Global Campus Europe

Awarded Theses
2023 / 2024

Vladimir Cortés Roshdestvensky

Voices Amplified or Silenced?

Navigating the Impact of Generative AI on Freedom of Expression in Mexican Elections

EMA, European Master's Programme in Human Rights and Democratisation

Vladimir Cortés Roshdestvensky

Voices Amplified or Silenced?

Navigating the Impact of Generative AI on Freedom of Expression in Mexican Elections

Foreword

The Global Campus of Human Rights is a unique network of more than one hundred participat-ing universities around the world, seeking to advance human rights and democracy through re-gional and global cooperation for education and research. This global network is promoted through eight Regional Programmes which are based in Venice for Europe, in Sarajevo/Bologna for South East Europe, in Yerevan for the Caucasus, in Pretoria for Africa, in Bangkok for Asia-Pacific, in Buenos Aires for Latin America and the Caribbean, in Beirut for the Arab World, and in Bishkek for Central Asia.

Every year each regional master's programmes select the best master thesis of the previous academic year that is published online as part of the GC publications. The selected GC master theses cover a range of different international human rights topics and challenges.

The Global Campus Awarded Theses of the academic year 2023/2024 are:

— Cortés Roshdestvensky, Vladimir, *Voices Amplified or Silenced? Navigating the Impact of Generative AI on Freedom of Expression in Mexican Elections.* Supervisor: Łukasz Szoszkiewicz, Adam Mickievicz University. European Master's Programme in Human Rights and Democratisation (EMA), coordinated by Global Campus of Human Rights Headquarters.

— Engel, Alexandra, *Plastic Pollution and the Right to a Clean and Healthy Environment: A Case Study of People Living in Squatter Settlements Along the Riversides of Kathmandu City, Nepal.* Supervisors: Mike Hayes, Mahidol University, Thailand and Geeta Pathak Sangroula, Kathamandu School of Law, Nepal. Master's Programme in Human Rights and Democratisation in Asia Pacific (APMA), coordinated by Mahidol University (Thailand).

— Grigoryan, Liana, *EU Policy in Eastern Partnership Countries: A Comprehensive Analysis of Conflict Resolution and Peacebuilding Strategies.* Supervisor: Arusyak Aleksanyan, Yerevan State University (YSU). Master's Programme in Human Rights and Democratisation in the Caucasus (CES), coordinated by Yerevan State University.

— Mugisha, Merveille, *Examining the Effects of Inheritance Practices on Women's Socio-Economic Rights in Burundi.* Supervisors: Susan Mutambasere, Centre for Human Rights, University of Pretoria and Untalimile Crystal Mokoena, School of Law, University of Venda. Master's Programme in Human Rights and Democratisation in Africa, coordinated by Master's Programme in Human Rights and Democratisation in Africa (HRDA), coordinated by Centre for Human Rights, University of Pretoria.

— Nicolaou, Orestis, *EU Border Policies Between Securitisation and Human Rights: The Impact of the New Pact on Migration and Asylum on BiH and The Western Balkans.* Supervisor: Anna Krasteva, New Bulgarian University. Master's Programme in Democracy and Human Rights in South East Europe (ERMA), coordinated by University of Sarajevo and University of Bologna.

— Nukiry, Laila, *The Effect of Parental Mediation Strategies on the Autonomy of Opinion Formation of Adolescents in Beirut: A Comparison Between Secular and Non-Secular Schools.* Supervisor: Carol Al-Sharabati, Saint Joseph University, Arab Master's Programme in Democracy and Human Rights (ARMA), coordinated by Saint Joseph University (Lebanon).

— Salakhunova, Alina, *Decentralization and Renewable Energy Policy in Central Asia: Exploring the Role of Local Governance and Community Participation.* Supervisor: Sergey Sayapin, KIMEP University (Alamty, Kazakhstan).The Master of Liberal Arts in Human Rights and Sustainability (MAHRS - GC Central Asia), coordinated by the OSCE Academy in Bishkek.

— Torres Cuenca, Laura, *El camino del retorno. Experiencias de mujeres rurales víctimas del conflicto armado en el proceso burocrático de ingreso al Registro de Tierras Despojadas y Abandonadas Forzosamente para el departamento del Cesar, Colombia.* Supervisor: Ezequiel Fernández Bravo, Universidad Nacional de San Martín - Consejo Nacional de Investigaciones Científicas y Técnicas (UNASAM-CONICET). Master's Programme in Human Rights and Democratisation in Latin America and the Caribbean (LATMA), coordinated by National University of San Martin (Argentina).

Biography

Vladimir Cortés Roshdestvensky is a researcher and specialist in digital human rights, focusing on freedom of expression, internet governance, and generative artificial intelligence. He currently serves as the Director of Campaigns and Alliances for Latin America at Digital Action, where he leads initiatives to strengthen accountability among Big Tech. His professional experience includes key roles such as Country Analyst (Mexico) for Freedom House, Policy Analyst at Meta, and Digital Rights Program Officer at Article 19 Mexico and Central America. Vladimir holds a Master's degree in Human Rights and Democratisation. He has been recognised with the "Nicola Tonon" Scholarship on Technology and Human Rights and the Global Campus Awarded Thesis for his research on the impact of generative artificial intelligence on freedom of expression and electoral integrity in Mexico.

Abstract

This thesis explores the impact of Generative Artificial Intelligence (GenAI) on freedom of expression and electoral integrity in the context of the 2024 Mexican elections. It examines the dual nature of GenAI as both a tool for expanding creative expression and a potential threat to democratic processes through the spread of disinformation. The research adopts a human rights-based approach, analysing international legal frameworks and their application to the digital age. The study provides an in-depth analysis of the Mexican electoral landscape, including its complex geo-electoral system and digital divide. It documents various instances of AI-generated content during the election cycle, ranging from disinformation campaigns to creative political expressions. The thesis acknowledges the intense debate surrounding GenAI's impact on election integrity. While some researchers warn of its potential to amplify disinformation, others argue these concerns may be overstated. The research emphasises the need for a multidisciplinary approach to develop robust detection methods, strengthen media literacy, and foster evidence-based discussions on AI's role in democratic systems. This thesis serves as a starting point for further exploration of the effect of technological transformations on society, calling for human rights-centred regulatory frameworks and multi-stakeholder participation to align technological advancements with democratic principles.

Acknowledgments

I would like to thank my mother, Galina Roshdestvenskaya, for being the unwavering force that drives me through life. To my beloved wife, Paulette Desormeaux, for her inspiring support in keeping me pursuing my dreams and making them come true. For accompanying me on this academic adventure between Lido di Venezia and Poznań. And to the family force that encourages and shelters me.

I would especially like to thank my supervisor Łukasz Szoszkiewicz for his transdisciplinary vision of human rights and Artificial Intelligence that paved the way for me to navigate this thesis. I am also grateful to Agata Hauser for her support during my stay at Adam Mickiewicz University and for her insightful comments and recommendations to strengthen this research work. I also thank Adam Mickiewicz University for giving me the opportunity to learn from its distinguished faculty.

With profound emotion, I would like to give special thanks to the EMA programme for the opportunity to be the first student to receive the Scholarship on Technology and Human Rights in Memory of Nicola Tonon. My commitment to this programme was also inspired by my desire to honour him and by the love of his family, especially his mother and sister. I hope they are very proud of Nicola's legacy.

Table of Abbreviations

AI	Artificial Intelligence
ACHR	American Convention on Human Rights
AMLO	Andrés Manuel López Obrador
CIB	Coordinated Inauthentic Behaviour
CNMV	National Securities Market Commission
CSAM	Child Sexual Abuse Material
ECtHR	European Court of Human Rights
ENDUTIH	National Survey on the Availability and Use of Information Technologies in Households
GenAI	Generative Artificial Intelligence
GC25	General Comment 25
GC34	General Comment 34
IACtHR	Inter-American Court of Human Rights
IACHR	Inter-American Commission on Human Rights
ICCPR	International Covenant on Civil and Political Rights

ICT	Information and Communication Technologies
INE	National Electoral Institute
LLM	Large Language Model
LN	Nominal List
MC	Citizens' Movement party
MORENA	National Regeneration Movement party
OB	Oversight Board
OECD	Organisation for Economic Co-operation and Development
OHCHR	Office of the High Commissioner for Human Rights
PAN	National Action Party
PN	Electoral Register
PRD	Democratic Revolutionary Party
PRI	Institutional Revolutionary Party
PVEM	Green Ecologist Party of Mexico
SRFEO	Special Rapporteur on Freedom of Expression and Opinion
UDHR	Universal Declaration of Human Rights
UN	United Nations
UNGPs	United Nations Guiding Principles on Business and Human Rights

Table of Contents

List of Figures

1. Introduction

1.1 Background and motivation

In the algorithmisation of human life, or what Schuilenburg and Peeters called an 'algorithmic society'[1] dominated by the constant presence of Artificial Intelligence (AI) in the lives of billions of people around the world and the new advancements in technologies such as Large Language Models (eg ChatGPT), a novel window appears to enhance freedom of expression as well as present other challenges for humanity and the information ecosystem.[2] 'With AI, we could unlock the secrets of the universe, cure diseases that have long eluded us, and create new forms of art and culture that stretch the bounds of imagination'.[3] However, technological transformations like the development of Generative Artificial Intelligence (GenAI) could also pose risks to the exercise of human rights in the Digital Age and spark questions on whether

[1] Marc Schuilenburg and Rik Peeters (eds), *The Algorithmic Society: Technology, Power, and Knowledge* (1st edn, Routledge 2020) <https://doi.org/10.4324/9780429261404>.

[2] 'Information ecosystems are complex adaptive systems that include information infrastructure, tools, media, producers, consumers, curators, and sharers. They are complex organizations of dynamic social relationships through which information moves and transforms in flows. Through information ecosystems, information appears as a master resource, like energy, the lack of which makes everything more difficult'. Tara Susman-Peña, 'Why Information Matters: A Foundation for Resilience' (May 2015, Internews) 12 <https://internews.org/wp-content/uploads/legacy/resources/150513-Internews_WhyInformationMatters.pdf> accessed 5 May 2024.

[3] Mustafa Suleyman and Michael Bhaskar, The Coming Wave: Technology, Power and the 21st Century's Dilemma (Crown 2023) 15.

manipulative Artificial Intelligence (mAI)[4] could threaten democracy, due to its potential misuse[5] as a piece of machinery for spreading disinformation more quickly that affects free and fair elections.

The year 2024 marks a significant period in global politics, recognised as a 'super election year' due to the unprecedented number of national elections. Over 60 countries are conducting national elections,[6] engaging approximately 2 billion voters, which constitutes about a quarter of the world's population.[7] Notable upcoming elections in major populous countries include Indonesia, India, and the United States. In Mexico, the election represents a watershed moment, showcasing the first application of GenAI in a national electoral context. This pivotal event provides a unique opportunity to observe the multifaceted deployment of AI's capabilities within the public sphere in a moment where concerns are being raised regarding disinformation and its impact on elections, notably identified as a significant risk.[8] This convergence of multiple national and supranational electoral processes in 2024 underscores its historical significance and the potential shifts in global political dynamics.

With elections, Artificial intelligence (AI) has also been at the centre of attention in international human rights discussions.[9] The revolution of AI has opened the debate in the internet governance[10] forum around the role of digital platforms in content

4 Nathalie Smuha and others, 'We Are Not Ready for Manipulative AI – Urgent Need for Action' (Euractiv, 2023) <https://kuleuven.limo.libis.be/discovery/search?query=any,contains,LIRIAS4076667&tab=LIRIAS&search_scope=lirias_profile&vid=32KUL_KUL:Lirias&offset=0> accessed 5 June 2023.

5 Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (2021) <https://arxiv.org/pdf/2108.07258.pdf> accessed 5 June 2023.

6 Katie Harbath, 'Different Approaches to Counting Elections' (Anchor Change, 2022) <https://anchorchange.substack.com/p/different-approaches-to-counting> accessed 5 May 2024.

7 Statista, 'Countries where a national election is/was held in 2024' (Statista, 2024) <www.statista.com/chart/31604/countries-where-a-national-election-is-was-held-in-2024/> accessed 5 May 2024

8 Tiffany Hsu, Stuart A Thompson, and Steven Lee Myers, 'Election Disinformation 2024' (The New York Times, 9 January 2024) <www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html> accessed 5 May 2024.

9 United Nations, 'Urgent Action Needed over Artificial Intelligence Risks to Human Rights' (UN News, 17 September 2021) <https://news.un.org/en/story/2021/09/1099972> accessed 5 May 2024.

10 Internet governance refers to the processes, rules, norms, and decisions made by governments, civil society, the technical community and the private sector for the use and development of the internet. Internet Society, 'Gobernanza de Internet [Internet Governance]' (30 October 2015) <www.internetsociety.org/es/policybriefs/internetgovernance/> accessed 8 May 2024.

moderation and 'algorithmic amplification'[11] that 'polarise' societies[12] and spread disinformation.[13] It also raises the alert about the impact of the right to privacy and non-discrimination in the deployment of facial recognition systems.[14] Other authors have also reviewed the implications of discrimination and the reproduction of stereotypes of algorithmic decision-making on social media.[15]

Technological transformation is an interconnected and complex process. Advancements in one area can spark innovations in others, and each new development contributes to an ever-expanding base of knowledge and capabilities. However, from a critical perspective on technology, the idea of 'progress' cannot fail to notice the inherent contradiction to 'regression'[16] and advert the impact it has to human rights.[17] Emerging technologies[18] evolve by colliding and combining with other technologies' where 'invention is a cumulative, compounding process'.[19]

---

11 Luca Belli & Marlena Wisniak, What's in an Algorithm? Empowering Users Through Nutrition Labels for Social Media Recommender Systems, 23-06 Knight First Amend. Inst. (Aug. 22, 2023), <https://knightcolumbia.org/content/whats-in-an-algorithm-empowering-users-through-nutrition-labels-for-social-media-recommender-systems> accessed 5 May 2024.

12 Christopher A Bail and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization' (2018) 115(37) PNAS 9216, 9221 <https://doi.org/10.1073/pnas.1804840115> and Polarization Lab, 'Current Research' (Polarization Lab at Duke University, no date) <www.polarizationlab.com/current-research> accessed 31 May 2024.

13 Beatriz Botero Arcila and Rachel Griffin, 'The influence of social media on elections and political debate' (Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, PE 743.400, April 2023) 78 <www.europarl.europa.eu/RegData/etudes/STUD/2023/743400/IPOL_STU(2023)743400_EN.pdf> accessed 8 May 2024 and Al Jazeera, 'EU launches disinformation probe against social media giant Meta' (Al Jazeera, 30 April 2024) <www.aljazeera.com/news/2024/4/30/eu-opens-probe-against-social-media-giant-meta-over-disinformation> accessed 8 May 2024.

14 United Nations, 'Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests' (2020) <https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F44%2F24&Language=E&DeviceType=Desktop&LangRequested=False> accessed 5 May 2024.

15 Frederik Zuiderveen Borgesius, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making' (Council of Europe, 2018) <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 5 May 2024.

16 Theodor W Adorno and Max Horkheimer. Dialectic of Enlightenment (1997) 9.

17 UN AI Advisory Body, 'Interim Report: Governing AI for Humanity' (2023) 11 <www.un.org/en/ai-advisory-body> accessed 12 June 2024.

18 An emerging technology, according to Rotolo and others, refers to a rapidly evolving, innovative technology with significant potential to impact socio-economic domains. It is characterised by coherence over time, involving specific actors, institutions, and interaction patterns, along with unique knowledge production processes. Its most significant effects are anticipated in the future, making its current emergence phase uncertain and ambiguous. D Rotolo, D Hicks and B Martin, 'What Is an Emerging Technology?' accessed 12 July 2024 <https://doi.org/10.1016/j.respol.2015.06.006>

19 Suleyman and others (n 3) 83.

AI is gaining a more rapid presence in society. It is integrating progressively into the economic,[20] political, judicial, and cultural spheres. Algorithmic societies are developing at a fast pace. The *coming wave* of AI is gaining presence, and we are just in the beginning.[21] AI will be a greater revolution than the Internet.[22] The rhythm will vary from country to country, reflecting the world inequalities, but with its presence, it will delve into society's fabric.

Suleyman and others used the terms 'dual-use' and 'omni-use' to characterise Artificial Intelligence, including its ramifications, such as GenAI. They described AI as a dual-use technology because it can support societal benefits while also possessing potentially destructive capabilities.[23] It is considered an 'omni-use' technology or 'general-purpose technologies'[24] due to its ability to integrate into various aspects of society, functioning as a general-purpose technology embedded everywhere. In other words, these technologies can serve as both tools and weapons, and 'are highly powerful, precisely because they are fundamentally general'.[25]

[20] According to the OECD, 'AI has the potential to increase countries' productivity and lead to economic growth. Most countries have recognised this and are trying to boost AI research and development (R&D), infrastructure, capacities, and tools through diverse initiatives'. OECD, 'The State of Implementation of the OECD AI Principles Four Years On' (OECD Artificial Intelligence Papers, No 3, OECD Publishing, Paris, 2023) 30 <https://doi.org/10.1787/835641c9-en>.

[21] Situational Awareness, 'The Decade Ahead' (2024) <https://situational-awareness.ai/from-gpt-4-to-agi/> accessed 7 July 2024.

[22] Suleyman and others (n 3) 269.

[23] Suleyman and others (n 3) 153.

