecting and mitigating disinformation, helping to preserve the integrity of information on the internet.

Empowering society is equally essential to confront this challenge. Fostering public awareness of disinformation tactics, providing media literacy education and offering tools that enable individuals to discern between truthful and manipulated information are fundamental elements for building a society resilient to disinformation. By equipping the population with the necessary skills and knowledge, a robust defence line is created against digital manipulation, thereby promoting informed and active participation.

The pernicious use of AI in the disinformation industry and its negative effects on culture and civilization are a cause for great concern. In one of the early definitions of media ecology, Postman (1970) emphasized the responsibility of this complex metadiscipline in shaping the imagination of human survival. Postman's proposal takes on particular significance in our tumultuous days.

REFERENCES

Alandete, D. (2020), *Fake News: La Nueva Arma de la Destrucción Masiva*, Barcelona: Ediciones Deusto.

Allcott, H. and Gentzkow, M. (2017), 'Social media and fake news in the 2016 election', *Journal of Economic Perspectives*, 31:2, pp. 211–36, https://doi.org/10.1257/jep.31.2.211.

Allison, G. T. and Geers, K. (2013), *The Future of Warfare: Debating the Shape of Things to Come*, Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School.

Amit Roy, C., Hany, F. and Philip, H. (2019), 'Deepfakes: A looming challenge for society', *Journal of Cybersecurity*, 5:1, pp. 1–12.

Amoros, M. (2018), *Fake News: La verdad de las noticias falsas,* Barcelona: Plataforma Editorial.

Andrzejeweski, C. (2023), 'Team Jorge: In the heart of a global disinformation machine', Forbidden Stories, 15 February, https://forbiddenstories.org/story-killers/team-jorge-disinformation/. Accessed 16 October 2023.

Ansari, J. A. N., Azhar, M. and Akhtar, M. J. (2022), 'The spread of misinformation on social media: An insightful countermeasure to restrict', *Studies in Economics and Business Relations*, 3:1, pp. 1–14, https://sabapub.com/index.php/sebr/article/view/401. Accessed 27 November 2023.

Aral, S., Muchnik, L. and Sundararajan, A. (2018), 'The spread of mistruths online', *Science*, 363:6425, pp. 363–66.

Article 19 (2023), 'Libertad de expresión en Cuba. Análisis semestral', Article 19, 24 August, https://articulo19.org/article-19-analiza-la-libertad-de-expresion-en-cuba-durante-el-primer-semestre-de-2023/. Accessed 17 October 2023.

Aziz, S. and Leung, L. (2018), *Internet Shutdowns: A Global Overview*, Cambridge, MA: Berkman Klein Center for Internet & Society.

Benkler, Y., Faris, R., Roberts, H. and Etzioni, O. (2018), *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford: Oxford University Press.

Bertoni, E. and Collin K. (2017), 'Surveillance and privacy protection in Latin America: Examples, principles, and suggestions', in H. F. Cate and J. X. Dempsey (eds), *Bulk Collection: Systematic Government Access to Private-Sector Data*, New York: Oxford Academic, https://doi.org/10.1093/oso/9780190685515.003.0016.

Bradshaw, S. and Howard, P. N. (2017), 'The global disinformation order: Mapping the spread of misinformation and the battle for control of the online world', Computational Propaganda Research Project, Oxford Internet Institute, 1 September, https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2019/09/CyberTroop-Report19.pdf. Accessed 16 October 2023.

Bradshaw, S. and Howard, P. N. (2023), *The Global Disinformation Order: 2023 Report*, Oxford: Computational Propaganda Research Project, Oxford Internet Institute.

Bradshaw, S., Bailey, H. and Howard, P. (2020), 'Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation', Oxford Internet Institute, University of Oxford, 13 January, https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/. Accessed 16 October 2023.

Deibert, R., Palfrey, J., Rohozinski, R. and Zittrain, J. (2008), *Access Denied: The Practice and Policy of State Control over the Internet*, Cambridge, MA: MIT Press.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G. and Quattrociocchi, W. (2016), 'Echo chambers: Emotional contagion and group polarization on Facebook', *Scientific Report*, 6:37825, pp. 1–12, https://doi.org/10.1038/srep37825.

Freedom House (2023), 'Freedom the net 2023: The repressive power of artificial intelligence', Freedom House, 21 November, https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence. Accessed 17 October 2023.

Frenkel, S. and Kang, C. (2021), *Manipulados: La batalla de Facebook por la dominación mundial*, Mexico: Debate.

Gehlbach, S. and Lorenz, J. (2020), *The Disinformation State: How Russia Uses Propaganda to Control Its Population*, Oxford: Oxford University Press.

Grupo Reforma (2023), 'Ejército mexicano es el que más espía con Pegasus – NYT', Reforma, 18 April, https://www.reforma.com/

ejercito-mexicano-es-el-que-mas-espia-con-pegasus-nyt/ar2589169. Accessed 16 October 2023.

Haugen, F. (2023), *La verdad sobre Facebook,* Barcelona: Deusto.

Hillebrecht, C. (2018), *Digital Censorship in Authoritarian Regimes,* New York: Routledge.

Howard, P. N., Kollanyi, B., Bradshaw, S. and Neudert, L. M. (2017), 'Social media, news and political information during the US election: Was polarizing content concentrated in swing states?', Computational Propaganda Research Project, Data Memo, 28 September, https://arxiv.org/pdf/1802.03573. Accessed 16 October 2023.

Kaiser, B. (2019), *La dictadura de los datos*, New York: HarperCollins.