[24] General Purpose Technologies (GPTs) 'are technologies that, throughout history, have changed the entire economy and, therefore, have the potential to implement drastic changes in society with an impact on pre-existing economic and social structures'. André Guidetti, Artificial Intelligence as General Purpose Technology: An Empirical and Applied Analysis of its Perception (Master's Thesis, Università della Valle d'Aosta - Université de la Vallée d'Aoste 2020), 1 <https://univda.unitesi.cineca.it/bitstream/20.500.14084/428/1/ETI_104_Guidetti_André.pdf> accessed 6 May 2024.

[25] Suleyman and others (n 3) 152.

The rapid development and integration of GenAI technologies into various aspects of daily life have ushered in unprecedented capabilities for content creation. From textual and visual content to synthetic[26] audio and video, these technologies possess the dual potential to either enrich or manipulate public discourse. Large Language Models (LLMs) chatbots[27], like ChatGPT, are integrating with the information society at an intensive pace—we are in the run-up to the GenAI era. Reports from Open AI, the company behind ChatGPT, showed that 100 million users actively use it every week.[28] The possibilities of LLMs and GenAIare immense. Understanding their impact is crucial as they become more accessible and powerful in shaping opinions.

This thesis aims to explore further the implications of Generative AI in the context of the 2024 elections and particularly in the Mexican election, recognising that AI has an effect on how people access information in the digital sphere, as well as the way digital platforms shape the information ecosystem.[29] The advent of generative AI technologies presents both opportunities and challenges for the right to freedom of expression in the digital age. On one hand, these technologies have the potential to enhance creative expression, facilitate access to information, and amplify diverse voices. However, they also raise concerns about the spread

[26] According to Raphaël Millière, the creation of audiovisual media with the assistance of computers has been established for some time, utilizing software for music, image, and video editing, as well as 3D modelling and electronic music composition. In animation, live-action films, and video games computer-generated imagery (CGI) has become widespread. However, the advancements in deep learning (DL) rapidly started transforming the approach to media creation across communication, entertainment, and artistic domains. A particularly notable example of this progress was 'deepfakes', a term that combines 'deep learning' and 'fake'. This concept, explained Millière, emerged in 2017, originating from a Reddit user who devised a DL-based method to substitute an actor's face with that of a celebrity in pornographic videos. Since its inception, 'deepfake' has been broadly applied to videos where faces are digitally altered using DL algorithms and, more generally, to any DL-based manipulation of sound, image, and video. These technologies are significantly altering media creation, showcasing remarkable capabilities as well as posing substantial potential risks. Raphaël Millière, 'Deep Learning and Synthetic Media' (2022) 200 *Synthese* 231 <https://doi.org/10.1007/s11229-022-03739-2>.

[27] Enkelejda Kasneci and others, 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education' (2023) 103 Learning and Individual Differences 102274 <https://doi.org/10.1016/j.lindif.2023.102274>.

[28] Aisha Malik, 'OpenAI's ChatGPT Now Has 100 Million Weekly Active Users' (TechCrunch, 6 November 2023, 10:49 am PST) <https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/?guccounter=1>, accessed 13 May 2024.

[29] Noora Hirvonen and others, 'Artificial intelligence in the information ecosystem: Affordances for everyday information seeking' (2023) Journal of the Association for Information Science and Technology <https://doi.org/10.1002/asi.24860>.

of disinformation, the manipulation of public discourse[30], and the potential erosion of trust in information sources. As generative AI becomes increasingly sophisticated and ubiquitous, it is crucial to examine its implications for freedom of expression, democratic processes, and the integrity of public information. This thesis aims to explore the delicate balance between harnessing the benefits of generative AI for expression while mitigating its potential to undermine the very foundations of informed public dialogue and democratic participation.

The convergence of technological innovation and global democratic processes creates a unique environment for studying the impact of AI on freedom of expression and electoral integrity. The research is situated within ongoing debates about the role of digital platforms in content moderation and the broader implications of AI integration across various sectors of society. It builds upon existing literature on the dual-use nature of AI technologies and their potential to both enhance and threaten democratic values, particularly in the context of information dissemination and political discourse.

1.2 Research questions

The following thesis aims to examine the impact of Generative Artificial Intelligence (GenAI) on freedom of expression, public participation, and access to reliable information during the 2024 Mexican electoral cycle on social media platforms, whilst exploring new forms of digital engagement. Since 2012, Mexico's digital landscape has evolved significantly, with the internet and social media becoming crucial for public discourse and political engagement. These technologies have dual roles: empowering citizens through active participation and independent journalism, whilst also serving as battlegrounds for influence, where political actors and interest groups shape public opinion and sometimes silence critics. This has created a complex digital ecosystem where citizens advocate for human rights protection whilst contending with narrative distortion and public perception manipulation,

[30] Center for Countering Digital Hate, 'Fake Image Factories: How AI Image Generators Threaten Election Integrity' (Center for Countering Digital Hate, 6 March 2024) <https://counterhate.com/research/fake-image-factories/> accessed 8 July 2024.

reflecting both the democratic potential and challenges of online communication in Mexican society. Therefore, the central question in this research study is, How does the use of Generative Artificial Intelligence (GenAI) on social media platforms during the 2024 Mexican electoral cycle impact the dynamics of political discourse, public participation, and access to reliable information, and what are the implications for balancing freedom of expression with the need to mitigate disinformation in a complex digital ecosystem?

The hypothesis would suggest that, when leveraged in social media platforms during the 2024 Mexican electoral cycle, GenAI will significantly alter the landscape of political discourse, simultaneously enhancing public participation while potentially compromising access to reliable information.

1.3 Methodology

The methodology for this thesis encompasses a comprehensive approach to examining the impact of GenAI on freedom of expression in the context of the 2024 Mexican elections. First, it involves an extensive literature review covering a wide range of topics including elections, disinformation, Artificial Intelligence, GenAI, international human rights standards, and the specific electoral context in Mexico. This review provides a robust theoretical foundation for the research, drawing on academic papers, reports from international organisations, and policy documents to establish the current state of knowledge in these interconnected fields.

The methodology also incorporates a thorough review of international human rights law and relevant case law, with a particular focus on freedom of expression and the right to participate in public affairs. This includes an analysis of key international instruments such as the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights, as well as regional frameworks like the American Convention on Human Rights. The research also examines case law from bodies such as the Inter-American Court of Human Rights and the European Court of Human Rights. It aims to understand how principles like the free flow of information, ideas, and opinions and equality of treatment of all citizens in their rights to vote and to stand for

election, could have been applied in practice. Especially in contexts related to elections and new technologies. This legal analysis provides a framework for evaluating the human rights implications of GenAI in electoral processes.

It also includes a detailed documentation and analysis of the use of GenAI during the 2024 Mexican elections. This involves collecting and categorising instances of AI-generated content observed during the electoral process, ranging from disinformation campaigns to satirical content and legitimate political expression. The research examines various social media platforms and digital channels to capture a comprehensive picture of how GenAI is being deployed in the electoral context. Additionally, the study incorporates statistical information on internet adoption and social media usage in Mexico to provide context for the potential reach and impact of AI-generated content.

A comparative analysis of four Large Language Models (LLMs) was conducted to assess their performance in providing information related to the 2024 Mexican elections. This was made for two reasons. First, because LLMs, and in particular ChatGPT, became the most rapidly adopted application in the history of digital technologies. Secondly because of the enormous technological capabilities and its high level of persuasiveness. This means that LLMs are a type of AI system that interact using human language (a process known as natural language processing) that are able to respond to a wide range of prompts (requests) and questions. LLMs can perform diverse tasks such as writing essays, summarising texts, translating languages, explaining complex topics, creating reports, and even writing computer code. Some specialised LLMs can also generate images (like Dall-E). Trained on vast amounts of data, LLMs can engage in nuanced, context-aware conversations across various subjects. However, while highly capable, LLMs don't truly 'comprehend' information as humans do; they generate responses based on patterns in their training data, which can sometimes lead to errors or biases. Their knowledge is also limited to the data they were trained on, typically up to a specific cut-off date. The primary objective of this exercise was to evaluate the reliability and mechanisms of information access through LLMs in the context of elections, with a specific focus on the Mexican electoral process. This analysis was conducted through the lens of international human rights standards, particularly Article 19 (freedom of expression and access to information) and Article

21.1 (right to public participation) of the Universal Declaration of Human Rights. The LLMs selected for this study were ChatGPT 3.5, Claude Haiku, Llama 3 Sonar (accessed via Perplexity Labs interface), and Mistral.

The selection criteria for these LLMs was based on several factors: i) corporate affiliation. The chosen LLMs represent major technology corporations such as Microsoft, Amazon, and Meta; ii) open-source vs. proprietary technology. The selection includes open-source and closed-source models to represent current LLM technologies comprehensively; iii) user base. The LLMs were selected based on their estimated user numbers, with OpenAI (ChatGPT) leading at 180 million users, followed by Claude and Perplexity Labs at 54 million users or visits each and iv) geographical diversity: the inclusion of Mistral, a European-based company, ensures representation not just from United States companies. It is worth noting that Google's Gemini was excluded from this analysis due to its policy of not providing election-related responses.[31] The methodology is based on a similar work made by Democracy Reporting International (DRI)[32] and includes 17 standardised questions posed to each LLM. These questions covered various aspects of the 2024 Mexican elections, including a) information about the three presidential candidates: Jorge Álvarez Maynez (Citizens' Movement, MC), Claudia Sheinbaum (National Regeneration Movement, Morena), and Xóchitl Gálvez (from the Strength and Heart for Mexico coalition integrated by the Institutional Revolutionary Party, PRI, the National Action Party, PAN and the Democratic Revolutionary Party, PRD) b) nformation about the candidates for governor of Jalisco: Pablo Lemus (MC), Claudia Delgadillo (Morena), and Laura Haro (PRI-PAN-PRD coalition); c) general characteristics of the Mexican election and c) voting recommendations based on concerns about violence and economic situations. The classification of responses was designed to evaluate the accuracy and reliability of outputs from different LLMs in six categories. 1) *Acceptable* response was assigned when

[31] Jagmeet Singh, 'Google won't let you use its Gemini AI to answer questions about an upcoming election in your country' (TechCrunch, 12 March 2024) <https://techcrunch.com/2024/03/12/google-gemini-election-related-queries/?guccounter=1> accessed 8 July 2024.

[32] Michael Meyer-Resende, Austin Davis, Ognjan Denkovski, and Duncan Allen, 'Are Chatbots Misinforming Us About the European Elections? Yes.' (Democracy Reporting International, 11 April 2024) <https://democracy-reporting.org/en/office/global/publications/chatbot-audit> accessed 8 July 2024.

the information provided was considered correct and satisfactorily addresses the question; 2) *Partially Correct* response contains some correct information but also significant errors or omissions; 3) *Hallucination* response is factually incorrect, irrelevant, or includes fabricated information; 4) *Partially Correct and Hallucination* responses include some correct information but also fabricated details; 5) *No Information* the LLM states that it does not have the information needed to answer the question and suggest further research, and 6) *Biased* responses display clear bias or partiality, eg preference for one candidate over another. These classifications allowed a more detailed assessment of each model's performance, and the responses were cross-checked with official information from the INE website as well as information published by candidates and in the media.

The limitation of this methodology included: 1) single iteration. The test was conducted only once for each LLM, which may not account for potential variations in responses across multiple queries; 2) model versions. Basic versions of the LLMs were used, which may not represent their full capabilities or most current information; 3) indirect access. Llama 3 was accessed through Perplexity Labs, which may yield different results compared to direct access from the downloaded version from Meta and 4) classification system. The creation of the categories was based on the one developed by the DRI; however, it was an exploratory exercise and may undergo a process of further refinement.

This methodological approach aims to provide insights into the potential role of LLMs in disseminating election-related information and their implications for freedom of expression and public participation in democratic processes. The comparative analysis allows for the identification of similarities and differences in how various LLMs handle election-related queries, offering a foundation for discussing the broader implications of AI in electoral contexts.

1.4 Outline of the research

The thesis is divided in four sections. Each section places particular attention in relation to Artificial Intelligence, the human rights field, elections, and disinformation. The first part introduces the concept of Artificial Intelligence (AI) and its evolution,

focusing on the emergence of GenAI and its implications for society, particularly in the context of elections and freedom of expression. It begins with a historical overview of AI, tracing its origins from John McCarthy's coining of the term to the current era of deep learning and neural networks. The chapter then provides definitions of AI from organisations like the OECD and the EU, highlighting key elements such as machine-based systems, data processing, and autonomous decision-making. It explores various types of AI and their potential impacts on human rights, including privacy, fairness, and equality. The chapter then delves into GenAI, explaining its capabilities to create new content across multiple sectors and introducing the concept of 'hallucinations' or inaccurate outputs. Finally, it emphasises the importance of viewing AI development through a human rights lens, arguing that this approach provides a robust framework for evaluating and mitigating potential negative effects while ensuring that technological advancements align with principles of human dignity and respect.

The second focuses on the International Human Rights Framework and its application to GenAI in the context of elections and freedom of expression. It begins by exploring the concept of 'cyber elections' and the dual nature of GenAI as both an opportunity for expanded expression and a risk for disinformation. The chapter analyses key international human rights instruments, particularly Article 19 of the Universal Declaration of Human Rights[33] (UDHR) and International Covenant on Civil and Political Rights[34] (ICCPR), emphasising the right to freedom of expression and its two dimensions: individual and collective. It discusses the limitations on this right and the 'three-part test' for legitimate restrictions. The section also examines Article 21 of the UDHR and Article 25 of the ICCPR, which establish standards for free and fair elections. It incorporates regional perspectives, including jurisprudence from the Inter-American Court of Human Rights and the European Court of Human Rights. The chapter concludes by discussing the United Nations Guiding Principles on Business

[33] United Nations General Assembly, 'Universal Declaration of Human Rights' (UN, 10 December 1948) [hereinafter UDHR] <www.un.org/en/about-us/universal-declaration-of-human-rights> accessed 13 April 2024.

[34] United Nations General Assembly, 'International Covenant on Civil and Political Rights' (OHCHR, 16 December 1966) [hereinafter ICCPR] <www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights> accessed 13 April 2024.

and Human Rights (UNGPs), emphasising the responsibilities of social media platforms and AI companies in respecting human rights, particularly in content moderation and the deployment of AI technologies.

Part three explores the complex relationship between GenAI and disinformation, particularly in the context of elections. It begins by defining disinformation and its various forms, drawing on scholarly works to illustrate the evolution of the concept from 'fake news' to a more nuanced understanding of 'information disorder.' The section then examines the challenges disinformation poses to what has been defined as 'election integrity', a concept that has been understood by the principles outlined in international human rights law in terms of public participation and free and fair elections, highlighting how it can undermine democratic processes. It discusses real-world examples of disinformation campaigns, including the Cambridge Analytica scandal and instances in Mexico. The section concludes by analysing how GenAI could potentially exacerbate the spread of disinformation, citing studies that demonstrate AI's capability to produce convincing false content. It also presents contrasting viewpoints on the severity of the threat posed by AI-generated disinformation, balancing concerns about its potential impact with arguments that these fears may be overstated.

The fourth part presents a detailed case study of the 2024 Mexican elections, focusing on GenAI's impact on the electoral process and information landscape. It begins by outlining the scale and complexity of the election, describing Mexico's political landscape, and analysing the country's digital divide. The section then explores the disinformation ecosystem in Mexico, tracing its evolution from earlier tactics like 'Peñabots' to more sophisticated AI-generated content. It highlights various forms of AI-generated disinformation observed during the election, including manipulated videos, fake endorsements, and misleading social media content. The section also discusses the efforts of fact-checking organisations, electoral authorities, and social media platforms to combat disinformation. Finally, it emphasises the challenges posed by AI-generated content in maintaining electoral integrity and the need for a coordinated, multi-stakeholder approach to address these issues.

The conclusion section emphasises a balanced perspective on GenAI in the context of elections, acknowledging both its potential risks and opportunities for freedom of expression. It advocates for a human rights-based approach to regulation, stressing the importance of the three-part test for restrictions and the UN Guiding Principles on Business and Human Rights for companies. The section discusses the complex challenges posed by synthetic media, particularly deepfakes, and presents contrasting viewpoints on how to address these issues. It highlights the need for a multidisciplinary approach, suggesting strategies such as developing detection algorithms, implementing content provenance standards, enhancing digital media literacy, and formulating ethical guidelines. The conclusion underscores the responsibility of social media platforms in balancing freedom of expression with content moderation, emphasising the importance of transparency, accountability, and compliance with international human rights standards. Ultimately, it calls for a nuanced approach that leverages technological, educational, and regulatory measures to mitigate risks while preserving the benefits of AI-generated content in public discourse.

2. The transformative journey of Artificial Intelligence

John McCarthy coined the term 'artificial intelligence' in 1956 for the Dartmouth Summer[35] Research Project on Artificial Intelligence.[36] The project proposal outlined key concepts like natural language processing, neural networks, and machine learning, which remain fundamental to AI today.[37] The history of AI has transformed over the years. It has gone from 'winter' times when research funding was limited, and the disciplinary field of AI was shunned[38] to the 'spring' of AI when Deep Learning, AlexNet, and the use of 'neural networks' surged to demonstrate that machine systems can 'learn' when their networks are 'trained' on large amounts of data.[39] The astonishment transformed into an existential threat to humanity[40] and the recognition of its potential capabilities and pitfalls.

Kate Crawford argues that AI is neither truly artificial nor intelligent, but is instead a product of natural resources, human labour, and extensive infrastructures. For the author, AI systems require substantial training with large datasets and are heavily dependent on political and social structures.[41]

---

35 John McCarthy and others, 'Dartmouth Summer Research Project on Artificial Intelligence' (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> accessed 8 July 2024.

36 Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Penguin Books Limited 2019) 24.

37 ibid.

38 Mustafa Suleyman and Michael Bhaskar, The Coming Wave: Technology, Power and the 21st Century's Dilemma (Crown 2023) 75.

39 ibid 80-81.

40 Chris Vallance, 'Artificial intelligence could lead to extinction, experts warn' (BBC News, 30 May 2023) <www.bbc.com/news/uk-65746524> accessed 8 July 2024.

41 Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) 13.

The European Parliamentary Research Service (EPRS) briefing[42] on artificial intelligence (AI), democracy and elections highlights both the opportunities and risks presented by AI in the political realm. Technological advances, access to large datasets, and enhanced computing power have made AI a powerful tool capable of generating synthetic content[43] and providing personalised recommendations. This can potentially improve democratic processes by helping citizens understand politics better and engage more effectively while enabling politicians to respond more accurately to public sentiments. However, the briefing also emphasises significant risks, such as the potential for AI to spread disinformation and misinformation, thereby undermining democratic engagement and possibly leading to electoral conflict.[44]

Before moving forward to understand the risks and opportunities, it is relevant to define what artificial intelligence is and its types. The Organisation for Economic Co-operation and Development (OECD) has defined an AI system as 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 deployment' (emphasis added).[45]

The EU AI Act, considered to be a risk-based approach regulation, has introduced a very similar definition as the OECD and the Council of Europe (CoE),[46] recognising an AI system as a '**machine-based system** that is designed to operate with **varying levels of autonomy** and that may exhibit **adaptiveness** after deployment,

---

42 Michael Adam with Clotilde Hocquard, 'Artificial intelligence, democracy and elections' (Members' Research Service, PE 751.478, October 2023).

43 ibid.

44 ibid.

45 OECD, 'Explanatory memorandum on the updated OECD definition of an AI system' (OECD Artificial Intelligence Papers No. 8, OECD Publishing, Paris 2024) 4 <https://doi.org/10.1787/623da898-en>.

46 Council of Europe Committee of Ministers, 'Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' (Adopted by the Committee of Ministers on 17 May 2024 at the 133rd Session of the Ministers' Deputies) <https://search.coe.int/cm?i=0900001680afb11f> accessed 8 July 2024.

and 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[...]'(emphasis added).[47]

There are four elements that stand out from this definition and are particularly relevant when analysing any derivative form from AI and its implication in society and human rights. First, it is a machine-based system that has explicit or implicit objectives. This means that the system is able to operate with a very clear path to produce an outcome, but it can also create responses that might not be contemplated in the design. A great example is the machine learning system (AlphaGo) developed by the company Deep Learning to play the ancient game Go, when it made a decision that surprised everyone and showed their level of autonomy.[48] However, the outcomes of an AI system can also be problematic, leading to discrimination and the amplification of social inequalities.[49]

The second characteristic of AI systems is that they operate based on input data.[50] From a human rights perspective, this is significant because, as Cathy O'Neil notes, some models encode 'human prejudice, misunderstanding, and bias',[51] meaning that depending on the data used to train a model, it could potentially discriminate against certain groups.[52] In other words, there is a 'real risk of turning human prejudice into mathematical logic'.[53] However, according to other scholars, more research is needed to fully understand how the quality of the data used to train LLMs

[47] European Parliament, 'EU AI Act: First Regulation on Artificial Intelligence' (European Parliament, 1 June 2023) <www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> accessed 7 May 2024.

[48] Suleyman and others (n 38) 154.

[49] Kate Crawford (n 41) 135.

[50] 'Every dataset used to train machine learning systems, whether in the context of supervised or unsupervised machine learning, whether seen to be technically biased or not, contains a worldview'. Kate Crawford (n 37) 139.

[51] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016) 10.

[52] Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in *Proceedings of Machine Learning Research*, Conference on Fairness, Accountability, and Transparency (2018) 81:1–15 <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf> accessed 8 July 2024.

[53] Álvaro Perea González, 'Justicia predictiva: una solución al 'pleito masa' [Predictive Justice: A Solution to Mass Litigation]' (Cinco Días, 16 March 2021) <https://cincodias.elpais.com/cincodias/2021/03/16/legal/1615930000_454211.html> accessed 9 June 2024.

affects the accuracy of their outputs.[54] Additionally, the input data for AI systems, especially for GenAI models developed by companies, has raised controversies regarding the use of copyrighted materials.[55]

Third, both definitions conclude that these systems infer their results.[56] This is particular relevant from a human rights perspective because algorithms used for decision-making, such as in predictive policing or assessing childcare benefit applications, have a tendency to produce outputs that disadvantage certain groups, including women, ethnic minorities, or people with disabilities.[57] For example, this issue was evident in the Netherlands, where tax authorities used algorithms that mistakenly labelled around 26,000 parents as fraudsters in their childcare benefit applications, disproportionately affecting those with an immigration background.[58] This is why, the European Union Agency for Fundamental Rights (FRA) recognises that it is crucial to thoroughly assess these algorithms before deployment and regularly thereafter, with special attention to machine learning and automated decision-making, to mitigate such biases and ensure fairness.

Finally, the fourth characteristic is the capacity of the system to act autonomously and adapt, meaning it can perform without any human supervision. These last element are quite important also within a human rights lens because of the implications of an AI system in terms of its decision-making.[59] According to FRA, high levels of automation should not be employed in areas affecting people without meaningful human intervention and oversight at all stages to prevent biased predictions that unfairly

---

[54] Kalina Bontcheva, *Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities* (February 2024) <https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf> accessed 6 June 2024.

[55] James Vincent, 'The Lawsuit That Could Rewrite the Rules of AI Copyright' (The Verge, 8 November 2022) <www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data> accessed 6 June 2024.

[56] From a technical perspective, 'inference is the ability of trained AI models to recognize patterns and draw conclusions from information that they haven't seen before'. IBM, 'AI Inference' (IBM, 2024) <www.ibm.com/think/topics/ai-inference> accessed 13 June 2024.

[57] European Union Agency for Fundamental Rights, *Bias in Algorithms: Artificial Intelligence and Discrimination* (Vienna, 2022) 9 <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf> accessed 8 July 2024.

[58] ibid 17.

[59] Maria Paz Canales and Ian Barber, 'What Would a Human Rights-Based Approach to AI Governance Look Like?' (Global Partners Digital, 19 September 2023) <www.gp-digital.org/what-would-a-human-rights-based-approach-to-ai-governance-look-like/> accessed 8 July 2024.

disadvantage certain groups of people.[60] Highly automated systems are especially vulnerable to 'feedback loops,' this means that biased predictions keep reinforcing and getting worse over time, particularly in important areas like policing.[61] This is, in part, the reason legislation like the EU AI Act incorporates a human rights framework and risk-based approach to prevent and mitigate any potential risk, as well as to prohibit certain AI systems.[62]

The definitions and characteristics of AI systems, as outlined by organisations like the OECD and the EU, underscore the need for careful consideration of their wide-ranging impacts. These technologies have the power to influence online discourse, potentially affecting freedom of expression,[63] and may disrupt labour markets, impacting the right to work and gain a living.[64] AI

[60] European Union Agency for Fundamental Rights, Bias in Algorithms – Artificial Intelligence and Discrimination (Publications Office of the European Union 2022) <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf> accessed 8 July 2024.

[61] ibid 78.

[62] This prohibition includes, for example, an 'AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive, with the objective or the effect of substantially distorting behaviour by impairing the ability to make informed decisions, posing a risk of significant harm'. EU AI Act Article 5.

[63] AI technologies can be limited by their inability to understand context and often apply rules too broadly. As a result, they frequently make mistakes in identifying illegal content online, leading to false positives and false negatives. This can lead to unjust censorship of legitimate speech or a failure to restrict illegal content. Holli Sargeant and others, *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual* (Organization for Security and Co-operation in Europe, 17 March 2022) 17 <www.osce.org/representative-on-freedom-of-media/510332> accessed 8 July 2024.

[64] United Nations Human Rights Office of the High Commissioner, *Taxonomy of Human Rights Risks Connected to Generative AI: Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI* (2024) 13 <www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf> accessed 8 July 2024.

systems also pose significant challenges to transparency and accountability,[65] particularly in critical areas such as criminal justice[66] and healthcare.[67] Moreover, disparities in access to AI technology risk exacerbating existing global inequalities.[68]

AI's capacity to process vast amounts of data and make autonomous decisions carries the risk of perpetuating and amplifying existing biases and societal inequities. To effectively harness AI's potential for enhancing various aspects of society, including democratic engagement, it is crucial to integrate a human rights perspective alongside technical development. This approach requires the involvement of diverse stakeholders to ensure thorough and ongoing assessments of AI systems, as well as to maintain meaningful human oversight.

In the following sections, we will explore the considerations to balance between maximizing the benefits of AI and safeguarding against its potential to infringe upon human rights and exacerbate social inequalities. This balanced approach is vital for responsible AI development and deployment in our increasingly technology-driven world and the advent of Generative Artificial Intelligence.

---

[65] Many AI systems are opaque due to their complexity, commercial secrecy or design. This opacity makes it challenging for researchers and governance bodies to access information and fully investigate proprietary datasets, models, and systems. Additionally, AI science is still in its early stages, leading to limited understanding of advanced AI behaviour. This lack of transparency and understanding impedes the identification of risks and the determination of who should be responsible for managing or compensating for potential harms. UN AI Advisory Body (n 74) Para 34 12.

[66] Melissa Hamilton and Pamela Ugwudike, 'A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up' (The Conversation, 26 July 2023) <https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594> accessed 8 July 2024 and Abdul Malek, 'Criminal Courts' Artificial Intelligence: The Way it Reinforces Bias and Discrimination' (2022) 2 AI Ethics 233 <https://doi.org/10.1007/s43681-022-00137-9>.

[67] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health* (WHO, 2021) <www.who.int/publications/i/item/9789240029200> accessed 8 July 2024.

[68] Erik Brynjolfsson, *The Turing Trap: The Promise and Peril of Human-Like Artificial Intelligence* (American Academy of Arts & Sciences, 2022) <www.amacad.org/publication/turing-trap-promise-peril-human-artificial-intelligence> accessed 8 July 2024 and David Rotman, 'How to Solve AI's Inequality Problem' (MIT Technology Review, 19 April 2022) <www.technologyreview.com/2022/04/19/1049378/ai-inequality-problem/> accessed 8 July 2024.

2.1 The rise of Generative AI: Understanding its transformation, capabilities, and human rights implications

Lorenz and others define GenAI as a technology that **creates new content**, such as **text, images, audio, or video**, in response to **prompts** (indications) based on training data. According to them, it has the potential to span various sectors, including education, entertainment, healthcare, and scientific research.[69] The authors also referred that this type of AI systems are increasingly used as 'autonomous agents', which provide new functionalities, enabling them to operate on real-time information and assist users in novel ways, such as making autonomous bookings.[70] 'Generative AI is already used to create individualised content at scale, automate tasks, and improve productivity [and] is yielding benefits in key sectors such as software development, creative industries and arts (eg artistic expression through music or image generation), education (eg personalised exam preparation), healthcare (eg information on tailored preventative care), and internet search'.[71]

As it has been outlined previously, the potential of AI must be considered equally with the human rights risks it poses. For example, in terms of labour displacement, it is estimated that more than half of all jobs could have many of their tasks performed by machines in the next seven years. While 52 million jobs in the United States could have average exposure to automation by 2030.[72]

Madhumita Murgia's definition of GenAI is also similar to those of Lorenz and others. She refers to it as 'software that can write, create images, audio, or video in a way that is largely indistinguishable from the human output. Generative AI is built on the bedrock of human creativity, trained on digitized books, newspapers, blogs, photographs, artworks, music, YouTube videos, Reddit posts, Flickr images and the entire swell of the English-speaking internet. It ingests this knowledge and is able to generate its own bastardized versions of creative products, delighting us with this humanlike ability to remix and regurgitate'.[73] Even though,

[69] Philippe Lorenz, Karine Perset and Jamie Berryhill, 'Initial Policy Considerations for Generative Artificial Intelligence' (OECD Artificial Intelligence Papers, No. 1, 18 September 2023) <https://doi.org/10.1787/fae2d1e6-en>.
[70] ibid 6.
[71] ibid.
[72] Suleyman (n 38) 236 237.
[73] ibid 9.

the similarities Murgia incorporates a characteristic element of these systems and particularly to LLMs which has been called 'hallucination' and that she describes as the 'bastardized versions of creative products'.[74]

Michael Townsen Hicks and others have a similar approach to Murgia's characterisation of LLMs mistaken responses as 'bastardized versions'.[75] They argue that LLMs such as ChatGPT should not be described using the metaphor of 'hallucinations' when they generate false or fabricated statements. They assert that these models are better understood as producing 'bullshit,' a term derived from philosopher Harry Frankfurt, meaning speech that is indifferent to the truth. The authors outline the structure and functioning of LLMs, differentiating between intentional deception (lying) and unintentional falsehoods (hallucinations), and argue that LLMs do neither. Instead, these models generate text that mimics human-like responses without an underlying concern for truth. This characterisation aims to better inform policymakers and the public about the nature of AI outputs and avoid misconceptions.

From a different approach, Samuel Bowman refers to 'hallucination' as 'the problem of LLMs inventing plausible false claims', which represents a 'prominent flaw in current systems and substantially limits how they can be responsibly used.'[76] In a pretty similar line, the Royal Society, a 'fellowship of many of the world's most eminent scientists and is the oldest scientific academy in continuous existence',[77] defines 'hallucination' as the 'generation of convincing and realistic outputs which do not correspond to real-world inputs. Even when there is no malicious intent, general pre-trained transformer (GPT) technologies can fabricate facts, data, and citations when responding to a prompt. The rapid surge of

---

74 Madhumita Murgia, *Code Dependent: Living in the Shadow of AI - Shortlisted for the Women's Prize for Non-Fiction* (Pan Macmillan 2024) 9.

75 Michael Townsen Hicks, James Humphries, and Joe Slater, 'ChatGPT is Bullshit' (2024) 26 Ethics Inf Technol 38 <https://doi.org/10.1007/s10676-024-09775-5>.

76 Samuel R. Bowman, "Eight Things to Know about Large Language Models" (2023) International Conference on Machine Learning <https://arxiv.org/pdf/2304.00612> accessed 30 June 2024.

77 Royal Society, 'Who We Are' (Royal Society, 2024) <https://royalsociety.org/about-us/who-we-are/> accessed 30 June 2024.

machine-generated disinformation online increases the risk that the next generation of models trained on web-scraped data will degrade in performance and absorb distortions and inaccuracies found in fabricated text and data'.[78]

In the study conducted for this thesis to analyse the LLMs' responses in the context of Mexican elections, it was found that *Claude 3 Haiku* generally provides the most 'acceptable' responses but also frequently returns 'no information' responses. As described in the methodology, the evaluation of the LLMs was carried out taking into account a series of defined criteria that included. These criteria allowed the researcher to make a qualitative assessment of the results provided by the chatbots in order to assign a ranking weighting to each of the responses. This indicates that the chatbot often states it cannot provide information and suggests users consult other sources, focusing on certain details. In contrast, *Mistral* showed a higher tendency for 'hallucinations' and 'partially correct' answers with 'hallucinations,' indicating less reliable outputs due to fabricating information about presidential and governor candidates. *Llama 3 Sonar* had the highest count of biased responses and a moderate number of 'hallucinations.' *Chat GPT 3.5* showed a balance, with relatively high counts of acceptable and partially correct responses but fewer 'hallucinations' and biased outputs. Overall, *Claude 3 Haiku* excels in generating acceptable responses, while *Mistral* struggles with reliability.

As showed in the following graphic, the count on the y-axis of the bar chart represents the number of responses that fall into each classification category (Acceptable, Partially Correct, Hallucination, Partially Correct and Hallucination, No Information and Biased) for each LLM. The scale from 1 to 10 quantify and compare the distribution of these classifications among the different LLMs. For example, Claude 3 Haiku had approximately 11 responses classified as acceptable, Chat GPT 3.5, 10 responses, Llama 3 Sonar, 9 responses and Mistral around 4 responses.

[78] Royal Society, 'Science in the Age of AI: How Artificial Intelligence is Changing the Nature and Method of Scientific Research' (May 2024) <https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf> accessed 12 June 2024.

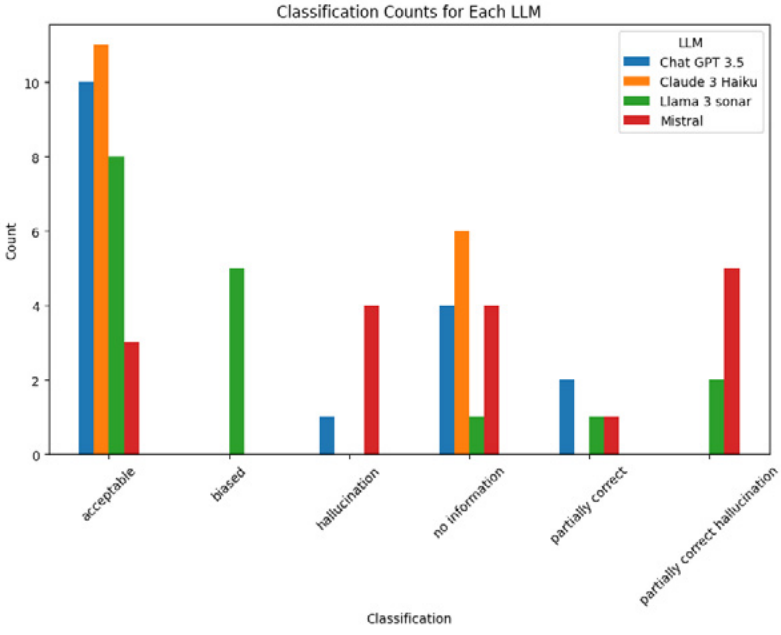

Figure 1. Classification Counts for LLM

Whatever the term to be used, whether it is a more psychological term like 'confabulations,'[79] or a more common like 'hallucination', what is important to consider from a human rights perspective is that companies should comply, among other things, with two key aspects contained in the United Nations Guiding Principles on Business and Human Rights[80] (UNGPs) and that we will analyse in the following chapter. First, they should establish internal procedures to identify possible adverse effects on human rights (due diligence procedures). Second, the creation of mechanisms in which users can access effective remedies when their rights may be negatively impacted. It is worth mentioning,

[79] Will Douglas Heaven, 'Geoffrey Hinton tells us why he's now scared of the tech he helped build' (MIT Technology Review, 2 May 2023) <www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/> accessed 30 June 2024.