Kitroeff, N. and Bergman, R. (2023), 'Cómo México se convirtió en el mayor usuario del programa de espionaje más conocido del mundo', *New York Times*, 18 April, https://www.nytimes.com/es/2023/04/18/espanol/pegasus-mexico-gobierno-ejercito.html. Accessed 16 October 2023.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F. and Watts, D. J. (2017), 'The science of fake news', *Science*, 359:6380, pp. 1094–96, https://doi.org/10.1126/science.aao2998.

Levinson, P. (1977), 'Toy, mirror, and art: The metamorphosis of technological culture', *ETC: A Review of General Semantics*, 34:2, pp. 151–67.

Logan, R. (2004), *The Alphabet Effect: A Media Ecology Understanding of the Making of Western Civilization*, Cresskill, NJ: Hampton Press.

McCombs, M. E. and Shaw, D. L. (1972), 'The agenda-setting function of mass media', *Public Opinion Quarterly*, 36:2, pp. 176–87, http://www.jstor.org/stable/2747787. Accessed 29 November 2024.

Malik, A., Kuribayashi, M., Abdullahi, S. M. and Khan, A. N. (2022), 'DeepFake detection for human face images and videos: A survey', *IEEE Access*, 10, pp. 18757–775, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9712265. Accessed 29 November 2023.

Meaker, M. (2023), 'Synthesia, la empresa de IA que vende deep fakes', *Wired*, 9 August, https://es.wired.com/articulos/synthesia-empresa-de-ia-que-vende-deepfakes. Accessed 17 October 2023.

Ndlovu, M. (2023), 'Media censorship deepens in Africa, further aggravating the state of press freedom', African Institute for Development, https://afidep.org/media-censorship-deepens-in-africa-further-aggravating-the-state-of-press-freedom/. Accessed 27 November 2023.

Nonnecke, B., Martín-del-Campo, A., Singh, S., Wu, W. and Crittenden, C. (2019), 'Women's reproductive rights computational propaganda in the United States: The role of bots in spreading disinformation, harassment, and divisiveness on Twitter in 2018', CITRIS Policy Lab, CITRIS and the Banatao Institute, 1 May, https://bit.ly/3dwdQVq. Accessed 15 October 2023.

Postman, N. (1970), 'The reformed English curriculum', in A. C. Eurich (ed.), *High School 1980: The Shape of the Future in American Secondary Education*, New York: Pitman, pp. 160–68.

Postman, N. (1998), 'Five things we need to know about technological change', UC Davis, 28 March, https://web.cs.ucdavis.edu/~rogaway/classes/188/materials/postman.pdf. Accessed 15 October 2023.

Reuters Institute for the Study of Journalism (2022), 'Digital News Report 2022', Reuters Institute for the Study of Journalism, 15 June, https://

reutersinstitute.politics.ox.ac.uk/digital-news-report/2022. Accessed 16 October 2023.

Sartori, G. (1998), *Homo videns: La sociedad teledirigida,* Spain: Taurus.

Senate Intelligence Committee (2019), 'Report of the select committee on intelligence United States senate on Russian active measures campaigns and interference in the 2016 U.S. Election: Volume 2: Russia's Use of Social Media, with Additional Views', *116th Congress 1st Session Senate*, Report 116-XX, pp. 1–86, https://digitalcommons.unl.edu/senatedocs/4/. Accessed 28 November 2023.

Serrano, M. M. (2007), *Teoría de la comunicación: La comunicación, la vida, la sociedad*, Spain: McGraw Hill.

Snowden, E. (2019), *Vigilancia permanente,* Spain: Ed. Planeta.

Soroush, V., Deb, R. and Sinan, A. (2018), 'The spread of true and false news online', *Science*, 359:6380, pp. 1146–51, https://www.science.org/doi/10.1126/science.aap9559. Accessed 27 November 2023.

Strate, L. (2017), *Media Ecology: An Approach to Understanding the Human Condition*, New York: Peter Lang, https://doi.org/10.3726/978-1-4331-4005-1.

SUGGESTED CITATION

CONTRIBUTOR DETAILS

José Octavio Islas-Carmona, Ph.D. is a professor at Universidad Autónoma del Carmen, and a member of the National Researchers System in Mexico.

Contact: Educación a Distancia, C.56 No.4 Esq. Avenida Concordia Col. Benito Juárez C.P. 24180. Ciudad del Carmen, Campeche, México
E-mail: joseoctavio.islas@gmail.com

https://orcid.org/0000-0002-6562-3925

Fernando Ignacio Gutiérrez-Cortés, Ph.D., is the director of the Division of Humanities and Education at Tecnológico de Monterrey, State of Mexico campus, and a member of the National Researchers System, Mexico.

Contact: Tecnológico de Monterrey, State of Mexico Campus, KM 3.5 Carr. Lago de Guadalupe, Atizapán, Estado de México 52926, México.
E-mail: fgutierr@tec.mx

https://orcid.org/0000-0003-1753-3527

Amaia Arribas-Urrutia, is an associate professor at the University of Valladolid (Spain), has a Ph.D. in information sciences and is the president of the Ibero-American Forum on Communication Strategies (FISEC) in Spain.

José Octavio Islas-Carmona | Fernando Ignacio Gutiérrez-Cortés | Amaia Arribas-Urrutia

Contact: Department of Modern, Contemporary, American History, and Journalism. Plaza del campus s/n 47011, Valladolid, Spain.
E-mail: amaya.arribas@uva.es

 https://orcid.org/0000-0001-9452-8364