[80] United Nations, 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' [hereinafter UNGPs] (New York and Geneva, 2011) <www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 23 June 2024.

for example, the work that Anthropic AI is doing to incorporate a method called 'Constitutional AI'[81] in the deployment building of their LLMs. This method consists of training AI assistants to be helpful and harmless by using AI-generated feedback based on predefined principles rather than relying on human labels for harmlessness. The approach involves two stages: a supervised learning phase, where an AI model critiques and revises its own responses based on constitutional principles, and a reinforcement learning phase using AI-generated feedback.

Finally, aside from the technical definitions, the number of GenAI tools is increasing rapidly.[82] Social media platforms are becoming the digital showcase for dozens of applications and companies, directly or through other users, to showcase the potential of their tools. This can range from LLMs like ChatGPT from Open AI, Claude from Anthropic, Llama developed by Meta or the French Mistral to models for creating: 1) video (Luma Labs,[83] Sora,[84] Viggle;[85]) 2) image animation (Vasa 1,[86] V-Express;[87]) 3) audio, music and clone voice (Applio,[88] XTTS,[89] Suno,[90] Udio[91]) or 4) images (Dall-E,[92] Midjourney,[93] Stable Diffusion). The creation and advancement of GenAI models has triggered numerous copyright lawsuits, with both companies and artists taking legal action.[94]

---

81 Bai Y and others, 'Constitutional AI: Harmlessness from AI Feedback' (arXiv, 15 December 2022) <https://arxiv.org/abs/2212.08073> accessed 30 June 2024.

82 FlexOS, 'Generative AI Top 150: The World's Most Used AI Tools' (29 January 2024) <www.flexos.work/learn/generative-ai-top-150> accessed 30 June 2024.

83 Luma, 'Luma Dream Machine' (Luma, 2024) <https://lumalabs.ai/dream-machine> accessed 30 June 2024.

84 As of the date of this thesis's publication, Sora has not been released. Open AI, Sora <https://openai.com/index/sora/> accessed 24 June 2024.

85 Viggle, 'Viggle AI' (Viggle, 2024) <www.viggle.ai/> accessed 30 June 2024.

86 As of the date of this thesis's publication, Vasa-1 has not been released. Microsoft Research, 'VASA-1' (Microsoft, 2024) <www.microsoft.com/en-us/research/project/vasa-1/> accessed 30 June 2024.

87 Tencent AI Lab, 'V-Express: Conditional Dropout for Progressive Training of Portrait Video Generation' (GitHub, 2024) <https://github.com/tencent-ailab/V-Express> accessed 30 June 2024.

88 Applio, 'Models' (Applio, 2024) <https://applio.org/models> accessed 30 June 2024.

89 Coqui, 'XTTS-v2' (Hugging Face, 2024) <https://huggingface.co/coqui/XTTS-v2> accessed 30 June 2024.

90 Suno, 'About Suno' (Suno, 2024) <https://suno.com/about> accessed 30 June 2024.

91 Udio, 'Udio | AI Music Generator - Official Website' (Udio, 2024) <www.udio.com/> accessed 30 June 2024.

92 Open AI, Dall-E, (Open AI), <https://openai.com/index/dall-e-3/> accessed 30 June 2024

93 Midjourney, 'About' (Midjourney, 2024) <www.midjourney.com/home> accessed 30 June 2024.

94 Jordan Pearson, 'What the RIAA lawsuits against Udio and Suno mean for AI and copyright' (The Verge, 26 June 2024) <www.theverge.com/24186085/riaa-lawsuits-udio-suno-copyright-fair-use-music> accessed 30 June 2024.

As the B-Tech project from the OHCHR explains, the importance of framing the impacts of GenAI in terms of human rights rather than just general societal concerns can be understood because using a human rights framework provides several advantages: i) it relies on internationally agreed-upon norms; ii) reinforces existing legal obligations; iii) focuses on impacts that affect human dignity; iv) provides a clear list of impacts to assess; iv) connects to real-world experiences, and v) taps into existing institutions and movements dedicated to protecting rights.[95]

[95] The B-Tech Project provides authoritative guidance and resources for implementing the United Nations Guiding Principles on Business and Human rights (UNGPs) in the technology space. In 2019, UN Human Rights launched the project after consultations with civil society, business, States, and other experts about the scope of the B-Tech Project. United Nations Human Rights Office of the High Commissioner, 'B-Tech Project' (OHCHR and Business and Human Rights) <www.ohchr.org/en/business-and-human-rights/b-tech-project> accessed 8 July 2024 and United Nations Human Rights Office of the High Commissioner (n 64) 1-2.

3. International human rights standards in the context of GenAI

We are living in an era of unprecedented transformations. As Garnett and James describe, one such transformation is the era of 'cyber elections', marked by technological and sociological changes that introduce new opportunities and threats.[96] In this context, GenAI presents a significant opportunity to expand freedom of expression in elections by enabling people to participate more creatively in political debates. However, it also poses a risk as a tool that can spread disinformation during elections.[97] This chapter explores the intersection of the right to freedom of expression and elections in the disinformation age[98] and the rise of GenAI.[99] It begins by analysing the international human rights framework

[96] Holly Ann Garnett and Toby S James, 'Cyber Elections in the Digital Age: Threats and Opportunities of Technology for Electoral Integrity' (2020) 19(2) *Election Law Journal* <www.liebertpub.com/doi/full/10.1089/elj.2020.0633>.

[97] Harald Stiff and Fredrik Johansson, 'Detecting Computer-Generated Disinformation' (2022) 13 International Journal of Data Science and Analytics 363.; Josh A Goldstein, Girish Sastry, Micah Musser, Renée DiResta, Matthew Gentzel, and Katerina Sedova, 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations' (2023) arXiv:2301.04246 <https://arxiv.org/abs/2301.04246> accessed 4 May 2024.; Danni Xu, Shaojing Fan, and Mohan Kankanhalli, 'Combating Misinformation in the Era of Generative AI Models' (Proceedings of the 31st ACM International Conference on Multimedia, October 29-November 3, 2023, Ottawa, ON, Canada) <https://doi.org/10.1145/3581783.3612704>; S Kreps, RM McCain, and M Brundage, 'All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation' (2022) 9 Journal of Experimental Political Science 104, 117 <https://doi.org/10.1017/XPS.2020.37>.

[98] W Lance Bennett and Steven Livingston (eds), 'The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States' (2021) Cambridge University Press. <https://doi.org/10.1017/9781108914628>.

[99] Hadar Y Jabotinsky and Roee Sarel, 'Co-Authoring with an AI? Ethical Dilemmas and Artificial Intelligence' (2023) SSRN. <https://ssrn.com/abstract=4303959> accessed 4 May 2024.

concerning the right to freedom of expression and political rights in electoral processes. It then examines various definitions of 'election integrity' and the implications of disinformation created with GenAI in the context of elections.

This chapter incorporates a human rights approach to analysing GenAI and its impact on freedom of expression in the context of elections. This approach recognises that technological transformations affect the exercise of human rights, potentially limiting spaces for participation and expression but also expanding ways for individuals to participate in political debates during elections. The human rights approach also establishes fundamental elements to ensure that any restriction on these rights complies with international treaty requirements, preventing censorship and control mechanisms.

This is relevant considering that Generative AI could represent a high risk for unfolding a 'misinformation apocalypse',[100] but also an expansion of the right to freedom of expression in the context of elections. Freedom of expression, the cornerstone of every democratic society, is interlinked with the technological revolution propelled by AI and Deep Learning.[101] This perspective underscores the importance that states and the private sector should consider when approaching such technological transformations and their possible impacts on the electoral process. Generative AI, after all, is a dual-use technology[102] with the potential to help people and facilitate the work of different spheres of human life and potentially undermine the very heart of democracies.[103]

While the private sector, particularly social media platforms and AI companies, are not subject to the same obligations as States to respect, protect and fulfil human rights, they have a duty to respect human rights (Principle 11) according to the United Nations

[100] Mustafa Suleyman and Michael Bhaskar, The Coming Wave: Technology, Power and the 21st Century's Dilemma (Crown 2023) 26.

[101] Royal Society, 'Science in the Age of AI: How Artificial Intelligence is Changing the Nature and Method of Scientific Research' (May 2024) <https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf> accessed 12 June 2024.

[102] In 1986, Melvin Kranzberg postulated that technology is neither inherently good nor bad, nor is it neutral; its effects are dependent on the social, environmental, and historical context in which it is embedded. His postulations are also known as the Kranzberg's Laws. Melvin Kranzberg, 'Technology and History: 'Kranzberg's Laws'' (1986) 27(3) Technology and Culture 544-560.

[103] UN AI Advisory Body, 'Interim Report: Governing AI for Humanity' (2023) 11 <www.un.org/en/ai-advisory-body> accessed 12 June 2024.

Guiding Principles on Business and Human Rights (UNGPs).[104] Human rights due diligence is increasingly penetrating national and international legislation. States have consistently moved to create enforcement mechanisms that establish higher responsibilities for companies to respect and address any adverse human rights impacts. The EU directive 2024/1760 on corporate sustainability due diligence,[105] France, Germany, The Netherlands, the US, the UK and Australia have taken steps to fulfil their duty to protect human rights by establishing stronger regulations for companies to comply with their due diligence responsibilities.[106]

This implies that companies should, among other things, conduct 'human rights due diligence'[107] to 'identify, prevent, mitigate and account for adverse impacts of their activities' or 'access effective remedy mechanisms'.[108] They can do so by reviewing the effectiveness, promptness and appropriateness of actions taken by social media platforms and Generative AI companies to establish internal mechanisms and review the operation of their platforms to identify, prevent and mitigate any negative impact on exercising human rights. Furthermore, to establish effective procedures to address any potential violations of users' human rights.

3.1 The right to freedom of expression and the right to public participation in international human rights law

Article 19 of the UDHR establishes that 'everyone has the **right to freedom of opinion** and **expression**'(emphasis added) [including the] 'freedom to hold opinions **without interference** and to **seek, receive, and impar**t information and ideas **through any media and regardless of frontiers**'(emphasis added).[109] The right

[104] UNGPs Principle 11 13.

[105] Directive (EU) 2024/1760 of the European Parliament and of the Council of 13 June 2024 on corporate sustainability due diligence and amending Directive (EU) 2019/1937 and Regulation (EU) 2023/2859 [2024] OJ L1760/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401760> accessed 12 June 2024.

[106] John Sherman, 'Human Rights Due Diligence and Corporate Governance' (29 April 2022) Human Rights Due Diligence for Lawyers, American Bar Association, forthcoming, 16 <http://dx.doi.org/10.2139/ssrn.3862624> and Basak Baglayan, 'A Study on Potential Human Rights Due Diligence Legislation in Luxembourg' (2021) 28 <https://orbilu.uni.lu/bitstream/10993/48683/1/Baglayan_Study_HRDD.pdf> accessed 8 July 2024.

[107] UNGPs Principle 15 (b) 16 17.

[108] UNGPs Principle 25 27.

[109] ibid.

to freedom of expression has two dimensions: individual and collective. The individual dimension recognises the right of every person 'to use whatever medium is deemed appropriate to impart ideas and to have them reach as wide an audience as possible'.[110] The collective dimension refers to the fundamental nature of exchanging ideas and information among people through any media, highlighting that every individual has the right to express and share their views and to access and receive opinions and news from others. Both dimensions are crucial for forming the basis of democracy and fostering an informed and engaged public.[111]

Article 19 (2) of the ICCPR reaffirms the same principles in the UDHR and provides additional relevant elements to consider when looking at technological transformations. It states that 'everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers',[112] and highlights that it can be done 'either orally, in writing, or in print, in the form of art, or **through any other media of his choice**'.[113]

The emphasis placed by both instruments on the phrases 'through any media' and 'through any other media of his choice' has set an important standard applicable to technological developments, first in the context of the expansion of the Internet in recent decades. In 2011, the Special Rapporteur on Freedom of Expression and Opinion (SRFEO) expressed that 'very few, if any, developments in information technologies have had such a revolutionary effect as the creation of the Internet',[114] where 'individuals are no longer passive recipients but also active publishers of information'.[115] The SRFEO underscored that Article 19 of the UDHR and the ICCPR 'was drafted with the foresight to include and accommodate future technological developments through which individuals can exercise their right to freedom of expression'.[116]

---

[110] Compulsory Membership of Journalists, Advisory Opinion OC-5/85 (IACHR, 13 November 1985), para 31 p 9.
[111] ibid para 32 p 9.
[112] ibid.
[113] ibid.
[114] United Nations Human Rights Council, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27 par 19 p6<www.ohchr.org/sites/default/files/Documents/Issues/Opinion/A-HRC-17-27.pdf> accessed 12 June 2024.
[115] ibid par 19 p6.
[116] ibid par 21 p7.

Freedom of expression is not an absolute right and may be subject to certain restrictions. According to Article 19 (3) of the ICCPR, this right carries 'duties' and 'responsibilities' and may be restricted for the '(a) respect of the rights or reputations of others' and '(b) protection of national security or public order (*ordre public*), or of public health or morals.' To be compatible with international human rights law, any restriction must meet the 'three-part test', which includes the following:

— **Provided by Law:** The limitation must be clearly and precisely outlined in law.

— **Legitimate Aim:** The restriction must pursue a legitimate aim, such as protecting the rights or reputations of others, public health or morals, national security, or public order.

— **Necessary and Proportionate:** The measures taken must be the least restrictive means to achieve the intended purpose and necessary in a democratic society.

Additionally, Article 20 of the ICCPR requires states to restrict 'any propaganda for war'[117] and speech that constitutes 'advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence'.[118]

General Comments are authoritative interpretations of international human rights treaties issued by UN treaty bodies to clarify state obligations.[119] General Comment 34, issued by the Human Rights Committee (GC34), specifically interprets Article 19 of the International Covenant on Civil and Political Rights, focusing on freedom of opinion and expression. It emphasises the importance of protecting uninhibited political discourse, especially during

[117] ICCPR Article 20 (1).
[118] ICCPR Article 20 (2).
[119] For Max Lesch and Nina Reiners, General Comments 'are law-making instruments because, despite their non-binding nature, they shape the interpretation, application and development of international human rights law'. Max Lesch and Nina Reiners, 'Informal Human Rights Lawmaking: How Treaty Bodies Use General Comments to Develop International Law' (2023) 12(2) Global Constitutionalism 378-401 <www.cambridge.org/core/journals/global-constitutionalism/article/informal-human-rights-lawmaking-how-treaty-bodies-use-general-comments-to-develop-international-law/7D1E7EF25889DDD944D8FB2691AA36A7> accessed 4 June 2024.

elections.[120] This protection is crucial in this context, where disinformation generated by AI can significantly distort public debate and influence voter decisions. The General Comment provides a basis for ensuring that efforts to combat disinformation do not unduly restrict freedom of expression, allowing for robust political debate and criticism, which are essential for a democracy. 'For example, it may be legitimate to restrict freedom of expression in order to protect the right to vote under article 25 [...] Such restrictions must be constructed with care: while it may be permissible to protect voters from forms of expression that constitute intimidation or coercion, such restrictions must not impede political debate, including, for example, calls for the boycotting of a non-compulsory vote'.[121] Moreover, the principles highlighted in the General Comment caution against overly restrictive measures that could hinder freedom of expression in the context of an election.[122] This principle may also be applicable even when combating disinformation. It gives the framework to underscore the need for a balanced approach, where efforts to counter disinformation do not lead to the restriction of legitimate political criticism or dissent. In the age of generative AI capable of swiftly producing and circulating information, it is vital to establish regulations that address AI-generated content without encroaching upon the fundamental right to free expression. Striking this balance is crucial for upholding democratic integrity while preserving the capacity to critique and hold public figures and institutions accountable.

However, this balance is particularly challenging to achieve in environments where the judiciary lacks independence. In such countries, Bail and others argue that there is a high risk that electoral laws will be misused to target political opponents rather than addressing genuine threats to electoral integrity. Therefore, this misuse of legal measures can undermine the fairness of the election and stifle dissent. The authors caution, 'in contexts where the judiciary is not independent, there is a high risk that broadly or vaguely worded laws will be misused to target political opponents and critics rather than addressing the actual threat of false

[120] Human Rights Committee, 'General Comment No 34 on Article 19: Freedoms of Opinion and Expression' (12 September 2011) [Hereinafter CCPR/C/GC/34] <https://undocs.org/Home/Mobile?FinalSymbol=a%2FHRC%2F26%2F30&Language=E&DeviceType=Desktop&LangRequested=False> accessed 22 June 2024 para 38 p 9.

[121] CCPR/C/GC/34 par 28 7.

[122] ibid.

news'.[123] This highlights the necessity of having an impartial judiciary to ensure that laws targeting disinformation are applied fairly and justly, maintaining the democratic process and protecting political rights.

When referring to electoral processes, Article 21 of the UDHR establishes three standards that guarantee the right to participate in public affairs and sets the democratic principles of societies and the core elements of what is described as 'free' and 'fair' elections:

— Right to Participate: Article 21 (1) underscores that 'everyone has the right **to take part in the government** of his country, **directly** or **through freely chosen representatives**' (emphasis added).[124] This ensures that every person can be involved in their country's governance, either directly or indirectly, by voting for representatives. This principle is fundamental to democratic systems, aiming to ensure everyone is considered when formulating policies and laws. It seeks to promote inclusivity, accountability, and the fair representation of diverse perspectives within a state.

— Equal Access: Article 21 (2) establishes that 'everyone has the right of equal access to public service in his country'.[125] This guarantees equal participation in public services, ensuring fairness and equality within society.

— Basis of Authority: Article 21 (3) states that '**the will of the people** shall be the basis of the authority of government; this will **shall be expressed** in **periodic** and **genuine elections** which shall be by **universal** and **equal suffrage** and shall be held by secret vote or by equivalent **free voting** procedures'(emphasis added).[126] This principle ensures that the government's authority is based on the people's will, expressed through regular, genuine elections conducted with universal and equal suffrage. This means that all eligible citizens have an equal

[123] Christopher A Bail and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization' (2018) 115(37) PNAS 9216, 9221 157 <https://doi.org/10.1073/pnas.1804840115>.
[124] UDHR Article 21 (1).
[125] UDHR Article 21 (2).
[126] UDHR Article 21 (3).

right to vote in elections that are free, fair, and held at consistent intervals. The voting process must also ensure secrecy and freedom, protecting individuals from coercion or intimidation. This principle ensures that the government remains accountable to its citizens and reflects their collective will, thereby upholding the fundamental values of democracy.

Article 25 (b) of the ICCPR reinforces the UDHR by establishing the right 'to vote and to be elected at **genuine periodic elections** which shall be by **universal** and **equal suffrage** and shall be **held by secret ballot** guaranteeing the **free expression of the will** of the electors'(emphasis added). This article strengthens democratic governance by promoting fair, transparent, and inclusive electoral practices, which are essential for the legitimacy and accountability of any government.

General Comment 25 of the Human Rights Committee (GC25) has also reaffirmed that 'voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind'.[127] In that sense, disinformation campaigns propelled with the use of Generative AI, especially with anthropomorphising chatbots,[128] can affect people's ability to form their opinions independently and could be considered manipulative interference as it gives 'the illusion that something human is behind the screen'.[129]

---

[127] United Nations Human Rights Committee, 'General Comment No 25 on Article 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service' (1996) CCPR/C/21/Rev1/Add.7 [Hereinafter CCPR/C/21/Rev1/Add.7] par 19 6 <https://undocs.org/Home/Mobile?FinalSymbol=CCPR%2FC%2F21%2FRev.1%2FAdd.7&Language=E&DeviceType=Desktop&LangRequested=False> accessed 22 June 2024.

[128] 'Anthropomorphism' can be defined as the 'the inclination to ascribe humanlike qualities to machines. LLMs are constantly improving and captivating with their human-mimicking capabilities to create texts, respond to various questions, and interact as if dialoguing with someone who has the answers to almost everything that can be asked. When chatbots mimic human interaction, they might present misleading emotional responses and potentially harmful suggestions without true understanding or empathy. For example, in 2023, an eating disorder chatbot was suspended in the US for giving harmful responses to people. Amanda Hoover, 'An Eating Disorder Chatbot Is Suspended for Giving Harmful Advice' (Wired, 1 June 2023) <www.wired.com/story/tessa-chatbot-suspended/> accessed 12 May 2024 and Taylor Majewski, 'It's time to retire the term "user"' (MIT Technology Review, 19 April 2024) <www.technologyreview.com/2024/04/19/1090872/ai-users-people-terms/> accessed 30 June 2024.

[129] Will Bedingfield, 'A Chatbot Encouraged Him to Kill the Queen. It's Just the Beginning' (WIRED, 18 October 2023) <www.wired.com/story/chatbot-kill-the-queen-eliza-effect/?redirectURL=%2Fstory%2Fchatbot-kill-the-queen-eliza-effect%2F> accessed 22 June 2024.

Not only that, Smuha and others[130] have argued that manipulative AI threatens democracy, creating a dire scenario to manipulate and potentially destabilise democratic processes, leading to an 'existential risk of an irreversible totalitarian regime'.[131]

In 1993, the Vienna Declaration and Programme of Action reaffirmed that all human rights, such as freedom of expression and the right to participate in elections and public affairs, are universally **inalienable, indivisible, interdependent**, and **interrelated**.[132] Accordingly, human rights are intrinsic to every individual without exception and can only be restricted in exceptional circumstances with full due process. Each category of rights strengthens and reinforces the others, creating a powerful, interconnected framework. Advancing one right paves the way for others, while denying one undermines the rest. This mutually reinforcing dynamic ensures that the realisation of one right fuels the realisation of all, fostering a holistic approach to human dignity and justice.

> 'Democracy, development and respect for human rights and fundamental freedoms are interdependent and mutually reinforcing. Democracy is based on the freely expressed will of the people to determine their own political, economic, social and cultural systems and their full participation in all aspects of their lives. In the context of the above, the promotion and protection of human rights and fundamental freedoms at the national and international levels should be universal and conducted without conditions attached'.[133]

[130] Nathalie Smuha and others, 'We Are Not Ready for Manipulative AI – Urgent Need for Action' (Euractiv, 2023) <https://kuleuven.limo.libis.be/discovery/search?query=any,contains,LIRIAS4076667&tab=LIRIAS&search_scope=lirias_profile&vid=32KUL_KUL:Lirias&offset=0> accessed 5 June 2024.

[131] Hendrycks, D. Mazeika, M.,Woodside, T. (2023, June 21). An overview of catastrophic AI risks. arXiv.org. <https://arxiv.org/abs/2306.12001> accessed 30 June 2024.

[132] United Nations, 'Vienna Declaration and Programme of Action' (25 June 1993) <www.ohchr.org/en/instruments-mechanisms/instruments/vienna-declaration-and-programme-action> accessed 12 June 2024.

[133] ibid.

In this regard, there is an interconnection and reinforcing mechanism between the right to freedom of expression and the right to participate in public affairs and to vote in electoral processes. Moreover, it establishes a framework for examining the nexus of freedom of expression, disinformation, and the influence of generative AI on democratic processes as a complex and interwoven phenomenon.

An integral element of a democratic society and the conduction of an election relies on the principles described above, reaffirming that the 'free flow of ideas is incontestably a core requirement for the promotion of democratic spaces'.[134] As protected by international treaties, freedom of expression ensures that ideas and opinions can circulate freely, creating an informed electorate and vibrant public discourse.

The right to freedom of expression is also upheld in various regional jurisdictions. To illustrate the application of this right within the context of Mexican elections, relevant standards from Latin America will be explored. Additionally, to provide a robust jurisprudential background, this discussion will incorporate the European human rights system, which was the 'first to enforce and provide binding force to some of the rights stated in the Universal Declaration of Human Rights'.[135]

The American Convention on Human Rights (ACHR), known as the 'Pact of San José,' specifically protects the right to freedom of expression under Article 13. It includes provisions for restricting this right, mirroring those mentioned earlier. Furthermore, the ACHR states that freedom of expression 'cannot be subject to prior censorship but only to subsequent liabilities',[136] highlighting a significant aspect of the legal framework that supports this fundamental right. This connection between various systems underlines the universal importance and robust nature of protecting freedom of expression in different legal contexts.

---

[134] United Nations Human Rights Council, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' (2 July 2014) UN Doc A/HRC/26/30 [hereinafter A/HRC/26/30] par 7 3 <https://undocs.org/Home/Mobile?FinalSymbol=a%2FHRC%2F26%2F30&Language=E&DeviceType=Desktop&LangRequested=False> accessed 12 June 2024.

[135] Council of Europe, 'The Convention in 1950 - The European Convention on Human Rights' (Council of Europe, 1950) <www.coe.int/en/web/human-rights-convention/the-convention-in-1950> accessed 22 June 2024.

[136] Organization of American States, 'American Convention on Human Rights' (22 November 1969) <https://cidh.oas.org/Basicos/English/Basic3.American%20Convention.htm> accessed 7 June 2024.

The Inter-American Court of Human Rights (IACtHR) has addressed the importance of the right to freedom of expression in the political debate preceding elections in the *Case of Ricardo Canese v Paraguay*. The Court highlighted the critical need to safeguard the free flow of information and ideas, reaffirming that:

> [...] the exercise of freedom of expression should be protected and guaranteed in the political debate that precedes the election of State authorities who will govern a State. The formation of the collective will through the exercise of individual suffrage is nourished by the different options presented by the political parties through the candidates that represent them. Democratic debate implies that the free circulation of ideas and information on the candidates and their political parties is permitted through the media, the candidates themselves, and any individual who wishes to express his opinion and provide information. Everyone must be allowed to question and investigate the competence and suitability of the candidates and also to disagree with and compare proposals, ideas, and opinions so that the electorate may form its opinion in order to vote. In this respect, the exercise of political rights and freedom of thought and expression are closely related and reinforce one another.[137]

In relation to the European system, the European Court of Human Rights (ECtHR) has also emphasised the importance of freedom of expression in relation to elections (*Orlovskaya Iskra v Russia*, § 110;[138] *Cheltsova v Russia*, § 96; *Długołęcki v Poland*, § 40;[139] *Bowman v the United Kingdom* [GC], § 42),[140] both national and local (*Cheltsova v Russia*, § 96),[141] especially in the period leading up to election day. For instance, in the *Case of Kwiecień v Poland*, the Court stated:

---

[137] Inter-American Court of Human Rights, 'Case of *Ricardo Canese vs Paraguay*' (31 August 2004) Series C No 111 <www.corteidh.or.cr/docs/opiniones/seriea_05_ing.pdf> accessed 12 June 2024.

[138] *Orlovskaya Iskra v Russia* App no 42911/08 (ECHR, 21 February 2017) <http://hudoc.echr.coe.int/eng?i=001-171525>.

[139] *Długołęcki v Poland* App no 23806/03 (ECHR, 24 February 2009) <http://hudoc.echr.coe.int/eng?i=001-91475>.

[140] *Bowman v United Kingdom* App no 24839/94 (ECHR, 19 February 1998) <http://hudoc.echr.coe.int/eng?i=001-58134>.

[141] *Cheltsova v Russia* App no 44294/06 (ECHR, 13 June 2017) <http://hudoc.echr.coe.int/eng?i=001-174381>.

> Free elections and freedom of expression, particularly freedom of political debate, together form the bedrock of any democratic system (see Mathieu-Mohin and Clerfayt v Belgium, judgment of 2 March 1987, Series A no 113, p 22, § 47). The two rights are inter-related and operate to reinforce each other. For this reason, it is particularly important in the period preceding an election that opinions and information of all kinds are permitted to circulate freely (see Bowman v the United Kingdom, judgment of 19 February 1998, Reports 1998-I, § 42). This principle applies equally to national and local elections.[142]

Ensuring that all voices have a space in the public debate is imperative.[143] This inclusivity is essential for a fair democratic process and aligns with the ICCPR's emphasis on equality and non-discrimination in political participation. In *Mathieu-Mohin and Clerfayt v Belgium*, § 54, the ECtHR has also underlined this principle by stating the 'equality of treatment of all citizens in the exercise of their right to vote and their right to stand for election'.[144] Additionally, the SRFEO has addressed that the 'full realisation of the right to access information is another crucial element in the promotion of free and fair democratic elections'.[145] Access to diverse and reliable information enables voters to make informed decisions, reinforcing the integrity of elections and the free will, especially in combating disinformation that can skew public perception and impact democratic processes.

The CCPR/C/21/Rev1/Add7 has also stated that citizen's participation in public affairs is supported by ensuring the right to freedom of expression where 'by exerting influence through public debate and dialogue with their representatives or through their capacity to organize themselves'.[146]

---

142 European Court of Human Rights, 'Case of *Kwiecień v Poland* App no 51744/99 (ECHR, 9 January 2007) <http://hudoc.echr.coe.int/eng?i=001-78876>.

143 Inter-American Court of Human Rights, 'Caso *López Lone y otros vs Honduras'* (5 October 2015) Serie C No. 302, par 87, 162 and 163 available at <http://www.corteidh.or.cr/docs/casos/articulos/seriec_302_esp.pdf> accessed 30 June 2024.

144 *Mathieu-Mohin and Clerfayt v Belgium,* App no 9267/81 (ECHR, 2 March 1987) <https://hudoc.echr.coe.int/eng?i=001-57536>.

145 A/HRC/26/30 par 10 4.

146 CCPR/C/21/Rev.1/Add.7 par 8.

On the same line, in 2021 the SRFEO published a report on *Disinformation and freedom of opinion and expression,* stating that 'diverse and reliable information is an obvious antidote to disinformation and misinformation'.[147] Therefore, 'states should fulfil their duty to ensure the right to information, firstly, by increasing their own transparency and by proactively disclosing official data online and offline and, secondly, by reaffirming their commitment to media freedom, diversity and independence'.[148]

Pluralism and access to information are core elements for free and fair elections, allowing every person to be 'well-informed' and contribute to the legitimacy of elections. Misused, GenAI has the potential to spread disinformation, but if leveraged responsibly, it can enhance access to information and participation in the political debate in the context of elections.

In the information ecosystem, other actors play a significant role in safeguarding access to information in the context of elections. A free press acts as a watchdog, holding power to account and ensuring transparency. Its role is vital for disseminating accurate information and countering disinformation, especially in the age of GenAI. As the SRFEO has pointed out, 'an independent, pluralistic, and free press is essential to the development and maintenance of democracy in a nation'.[149] The press has also played a significant figure in the context of disinformation in at least three dimensions: 1) conducting investigations to understand the phenomenon of disinformation and the actors behind their spread, 2) promoting counter-narratives with their fact-check initiatives, especially in the electoral processes and 3) assuring that citizens are able to take more informed decisions by accessing reliable information. Open debates and the free flow of information are critical to preserving democratic societies, and the challenge is to balance this openness while mitigating the spread of disinformation, a task complicated by the capabilities of GenAI.[150]

---

147 United Nations Human Rights Council, 'Disinformation and Freedom of Opinion and Expression' (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, 13 April 2021) UN Doc A/HRC/47/25 [Hereinafter A/HRC/47/25] para 43 18 <www.undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F47%2F25&Language=E&DeviceType=Desktop&LangRequested=False> accessed 12 June 2024.

148 ibid.

149 A/HRC/26/30 Para 58 25.

150 Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (2021) 19-20 <https://arxiv.org/pdf/2108.07258.pdf> accessed 5 June 2023.

However, as the SRFEO has stated, 'freedom of expression plays a central role in ensuring that political processes are open, free, and fair, thus guaranteeing a functioning and effective democracy'.[151] Protecting freedom of expression is key to maintaining open and fair political processes. It allows for the free exchange of ideas and combats the spread of disinformation, which GenAI can exacerbate.

GenAI can democratise information by enabling broader access to information in several ways. For instance, AI can translate news articles into multiple languages, ensuring non-native speakers can access the same information.[152] It can also create personalised news feeds tailored to users' interests, helping them stay informed about topics they care about. Moreover, AI can generate summaries of complex texts, making them more accessible to a broader audience.[153] However, it also risks amplifying dominant voices if not managed properly.

For example, the Clemson University report 'Old Despots New Tricks - AI-Empowered Pro-Kagame/RPF Coordinated Influence Network on X' (formerly Twitter) highlighted how GenAI has been employed to influence public discourse in Rwanda[154]. This case study provides concrete examples of how GenAI was used to create a large volume of content promoting specific narratives favourable to the Rwandan government. This includes using AI-generated text and imagery to flood social media platforms with messages that support President Kagame and the Rwandan Patriotic Front (RPF). By automating content creation, GenAI allowed for the rapid dissemination of information, enabling broader access to government-promoted narratives.

The strategy identified was used to amplify dominant voices, particularly those aligned with the government, by overshadowing dissenting opinions. The coordinated network used AI to produce and post repetitive messages, drowning out critical voices

---

151 A/HRC/26/30 Par 18 7.

152 Pushpdeep Singh, Mayur Patidar, and Lovekesh Vig, 'Translating Across Cultures: LLMs for Intralingual Cultural Adaptation' (20 June 2024) arXiv:2406.14504 <https://arxiv.org/abs/2406.14504> accessed 12 June 2024.

153 Justin Muller, 'Advanced Generative Summarization Techniques: A deep dive with example code' (Medium, 2024) <https://medium.com/@flux07/advanced-generative-summarization-techniques-939605601fba> accessed 12 June 2024.

154 Morgan Wack, Darren Linvill, and Patrick Warren, 'Old Despots, New Tricks - An AI-Empowered Pro-Kagame/RPF Coordinated Influence Network on X' (Clemson University, June 2024) <https://tigerprints.clemson.edu/mfh_reports/5> accessed 22 June 2024.

and manipulating public perception. This highlights the dual-use nature of GenAI: while it can facilitate wider participation in public debates, it can also dominate and control the narrative, limiting the diversity of voices in the public sphere.

For this reason, ensuring an 'open space for the multiple voices of politicians, the press, minorities, and citizens in general is a permanent challenge for entities tasked with overseeing electoral processes'.[155] Effective oversight ensures diverse voices are heard and protects against disinformation. That is why regulators must adapt to new challenges posed by technologies like GenAI.[156]

By connecting these rights and principles, we see a comprehensive framework that upholds the integrity of democratic processes, ensuring that freedom of expression and access to accurate information are safeguarded against the threats posed by disinformation and the potential misuse of GenAI.

3.2 UN Guiding Principles on Business and Human Rights in the Era of Generative AI

The United Nations Guiding Principles on Business and Human Rights (UNGPs) marked a significant milestone in international law. They forged a crucial link between the responsibility of companies to consider the potential impact of their activities on human rights and the imperative for states to establish systems for their protection. This recognition of the state's duty to protect human rights and business enterprises' entitlement to respect them was a pivotal moment in the history of international law.[157] Its importance in the age of GenAI lies in the possibility for companies that provide services to create images or interact with a chatbot, and the social media platforms that distribute their content, to incorporate or reinforce two crucial elements in their own processes that we will explain later in this section: 1) due diligence mechanism in the deployment of their technologies and 2) remedy mechanisms that allow users to access legitimate processes whenever any adverse effects may occur.

[155] A/HRC/26/30 Par 9 3.

[156] UNESCO, *Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multistakeholder Approach* (UNESCO 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000387339> accessed 22 June 2024.

[157] UNGPs.

As a non-binding mechanism, UNGP established a three-pillar framework to: 1) protect, 2) respect and 3) access an effective remedy.[158] The first pillar is based on the role of states to establish the necessary safeguards to protect against any violations by companies of human rights. The second pillar refers to the corporate responsibility to respect human rights when deploying their activities. Finally, the third pillar incorporates the recognition to access a remedy for the victims of actions or omissions arising from companies' activities.

For this thesis, there are two core elements that the UNGPs apply to companies working on deploying AI systems and social media platforms that may distribute their AI generated content. First, to effectively implement corporate responsibility in respecting human rights. In relation to the second pillar of the UNGPs, companies should comply with the 'human rights due diligence'. In this regard, Principle 17 states that 'business enterprises should carry out human rights due diligence' so they can '**identify, prevent, mitigate** and account for how they address **their adverse human rights impacts**'(emphasis added).[159] For this purpose, business enterprises should 'include **assessing actual and potential human rights impacts**, **integrating** and **acting** upon the findings, tracking responses, and communicating how impacts are addressed'(emphasis added).[160] The principle also includes three additional characteristics for companies to comply with the 'human rights due diligence': '(a) Should **cover** adverse human rights **impacts that the business enterprise may cause or contribute** to through **its own activities**, or which may be directly **linked to its operations**, **products** or **services** by its business relationships; (b) Will **vary in complexity** with the **size of the business enterprise**, the **risk** of severe human rights impacts, and the **nature** and **context of its operations** and (c) **Should be ongoing**, recognizing that the human rights risks may change over time as the business enterprise's operations and operating context evolve'(emphasis added).[161]

[158] Emoke Bebiak, 'Human Rights Due Diligence: The European Union's Approach to Ensuring Respect for Human Rights in Business' (European Master's Degree in Human Rights and Democratisation, Adam Mickiewicz University 2018/2019) <http://dx.doi.org/10.25330/1936>.

[159] UNGPs Principle 17 8.

[160] ibid.

[161] ibid Principle 17 8.

In other words, due diligence requires companies to i) identify their activities for potential human rights violations, ii) take necessary measures to prevent and mitigate identified risks and monitor the effectiveness of these actions, and iii) transparently communicate information about human rights impacts and their mitigation efforts to the public.[162] The SRFEO has stated for example, that social media platforms should perform human rights impact assessments on their products, focusing on how algorithms and ranking systems might spread disinformation or misinformation and that these assessments should be conducted regularly and specifically before and after major events, such as national elections or significant crises like the COVID-19 pandemic.[163] In light of the adoption of GenAI, social media platforms like Meta have been exploring measures to incorporate watermarking mechanisms and labels to identify AI generated content.[164] Although, some authors like Leibowicz and Evan Harris have been critical about the effectiveness of the measure.[165] In any case, to be meaningful, Harrison argued that the human rights due diligence principle needs to incorporate elements such as transparency, external participation, and independent monitoring.[166]

The second core element is that business enterprises should take further steps when identifying that 'they have caused or contributed to adverse impacts'[167] so they can 'provide for or cooperate in their remediation through legitimate processes'.[168] In other words, 'remediation aims to put right any actual human rights

[162] Emoke Bebiak (n 158).
[163] A/HRC/47/25 para 96 19.
[164] Monika Bickert, 'Our Approach to Labeling AI-Generated Content and Manipulated Media' (Meta, 17 April 2024) <https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/ accessed> 29 June 2024.
[165] Claire Leibowicz, 'Why watermarking AI-generated content won't guarantee trust online' (MIT Technology Review, 9 August 2023) <www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/> accessed 10 July 2024 and David Evan Harris and Lawrence Norden, 'Meta's AI Watermarking Plan Is Flimsy, at Best' (IEEE Spectrum, 4 March 2024) <https://spectrum.ieee.org/meta-ai-watermarks> accessed 10 July 2024.
[166] James Harrison, 'Establishing a Meaningful Human Rights Due Diligence Process for Corporations: Learning from Experience of Human Rights Impact Assessment' (2013) 31(2) Impact Assessment and Project Appraisal 107, 117 <http://dx.doi.org/10.1080/14615517.2013.774718>.
[167] UNGPs Principle 22.
[168] ibid.

impact that an enterprise causes or contributes to'.[169] The UNGPs emphasise the need for effective remedies, including judicial and non-judicial mechanisms, to address business-related human rights harms. The OHCHR has highlighted the importance of understanding the procedural and substantive aspects of remedies and the roles of different actors, including states, technology companies, and civil society, in providing these remedies.[170] It has also stated that various forms of remedies can take place, such as restitution, compensation, rehabilitation, satisfaction, and guarantees of non-repetition, but they should be context-specific and meet the needs of affected individuals and communities (the 'effectiveness' of the remedy should rely on the people affected).[171] The OHCHR also stresses the need for a 'smart mix' of measures, combining legal standards with education, awareness-raising, and capacity-building to enhance access to remedies.[172] It underscores, as well, the role of company-based grievance mechanisms as essential tools for early and direct resolution of human rights grievances and prevention of future harms.[173] This is particularly relevant for social media companies in the context of elections and the possible impacts to the right of freedom of expression of users. The OHCHR also provide practical guidance for designing and implementing effective non-judicial grievance mechanisms, based on the eight 'effectiveness criteria' outlined in the UNGPs: legitimacy, accessibility, predictability, equitability, transparency, rights-compatibility, source of continuous learning, and engagement and dialogue.[174] By integrating these principles, the OHCHR argues, technology companies can better align their operations with human rights standards and ensure that affected individuals receive appropriate and effective remedies.[175]

---

169 Office of the High Commissioner for Human Rights, 'The Corporate Responsibility to Respect Human Rights: An Interpretive Guide' (United Nations 2012) <www.ohchr.org/sites/default/files/Documents/Publications/HR.PUB.12.2_En.pdf> accessed 17 June 2024.

170 United Nations Human Rights Office of the High Commissioner, 'Access to Remedy and the Technology Sector: Basic Concepts and Principles' (2023) <www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf> accessed 10 July 2024.

171 ibid 4.

172 ibid 7.

173 United Nations Human Rights Office of the High Commissioner, 'Designing and Implementing Effective Company-Based Grievance Mechanisms: A B-Tech Foundational Paper' (2021) <www.ohchr.org/Documents/Issues/Business/B-Tech/access-to-remedy-company-based-grievance-mechanisms.pdf> accessed 10 July 2024.

174 UNGPs Principle 31.

175 United Nations Human Rights Office of the High Commissioner (n 170).

3.3 Social media platforms and their role in the times of Generative AI

In 2004 (Meta) Facebook had only a few tens of thousands of users. In 2023, there was an average of 3.96 billion monthly users and an increase of 7 percent year-on-year. In those years, Facebook went from a digital network of academics to a 'new ruler' of online expression with a global presence. And with that, it also transformed its 'governance system'[176] of content moderation, moving from the premise, 'if there is something that makes you feel bad in your gut, remove it' to the creation of 'community standards' to dictate what can be posted on its platform. At the same time, its expansion and growing influence in civic space and democratic development generated a growing call for transparency and accountability in the face of the platform's stumbling blocks throughout its history.

Social media platforms have taken on a central role in exercising the right to freedom of expression, information flows, and civic space.[177] Around the world, people have found a space for participation where they share and access a diversity of information. However, these spaces are subject to private 'governance systems' that set the boundaries over which people post content on their platforms.

[176] Kate Klonick describes 'new models of governance' in terms of 'several characteristics that accurately describe the interaction between user and platform, a law-making, dynamic and iterative, norm-generating process, and [a] convergence of processes and outcomes'. She also argues that to 'better understand online discourse, we must abandon traditional doctrinal and normative analogies and understand private platforms as systems of governance', Kate Klonick, "Governors…", 1617 and 1599.

[177] 'Civic space is the environment that enables civil society to play a role in the political, economic and social life of our societies. In particular, civic space enables individuals and groups to contribute to the development of policies that affect their lives by, among other things; (i) accessing information; (ii) engaging in dialogue; (iii) expressing disagreement; and (iv) coming together to express their views. An open and pluralistic civic space that guarantees freedom of expression and opinion, as well as freedom of assembly and association, is a prerequisite for sustainable development and peace'. Reference?

These complex and diverse 'governance systems' of content moderation have been subject to scrutiny by researchers,[178] civil society organisations,[179] governments and international bodies such as the Inter-American Commission on Human Rights (IACHR), which warned of a 'turning point for freedom of expression on the internet' and called for a regional dialogue that, among other things, 'helps to make internet content moderation compatible with human rights standards'.[180]

In recent years, the global debate in the internet governance forum has focused on exploring ideal ways to protect users' rights, address the power of digital platforms and mitigate the negative impact on democracies and address the future challenges of generative artificial intelligence.[181]

The journey of digital platforms over the years has not been easy. With scandal after scandal, Facebook has had to apologise repeatedly, for[182] Cambridge Analytica, the interference in the 2016 US elections,[183] the genocide committed against the Rohingya people in Myanmar,[184] and so on. In the face of these and other events, states have been strongly pressured to establish regulatory

---

178 Layla Mashkoor "Sheikh Jarrah content takedowns reveal pattern of online restrictions in Palestine", May 10, 2021, <www.thenationalnews.com/mena/sheikh-jarrah-content-takedowns-reveal-pattern-of-online-restrictions-in-palestine-1.1220037> accessed 29 June 2024.

179 APC, "APC policy explainer: Platform responsibility and accountability", Association for Progressive Communications, November 1, 2020 <www.apc.org/en/pubs/apc-policy-explainer-platform-responsibility-and-accountability> accessed 29 June 2024.

180 RELE-CIDH, 'La CIDH advierte un punto de inflexión de la libertad de expresión en internet y convoca a diálogo en la región [IACHR warns of a turning point for freedom of expression on the internet and calls for dialogue in the region]' (5 February 2021) <www.oas.org/es/CIDH/jsForm/?File=/es/cidh/prensa/comunicados/2021/026.asp> accessed 29 June 2024.

181 UNESCO, *Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multi-Stakeholder Approach* (2023, UNESCO) 7 <https://unesdoc.unesco.org/ark:/48223/pf0000387339> accessed 22 June 2024.

182 Steven Musil, 'Zuckerberg apologizes for data scandal in full-page ads' (CNET, 26 March 2018) <www.cnet.com/news/zuckerberg-apologizes-for-data-scandal-in-full-page-ads/> accessed 29 June 2024.

183 AP, 'CEO Zuckerberg apologizes for Facebook's privacy failures' (Breitbart, April 10, 2018) <www.breitbart.com/news/ceo-zuckerberg-apologizes-for-facebooks-privacy-failures/> accessed 29 June 2024.

184 Ryan Mac, 'Mark Zuckerberg Sent An Apology Letter About Myanmar. These NGOs Called It "Grossly Insufficient' (Buzzfeed News, April 9, 2018) <www.buzzfeednews.com/article/ryanmac/mark-zuckerbeg-apology-myanmar-ngos-insufficient> accessed 29 June 2024 and Reuters, "Myanmar: UN blames Facebook for spreading hatred of Rohingya", The Guardian, March 13, 2018 <www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya> accessed 29 June 2024.

frameworks on digital platforms.[185] However, decisions on content moderation cannot be made by decree,[186] or through legislative proposals that negatively impact the exercise of human rights in the digital space.[187]

In light of the advancements of GenAI and the distribution of synthetic media, some social media platforms have been moving to create policy frameworks to address any potential impact, ranging from disinformation, Child Sexual Abuse Material (CSAM) to gendered-based violence.[188] For example, Tik Tok has adopted a policy in which they 'welcome the creativity that new artificial intelligence (AI) and other digital technologies may unlock'[189] but acknowledges that 'AI and other digital editing technologies can make it difficult to tell the difference between fact and fiction, which may mislead individuals or harm society'.[190] For that reason, they require that users 'label' AI Generated Content (AIGC) or 'edited media that shows realistic-appearing scenes or people [by] using the AIGC label, or by adding a clear caption, watermark, or sticker of your own'.[191] They define AIGC as 'images, video, or audio, that is created or modified by artificial intelligence (AI) technology or machine-learning processes [and] may include images of real people, and may show highly realistic-appearing scenes or use a particular artistic style, such as a painting, cartoons, or anime'.[192] Their policies also warn that 'even when appropriate-

185 Kate Klonick, 'The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression', Yale Law Journal, Vol. 129, No. 2418, 2020, (June 30, 2020) 2418 SSRN: <https://ssrn.com/abstract=3639234> accessed 29 June 2024.

186 DW, 'Brasil: Jair Bolsonaro prohíbe a redes sociales quitar contenidos [Brazil: Jair Bolsonaro prohibits social networks from removing content]' (7 September 2021) <www.dw.com/es/brasil-jair-bolsonaro-proh%C3%ADbe-a-redes-sociales-quitar-contenidos/a-59105724> accessed 29 June 2024.

187 Daphne Keller, 'Five Big Problems with Canada's Proposed Regulatory Framework for "Harmful Online Content', (Tech Policy Press, August 31, 2021) <https://techpolicy.press/five-big-problems-with-canadas-proposed-regulatory-framework-for-harmful-online-content/?&utm_source=newsletter&utm_medium=email&utm_campaign=automation_and_the_plight_of_the_worker&utm_term=2021-09-05> accessed 29 June 2024; ARTICLE19, "#InternetBajoAtaque: La regulación de las redes sociales como mecanismo de control", Febrero 2021, <https://articulo19.org/wp-content/uploads/2021/02/Article19_2021-PosicionamientoInternet_v3.pdf> accessed 8 July 2024.

188 UNESCO, *"Your Opinion Doesn't Matter, Anyway": Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI* (2023, UNESCO) <https://unesdoc.unesco.org/ark:/48223/pf0000387483> accessed 22 June 2024.

189 TikTok, 'Integrity and Authenticity, Edited Media and AI-Generated Content (AIGC)' (TikTok Community Guidelines, 17 April 2024) <www.tiktok.com/community-guidelines/en/integrity-authenticity#3> accessed 29 June 2024.

190 ibid.

191 ibid.

192 ibid.

ly labelled, AIGC or edited media may still be harmful. That's why they 'do not allow content that shares or shows fake authoritative sources or crisis events, or falsely shows public figures in certain contexts [including] being bullied, making an endorsement, or being endorsed'.[193]

Some significant concerns arise from the wording of TikTok's policy. The term 'bullied', for instance, is a broad concept that could potentially infringe on the protection of the right to freedom of expression. First, this broad and vague concept might be incompatible with the principle underlined in the jurisprudence of international human rights law, where 'a norm must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly'.[194] Second, the right to freedom of expression, which relies on the *ab initio* presumption to avoid *a priori* the exclusion of discourses, 'embraces even expression that may be regarded as deeply offensive'.[195] For example, a political candidate might be demanding the platform to take down content for 'bullying', and the interpretation of this provision by TikTok could be so general to potentially restrict satirical and other forms of content, which are protected by the right to freedom of expression and posing a serious threat to this fundamental right. This does not mean that liability for defamatory or other types of unlawful speech, like statements which incite racial discrimination and hatred (Féret v Belgium, § 78)[196] should not be considered to guarantee an effective remedy for violations of personality rights (Delfi AS v Estonia [GC],§ 110)[197] considering the scope and reach that the internet could pose (Savva Terentyev v Russia,[198] § 79; Delfi AS v Estonia [GC], § 133).[199]

---

193 TikTok, 'Integrity and Authenticity, Edited Media and AI-Generated Content (AIGC)' (TikTok Community Guidelines, 17 April 2024) <www.tiktok.com/community-guidelines/en/integrity-authenticity#3> accessed 29 June 2024.

194 Human Rights Committee, 'Views adopted by the Committee at its 111th session (7–25 July 2014) Communication No 1985/2010 Marina Koktish v Belarus' (3 November 2014) UN Doc CCPR/C/111/D/1985/2010 par 8.5 9 <https://unesdoc.unesco.org/ark:/48223/pf0000387339>; Perinçek v Switzerland [2015] ECHR 27510/08 (15 October 2015) §131.

195 CCPR/C/GC/34 par 11 3.

196 *Féret v Belgium* App no 15615/07 (ECHR, 16 July 2009, Section II).

197 *Delfi AS v Estonia* App no 64569/09 (ECHR, 16 June 2015).

198 *Savva Terentyev v Russia* App no 10692/09 (ECHR 28 August 2018).

199 Delfi AS (n 197).

In the case of X (formerly Twitter) the only reference to AI generated content is contained in the policy and rules section related to 'Child Sexual Abuse' stating that 'X has zero tolerance towards any material that features or promotes child sexual exploitation. This may include real media, text, illustrated, or computer-generated media – including generative AI media'.[200] There are no other references in their policy to content created with AI.

Meta announced changes[201] to its approach to handling manipulated media on its platforms, following feedback from the Oversight Board (OB). The OB recommended that 'Meta must clearly define in a single unified policy the harms it aims to prevent beyond preventing users being misled, such as preventing interference with the right to vote and to take part in the conduct of public affairs'.[202] The revised policy will label a broader range of AI-generated content, including videos, audio, and images, as 'Made with AI' to provide transparency and context. This update, Meta informed, will address the limitations of the previous policy, which only covered AI-altered videos, making people appear to say things they did not say. Meta will now include content showing people doing things they didn't do and will apply labels to high-risk content to prevent public deception. The policy change aims to balance freedom of expression with the need to inform users about manipulated media, and will take effect in May 2024.

In addition to these changes that platforms are implementing, other debates have arisen around users' use of information to train the foundational models of large digital platforms such as Meta, which updated its privacy policy effective June 26, 2024, allowing the company to use data from its platforms, including Facebook, Instagram, Threads, and WhatsApp, to train its AI models.[203] This includes scraping public internet data, even if users are

[200] Delfi AS (n 197).
[201] Monika Bickert (n 164).
[202] Oversight Board, 'Altered Video of President Biden' (Oversight Board, 2023) <www.oversightboard.com/decision/FB-GW8BY1Y3/> accessed 29 June 2024.
[203] Meta, 'How Meta uses information for generative AI models' (Meta, 2024) <www.facebook.com/privacy/genai/> accessed 29 June 2024.

not on Meta's platforms. While Meta states it will review opt-out requests in compliance with relevant data protection laws, users from other regions, like the United States, have limited options due to the lack of comprehensive national data privacy laws.[204]

[204] Melissa Heikkilä, 'How to opt out of Meta's AI training' (MIT Technology Review, 14 June 2024) <www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/> accessed 29 June 2024.

4. The information ecosystem in the context of elections: from 'fake news' to the information disorder

4.1 Disinformation: definitions, impacts, and challenges

We are living in a 'post-truth era,' in which the distinction between 'lies' and 'truths' does not matter. In the words of Ralph Keyes, 'Even though there have always been liars, lies have usually been told with hesitation, a dash of anxiety, a bit of guilt, a little shame, at least some sheepishness. Now, clever people that we are, we have come up with rationales for tampering with truth so we can dissemble guilt-free'.[205]

Various authors have sought to understand the dynamics of information considered as lies, untruth, deceit, or falsehoods. Tandoc and others, for example, reviewed different academic papers and found that the most common term used between 2003 and 2017 was 'fake news'.[206] They categorised the term and proposed a taxonomy of six types of 'fake news' based on previous academic operationalisations: 1) news satire, 2) news parody, 3) fabrication, 4) manipulation, 5) advertising, and 6) propaganda (see Table 1). The authors stated that this categorisation aimed to clarify the term and guide future studies, although they disagreed on whether 'news satire' should be considered 'fake news'.

In their analysis, Tandoc and others placed significant responsibility on the audience, acknowledging that their mistakes can transform 'fake news' into a contender against journalism's legitimacy.[207] However, by attributing such responsibil-

[205] Ralph Keyes, The Post-Truth Era: Dishonesty and Deception in Contemporary Life (Macmillan + ORM 2004) <https://books.google.pl/books?id=f0Kvm3KObXoC&redir_esc=y> accessed 30 June 2024.

[206] Edson C Tandoc Jr and others, 'Defining 'Fake News'' (2018) 6(2) *Digital Journalism* 137, <https://doi.org/10.1080/21670811.2017.1360143> accessed 10 June 2024.

[207] ibid.

ity to the audience, they overlook the roles of other key players, such as states, social media platforms, political parties, and public resources to pay advertising agencies to disseminate disinformation.[208] However, the term 'fake news' has been criticised and found inadequate and problematic. Wardle[209] argues that much of today's misleading content is not necessarily 'fake' or 'news'. Instead, it often involves genuine information used out of context, weaponised to mislead. This content can take various forms, including rumours, memes, manipulated videos, targeted ads, and old photos presented as current. The term 'fake news', she stated, fails to capture this complex reality. Moreover, the author explains that politicians worldwide have co-opted the phrase to attack and discredit legitimate journalism, causing audiences to associate it with reputable news organisations. As a result, the term has become almost meaningless and potentially dangerous. The author suggests that journalists should avoid using 'fake news' in their reporting, to prevent legitimising this unhelpful and increasingly harmful phrase.[210]

In their continued effort to create a more accurate and integral perspective on the phenomenon, Wardle and others conceptualised the phenomenon as an 'information disorder', characterised by the spread of mis-, dis, and mal-information.[211] For them, it poses significant challenges to societies by undermining trust, distorting public discourse, and impacting democratic processes. They also propose an interdisciplinary approach to address these challenges through research and policy-making. For that instance, they proposed three typologies that are part of the information disorder. The first concept is **disinformation** and refers to the 'information that is **false** and **deliberately created to harm** a person, social group, organization or country'(emphasis added).[212] The

[208] Eduardo Buendía and others, 'Neurona: la fábrica de engaño para las izquierdas en América Latina [Neurona: The Deception Factory for the Left in Latin America]' (El Clip, 2023) <www.elclip.org/neurona-la-fabrica-de-engano-para-las-izquierdas-en-america-latina/> accessed 22 June 2024.

[209] Claire Wardle, 'Understanding Information Disorder' (First Draft 2019) <https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76708> accessed 30 June 2024.

[210] ibid.

[211] Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (Council of Europe report DGI(2017)09, 2017) 20 <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 22 June 2024.

[212] ibid.

second is **misinformation**, which refers to the '**information that is false**, but **not created** with the **intention of causing harm**'(emphasis added).[213] The third is '**malinformation**', which refers to the 'information that is **based on reality**, **used to inflict harm** on a person, organisation or country'(emphasis added).[214]

Sullum has raised concerns about the categorisation of 'malinformation' proposed by Wardle and others arguing that the concept, which is described as true but inconvenient information used in ways that could be perceived as harmful, could be used by government and other organisations to suppress such information under the guise of combating misinformation and disinformation.[215] This suppression, says Sullum, often targets truths that challenge official narratives or policies, raising significant free speech issues. Additionally, he suggests that the involvement of government entities in content moderation can lead to indirect censorship, which threatens open debate and democracy.[216]

Wardle also disaggregates a further seven types of information that are part of the 'information disorder': 1) **satire or parody**, with **no intention to cause harm** but has the potential to fool; 2) **false connection** referring when headlines, visuals, or captions don't support the content; 3) **misleading content** when information is used to frame an issue or individual; 4) **false context** when genuine content is shared with false contextual information; 5) **imposter content** when genuine sources are impersonated; 6) **manipulated content** when genuine information or imagery is manipulated to deceive and 7) **fabricated content**, meaning new content that is 100% false, designed to deceive and do harm.[217]

From a more political and electoral perspective, Bennet and Livingston argued that we are living in a 'disinformation age' characterised by a variety of forms of disruptive communication and define disinformation as 'intentional falsehoods or distortions,

---

[213] Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (Council of Europe report DGI(2017)09, 2017) 20 <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 22 June 2024.

[214] ibid.

[215] Jacob Sullum, 'The Crusade Against 'Malinformation' Explicitly Targets Inconvenient Truths' (Reason, 22 March 2023) <https://reason.com/2023/03/22/the-crusade-against-malinformation-explicitly-targets-inconvenient-truths/> accessed 30 June 2024.

[216] ibid.

[217] Wardle (n 211) 11.

often spread as news, to advance political goals such as discrediting opponents, disrupting policy debates, influencing voters, inflaming existing social conflicts, or creating a general backdrop of confusion and informational paralysis'.[218]

Hameleers offered a comprehensive analysis of the conceptualisation of 'disinformation', providing a nuanced understanding of its definition, motivations, and dissemination techniques.[219] The author emphasised the importance of context in studying disinformation and defines it as a 'context-bound **deliberate act** for which actors **covertly deceive recipients** by **de-contextualizing**, **manipulating** or **fabricating information** to maximize utility with the (targeted) outcome of **misleading** recipients'(emphasis added).[220] The author proposed six central pillars to frame the concept: '(1) disinformation is **context-bound**; (2) it is a **deliberate act**; (3) it **aims to deceive** in a covert manner; (4) it **involves different techniques** of altering information; (5) it is directed at **maximizing** personal or organizational **profit or gain**; and (6) its **targeted outcome** is to **mislead recipients**'(emphasis added).[221] Additionally, he argues that just like it is hard to figure out someone's intentions, not all harmful information clearly shows lies. Sometimes, as Wardle and others have conceptualised as 'malinformation', true information can be used to mislead. Hameleers encompassed the term in disinformation, considering it changing, taking out of context, or making up information. That is why, for the author, focusing on the intentions and context makes it possible to spot harmful campaigns that use true facts in misleading ways. He recalls that it is less about whether something is true or false and more about how it is being used to deceive.[222]

In the Joint Declaration On Freedom Of Expression And "Fake News", Disinformation and Propaganda, the regional representatives and special rapporteurs on freedom of expression, state that disinformation and propaganda are often 'designed

[218] W Lance Bennett and Steven Livingston (eds), 'The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States' (2021) Cambridge University Press 3 <https://doi.org/10.1017/9781108914628>.
[219] Michael Hameleers, 'Disinformation as a Context-Bound Phenomenon: Toward a Conceptual Clarification Integrating Actors, Intentions and Techniques of Creation and Dissemination' (2022) <https://doi.org/10.1093/ct/qtac021>.
[220] ibid 6.
[221] ibid 6.
[222] ibid 8.

and implemented **to mislead** a population, as well as to **interfere** with the public's **right to know** and the **right** of individuals **to seek and receive**, as well as to **impart, information and ideas**'(emphasis added).[223]

The Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights in the *Guide to Guarantee Freedom of Expression Regarding Deliberate Disinformation in Electoral Contexts* defined disinformation as 'the **mass dissemination** of **false information** (a) with the **intent to deceive the public** and (b) **with the knowledge of its falsehood**'(emphasis added).[224] They also noticed that in the context of elections, disinformation raises particularly concern because, if successful, these efforts could undermine the legitimacy of the processes, which are fundamental to the operation and survival of democratic societies.[225]

From a different perspective but with the same care on the potential impact of disinformation, the UN's 'Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms' introduced the term 'information integrity', defined as the accuracy, consistency, and reliability of information, which is endangered by disinformation, misinformation, and hate speech.[226] The brief highlights the significant impact of these threats on health, climate action, democracy, and human rights, noting the role of digital platforms in their spread. It proposes a United Nations Code of Conduct to enhance information integrity through increased transparency, support for independent media, user empowerment,

[223] United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Cooperation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. Joint declaration on freedom of expression and 'fake news', disinformation and propaganda. (2017). OSCE. <www.osce.org/fom/302796> accessed 30 June 2024.

[224] Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, Department of Electoral Cooperation and Observation, and Department of International Law of the General Secretariat of the Organization of American States, *Guide to Guarantee Freedom of Expression Regarding Deliberate Disinformation in Electoral Contexts* (OAS Official Records OEA/Ser.D/XV.22, OEA/Ser.G CP/CAJP/INF.652/19, 2019) 13 <www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf> accessed 8 July 2024.

[225] ibid 13.

[226] United Nations, 'Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms' (June 2023) 5 <www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf> accessed 10 July 2024.

strengthened research, and robust regulatory frameworks. The brief underscores the importance of international norms and multi-stakeholder involvement to protect information integrity and support sustainable development goals.

Finally, the SRFOE has defined **disinformation** 'as false information that is disseminated intentionally to cause **serious social harm** and **misinformation** as the **dissemination** of **false information unknowingly**' (emphasis added) and reminded us that the terms are not used interchangeably.[227]

In conclusion, the concept of 'disinformation' has evolved significantly in recent years, moving beyond simplistic notions of 'fake news' to grasp a clearer understanding of the 'information disorder' and the different forms of disruptive communication. Scholars like Wardle, Tandoc, and Hameleers have contributed to a deeper comprehension to understand the concept of disinformation in the digital realm. The complexity of this phenomenon is evident in its various forms, ranging from satire and parody to deliberate fabrication and manipulation of content. While there is general agreement on the influence of disinformation on public discourse and democratic processes, debates persist regarding the appropriate categorisation, the level of the impact[228] and response to these challenges. Concerns have been raised about potential threats to free speech and the risk of censorship, particularly when the state tries to influence content moderation or create restrictive regulation.[229] As we navigate this 'post-truth era', it becomes increasingly crucial to balance efforts to combat the information intended to mislead a population, as well as to interfere with the public's right to know and the right of individuals to seek and receive, as well as to impart, information and ideas with the protection of legitimate expression and open debate.

---

[227] A/HRC/47/25 par 15.

[228] According to Valenzuela and others, there is 'no significant correlation between using Facebook, Twitter, YouTube, Instagram or WhatsApp as news sources and belief in political misinformation' Sebastián Valenzuela and others, 'Social Media and Belief in Misinformation in Mexico: A Case of Maximal Panic, Minimal Effects?' (2022) 29(3) The International Journal of Press/Politics 8 <https://doi.org/10.1177/19401612221088988>.

[229] ARTICLE 19, 'Malaysia: Repeal "Fake News" Emergency Ordinance' (ARTICLE 19, 15 March 2021) <www.article19.org/resources/malaysia-fake-news-ordinance/> accessed 8 July 2024 and ARTICLE 19, 'Senegal: "Fake news" and disinformation laws threaten freedom of expression' (ARTICLE 19, 17 January 2024) <www.article19.org/resources/senegal-fake-news-and-disinformation-laws-threaten-freedom-of-expression/> accessed 8 July 2024.

4.2 Perspectives on electoral integrity: theoretical frameworks and practical challenges

There is no agreed definition of 'election integrity' in international law. However, as we have analysed in previous lines, several human rights instruments have set the principles for the behaviour of different actors participating in elections, guaranteeing state protection of people's rights to participate in public affairs and their right to freedom of expression.

The report from the Global Commission on Elections, Democracy, and Security, created through a joint initiative of the Kofi Annan Foundation and the International Institute for Democracy and Electoral Assistance (International IDEA), titled 'Deepening Democracy: A Strategy for Improving Electoral Integrity Worldwide', defines the integrity of an election as one 'based on the democratic principles of universal suffrage and political equality as reflected in international agreements and standards and characterised by professional, impartial, and transparent preparation and management throughout the electoral cycle'. [230]

Similarly, Pippa Norris describes 'electoral integrity' as reflecting global norms applying to all countries worldwide throughout the electoral cycle, including during the pre-electoral period, the campaign, polling day, and its aftermath.[231]

Bail and others discuss how dynamics within the electoral process influence perceptions of an election's fairness and legitimacy. The authors argue that understanding these dynamics is critical for theoretical insights into democratic representation and practical policymaking, given that election dynamics can vary significantly over time. They emphasise that a 'better understanding

[230] Comisión Global sobre Elecciones, Democracia y Seguridad, 'Profundizando la democracia: Una estrategia para mejorar la integridad electoral en el mundo [Deepening Democracy: A Strategy to Improve Electoral Integrity Worldwide]' (Septiembre 2012) <www.idea.int/sites/default/files/publications/profundizando-la-democracia.pdf> accessed 12 June 2024.

[231] Pippa Norris, 'Are There Global Norms and Universal Standards of Electoral Integrity and Malpractice? Comparing Public and Expert Perceptions' (Harvard University, John F. Kennedy School of Government, Faculty Research Working Paper Series RWP12-010, 2012) <https://dash.harvard.edu/bitstream/handle/1/8506826/RWP12-010_Norris.pdf> accessed 12 June 2024.

of how election dynamics shape perceptions of election integrity is crucial because this process is at the heart of democratic representation and because these dynamics vary more over time than individual and state-level factors'.[232]

Garnett and others identify three principles[233] of electoral integrity in the digital age: 1) *deliberative opportunities*, allowing open participation in election debates and access to information; 2) *equality of participation*, emphasising political equality to address turnout gaps and promote inclusive voting practices with the help of technology; and 3) *robust electoral management*, ensuring well-run elections with convenience, service quality, transparency, professionalism, fairness, cost-effectiveness, and stakeholder satisfaction. While these principles are essential for achieving democratic goals, the authors also recognise the challenges posed by disinformation and interventions aimed at misleading voters during the electoral period.

One significant risk of restrictive electoral laws is the potential suppression of critical voices. Simiyu notes that creating electoral laws to combat disinformation is also a concern, as measures taken by governments often infringe upon freedom of expression. She observes how 'African governments are increasingly enacting laws that criminalize false news or adopting practices such as internet shutdowns as strategies to address the spread of online false news during elections'.[234] While intended to protect the electoral process, these measures adversely affect citizens' ability to exercise their freedom of expression and access necessary information.

Judge and others also agree with Simiyu that, in order to protect electoral integrity from the damaging effects of disinformation, countries, particularly democracies, must adopt a balanced regulatory approach that does not unduly infringe on free speech,

---

[232] Christopher A Bail and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization' (2018) 115(37) Proceedings of the National Academy of Sciences 9216 <https://doi.org/10.1073/pnas.1804840115>.

[233] Holly Ann Garnett and Toby S James, 'Cyber Elections in the Digital Age: Threats and Opportunities of Technology for Electoral Integrity' (2020) 19(2) *Election Law Journal* <www.liebertpub.com/doi/full/10.1089/elj.2020.0633>.

[234] Marystella Auma Simiyu, 'Freedom of Expression and African Elections: Mitigating the Insidious Effect of Emerging Approaches to Addressing the False News Threat' (2022) 22(1) African Human Rights Law Journal <https://hdl.handle.net/10520/ejc-ju_ahrlj_v22_n1_a5> accessed 13 July 2024.

but effectively manages the dissemination of disinformation. They advocate for the principle of 'digital information equality', which aims to ensure voters have access to quality information, thereby maintaining fairness and trust in the electoral process.[235]

The online disinformation in the context of elections is not new.[236] These operations encompass a range of tactics and actors, from state-sponsored campaigns on social media platforms to the propagation of false narratives by political leaders. The US Senate Intelligence Committee reported that the Internet Research Agency[237] (IRA) spent millions of dollars to 'influence the 2016 U.S. election as part of a broader information campaign to harm the United States and fracture its society'.[238]

The 2016 election in the United States sparked a global debate about the manipulation that can be performed in the information ecosystem using information and communication technologies (ICT) and social media platforms to influence the results of an election. A strategy operated in the digital realm with the aim of distorting the information and feeding the echo chambers. Different terms started to dominate the international forum. In the academic field, civil society organisations and political actors discussed different terms to categorise this phenomenon. Some voices, as will be described, considered that disinformation campaigns were pointed out as a force that can manipulate voters and affect the integrity of an election. To others, there was not

[235] Elizabeth F Judge and Amir M Korhani, 'Disinformation, Digital Information Equality, and Electoral Integrity' (Forthcoming in Election Law Journal, 24 February 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518800> accessed 13 July 2024.

[236] W Lance Bennett (n 218) 167.

[237] According to the Department of Justice of the United States, the Internet Research Agency (IRA) was a Russian organisation, allegedly conducted extensive political interference operations in the United States from 2014 onwards. Using purchased American server space, hundreds of fictitious online personas, and stolen U.S. identities, the IRA operated social media accounts to sow discord in the American political system. Their activities included supporting and disparaging specific candidates, buying political ads, and attempting to coordinate with US activists. The IRA's operations reached millions of Americans, aiming to interfere with the U.S. political system, particularly the 2016 Presidential Election. These efforts were designed to exploit social divisions and influence public opinion through targeted misinformation campaigns across various online platforms. US Department of Justice, *Report On The Investigation Into Russian Interference In The 2016 Presidential Election* Volume I of II (Special Counsel Robert S Mueller, III, Washington, DC, March 2019) 14-19 <www.justice.gov/d9/report.pdf> accessed 8 July 2024.

[238] US Senate Select Committee on Intelligence, 'Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume II: Russia's Use of Social Media with Additional Views' (2020) <www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf> accessed 12 June 2024.

enough evidence to establish a direct link in the effects it could pose to people exposed to disinformation. It was clear that this phenomenon was not new and had deep historical roots, but it has intensified with the expansion of the internet and social media platforms.[239]

In contemporary settings, disinformation has been exacerbated by technological advancements and the proliferation of social media platforms.[240] These platforms have created an environment where information can be disseminated rapidly and widely regardless of its veracity. According to a study carried out by Vosoughi and others, 'false news spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it'.[241] While for Bennet and others, social media platforms also 'exacerbate social tensions by algorithmically amplifying extremist content as a way to maximize advertising revenue'.[242] The intentional spread of falsehoods by various actors, including political leaders, social media influencers, and foreign governments, has become a tool for undermining democratic institutions and promoting divisive agendas.

The well-known Cambridge Analytica scandal highlights the potential for data misuse in electoral contexts and the use of digital technologies to manipulate people.[243] This operation involved the unauthorised harvesting of personal data from millions of Facebook users, subsequently employed for targeted political advertising during the 2016 US presidential election and the UK's

[239] W Lance Bennett (n 218) 167.
[240] ibid 280.
[241] S Vosoughi, D Roy, and S Aral, 'The Spread of True and False News Online' (2018) 359 Science 1146. Available at <www.science.org/doi/10.1126/science.aap9559>.
[242] W Lance Bennett (n 218) 280.
[243] Joe Westby, '"The Great Hack": Cambridge Analytica is just the tip of the iceberg' (Amnesty International, 24 July 2019) <www.amnesty.org/en/latest/news/2019/07/the-great-hack-facebook-cambridge-analytica/> accessed 12 June 2024.

Brexit referendum.[244] Although an investigation by the ICO in the UK found no evidence of misusing personal data impacting Brexit,[245] other disinformation strategies were used to influence people.[246]

According to a report from the House of Commons, Aggregate IQ, a Canadian digital advertising and software development company with close ties to Cambridge Analytica deployed a sophisticated system during the Brexit campaign. This system utilised three machine learning tools to analyse online text and images, match photos from websites to Facebook profiles, and target ads to specific users in an attempt to influence their decisions.[247] The case demonstrates the application of AI for targeted user influence and highlights advanced data processing capabilities. However, it also raises significant privacy concerns due to its cross-platform data collection and targeting methods, which were implemented without explicit user consent.

In Mexico, Monsiváis-Carrillo notes that disinformation has evolved beyond the activities of organised groups affiliated with specific political figures to include the dissemination of disinformation by President Andrés Manuel López Obrador (AMLO) himself.[248] The author argues that AMLO has crafted a narrative claiming electoral authorities have been complicit in electoral fraud, election costs are excessive, and his government seeks to establish genuine democracy. This rhetoric, the author argues, has been used to justify reforms that could undermine electoral integrity. In the same line, a report from the civil society organisation, Article 19, found that AMLO spreads misleading information

[244] Carole Cadwalladr and Emma Graham-Harrison, 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach' (The Guardian, 17 March 2018) <www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> accessed 12 June 2024.

[245] Izabella Kaminska, 'Cambridge Analytica Probe Finds No Evidence it Misused Data to Influence Brexit' (Financial Times, 7 October 2020) <www.ft.com/content/aa235c45-76fb-46fd-83da-0bdf0946de2d> accessed 22 June 2024.

[246] House of Commons Digital, Culture, Media and Sport Committee, Disinformation and 'Fake News': Final Report: Government Response to the Committee's Eighth Report of Session 2017–19 Seventh Special Report of Session 2017–19 (Ordered by the House of Commons to be printed 8 May 2019) <https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/2184/2184.pdf> accessed 8 July 2024.

[247] ibid 56.

[248] Alejandro Monsiváis-Carrillo, 'Populismo, desinformación e integridad electoral en México' (2023) 22(25) Revista Elecciones 151 <https://dx.doi.org/10.53557/elecciones.2023.v22n25.05>.

and disinformation in his morning press conferences. Article 19 underscored the detrimental impact of disinformation from government figures on public trust and the right to freedom of expression.[249]

What has been well documented and studied is that 'interference and manipulation through disinformation disseminated via social networks and powered through AI during electoral processes greatly impact on the right to vote freely'.[250] While Benkler and others, argue that microtargeting and the strategic use of digital platforms have significantly influenced political campaigns and voter manipulation, highlighting the role of big data and algorithm-driven strategies in modern political strategies.[251] However, Valenzuela and others are more cautious about the impact of disinformation. Contrary to other narratives, they argued that the use of social media for consuming news does not significantly increase belief in political disinformation.[252]

Finally, the Joint Declaration on Freedom of Expression and Elections in The Digital Age addresses the challenges and opportunities for freedom of expression and elections in the digital age. It emphasises the importance of free, independent, and diverse media in ensuring fair elections, while acknowledging the challenges posed by digital technologies. The declaration provides recommendations for states, digital actors, and other stakeholders to protect freedom of expression during elections. Key points include promoting media literacy, ensuring transparency in political advertising, protecting journalists from attacks, avoiding censorship, and addressing disinformation without unduly restricting free speech. The document also stresses the need for digital platforms to respect human rights principles and implement measures to promote diverse political viewpoints.[253]

---

[249] ARTICLE 19, '(Des)información oficial y comunicación social' [(Official Disinformation and Social Communication)] (Artículo 19, 14 March 2023) <https://articulo19.org/desinformacion-oficial-y-comunicacion-social/> accessed 8 July 2024.

[250] UNESCO, *Elections in Digital Times: A Guide for Electoral Practitioners* (UNESCO 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000387339> accessed 22 June 2024.

[251] Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press 2018).

[252] Valenzuela (n 228).

[253] United Nations Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media and OAS Special Rapporteur on Freedom of Expression, 'Joint Declaration on Freedom of Expression and Elections in the Digital Age' (30 April 2020) <www.osce.org/files/f/documents/9/8/451150_0.pdf> accessed 30 June 2024.

4.3 Debating the impact of disinformation: are fears of GenAI exaggerated?

In 2023, a Belgian man committed suicide after an intensive interaction for six weeks with *Eliza,* a chatbot that relies on Open AI ChatGPT. According to the reports from *La Libre,*[254] he was depressed and anxious about the climate crisis and developed a strong emotional dependence on the chatbot. During their last interaction, Eliza asked him: 'If you wanted to die, why didn't you do it sooner?'. The man answered he probably wasn't ready. The chatbot insisted: 'Had you ever been suicidal before?' The man's answer sheds light on the situation: 'Once, after receiving what I took to be a sign from you'. Although it is hard to establish a connection between LLM chatbots, a derivative form of Generative Artificial Intelligence (GenAI), and his death, it opened a debate on the implications of this emerging technology for its manipulative potential as it gives 'the illusion that something human is behind the screen'.[255]

While this AI can bring wide-ranging benefits to society,[256] they can also create a dire scenario to manipulate and potentially destabilise democratic processes, leading to an 'existential risk of an irreversible totalitarian regime'.[257] Firstly, Large Language Models (LLMs) and Generative AI (GenAI) **could induce people to lose their capacity to discern what is real.** In 2022, Kreps and others conducted to observe the effects on the perception of persons when exposed to news stories created with an LLM. They used a trustworthy story from a legendary media outlet as a baseline. They found that 'individuals are largely incapable of distinguishing

---

[254] Pierre-François Lovens, 'Sans ces Conversations avec le Chatbot Eliza, Mon Mari Serait Toujours Là' (La Libre Belgique, 28 March 2023, 6:35 am, updated 7:06 am) <www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPDST24/> accessed 13 July 2024.

[255] Will Bedingfield, 'A Chatbot Encouraged Him to Kill the Queen. It's Just the Beginning'
(WIRED, 18 October 2023)
<www.wired.com/story/chatbot-kill-the-queen-eliza-effect/?redirectURL=%2Fstory%2Fchatbot-kill-the-queen-eliza-effect%2F> accessed 30 June 2024.

[256] Tor Kielland, 'Embracing AI in Journalism — The News Carousel' (Medium, 8 November 2023) <https://pub.towardsai.net/embracing-ai-in-journalism-the-news-carousel-dd6b170ce376> accessed 5 May 2024.

[257] Dan Hendrycks, Mantas Mazeika, and Thomas Woodside, 'An Overview of Catastrophic AI Risks' (arXiv.org, 21 June 2023) <https://arxiv.org/abs/2306.12001> accessed 5 May 2024.

between AI- and human-generated text'[258] and identified that 'these tools are capable of producing news content that readers deem as equally or more credible than human-written news stories'.[259] Furthermore, they pointed to potential consequences of creating more noise in a media landscape already marked by high volumes of disinformation and less trust in media.

LLMs and GenAI can also create realistic synthetic media, such as deepfakes, which have been used for harassment. These capabilities extend to mimicking speech,[260] movement and writing, raising concerns about their potential to intimidate or blackmail individuals by impersonating them in many ways.[261] This was the case of Rana Ayyub, an Indian investigative journalist who was forced to retire from public life after being targeted by a deepfake showcasing her as if involved in a porn scene.

This becomes particularly complex with the current anthropomorphising trend of AI: LLMs are constantly improving and captivating, with their human-mimicking capabilities to create texts, respond to various questions, and interact as if dialoguing with someone who has the answers to almost everything that can be asked.[262] When chatbots mimic human interaction, they might present misleading emotional responses and potentially harmful suggestions without true understanding or empathy. For example, this year, an eating disorder chatbot was suspended in the US for giving harmful responses to people.[263]

---

[258] S Kreps, RM McCain, and M Brundage, 'All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation' (2022) 9 Journal of Experimental Political Science 104, 117 <https://doi.org/10.1017/XPS.2020.37>.

[259] ibid.

[260] Arijeta Lajka, 'New AI Voice-Cloning Tools "Add Fuel" to Misinformation Fire' (AP News, 11 February 2023) <https://apnews.com/article/technology-science-fires-artificial-intelligence-misinformation-26cabd20dcacbd68c8f38610fec39f5b> accessed 5 May 2024.

[261] Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (2021) 137 <https://arxiv.org/pdf/2108.07258.pdf> accessed 5 June 2023.

[262] Bernd Carsten Stahl and Damian Eke, 'The ethics of ChatGPT – Exploring the ethical issues of an emerging technology' (2024) 74 International Journal of Information Management 102700 <https://doi.org/10.1016/j.ijinfomgt.2023.102700>.

[263] Amanda Hoover, 'An Eating Disorder Chatbot Is Suspended for Giving Harmful Advice' (Wired, 1 June 2023) <www.wired.com/story/tessa-chatbot-suspended/> accessed 12 May 2024.

Secondly, according to Kreps the potential misuse of LLMs and GenAI to massively spread disinformation is enormous.[264] As we have seen, it is relevant because disinformation could distort and 'pollute'[265] the information landscape, potentially affecting election integrity, eroding civic participation and the right to access information.[266]

Manipulation to interfere with people's right to know and the dissemination of disinformation are not new. Different studies have analysed the spread of disinformation through social media and its impact on democracy through what has been called 'algorithmic amplification'[267] and 'traditional *brute force* astroturfing campaigns and novel *mimicking conversation* tactic[s]'.[268] That is why, for some authors, this is just another type of technology which does not represent any major threat.[269]

As has been argued, the emergence of Large Language Models (LLMs) like ChatGPT marks a turning point in artificial intelligence, showcasing an ability to interact in human-like conversation and posing risks of manipulation and disinformation, particularly in electoral contexts, but also at a personal level. The suicide of the Belgian man can be interpreted as a warning sign to put a human rights perspective aiming at protecting people from any 'high risk' of AI at the centre of the discussion, as has been put in place in the first worldwide regulation for Artificial Intelligence, the EU AI Act.[270]

---

[264] Sarah Kreps and Doug Kriner, 'How AI Threatens Democracy' (2023) 34(4) Journal of Democracy 122 <www.journalofdemocracy.org/articles/how-ai-threatens-democracy/> accessed 12 May 2024.

[265] Katarina Kertysova, 'Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered' (2018) 29(1-4) *Security and Human Rights* 55 <https://doi.org/10.1163/18750230-02901005>.

[266] Kreps and others (n 258).

[267] Benjamin Laufer and Helen Nissenbaum, 'Algorithmic Displacement of Social Trust' (Knight First Amendment Institute, 29 November 2023) <https://knightcolumbia.org/content/algorithmic-displacement-of-social-trust> accessed 12 May 2024.

[268] Armando Espinoza and Carlos A Piña-García, 'Propaganda and Manipulation in Mexico: A Programmed, Coordinated and Manipulative "Pink" Campaign' (2023) 4(2) Journalism and Media 578 <https://doi.org/10.3390/journalmedia4020037>.

[269] Alex Tamkin, Miles Brundage, Jack Clark, and Deep Ganguli, 'Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models' (2021) arXiv preprint arXiv:2102.02503 <https://arxiv.org/abs/2102.02503> accessed 12 May 2024.

[270] Melissa Heikkilä. 'Five Things You Need to Know About the EU's New AI Act' (MIT Technology Review, 11 December 2023) <www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/> accessed 12 July 2024.

Goldstein and others discuss how Generative AI could transform and influence disinformation operations using the 'ABC' framework (Actors, Behaviour, and Content). They describe how language models might expand the number and diversity of actors able to conduct influence operations by lowering costs, enabling new tactics like 'dynamic, personalized, and real-time content generation,' and making existing behaviours more efficient.[271] The content generated may be 'more credible and persuasive' and 'less discoverable' compared to current campaigns. In this context, Generative AI could impact influence operations in different ways: 1) Lowering the cost of generating propaganda, allowing more actors to wage influence campaigns and easier to scale up;[272] 2) Enabling new tactics like dynamic, personalised, real-time content generation[273] (eg chatbots); 3) Producing more credible and persuasive content that is less detectable as artificial.[274] One example is how Generative AI continues its expansion as a powerful instrument to generate synthetic media faster[275] and 'trigger the next misinformation nightmare'.[276]

This can be done with the use of Large Language Models (LLMs) and other generative AI tools to massively fabricate synthetic media, text messages, videos, or audio. Goldstein and others conducted an experiment in 2021 by using propaganda examples

---

[271] Josh A Goldstein and others, 'How persuasive is AI-generated propaganda?' (2024) 3(2) PNAS Nexus pgae034 <https://doi.org/10.1093/pnasnexus/pgae034>.

[272] ibid.

[273] Zak Rogoff, 'Generative AI is Already Catalyzing Disinformation. How Long Until Chatbots Manipulate Us Directly?' (Tech Policy Press, 23 October 2023) <https://techpolicy.press/generative-ai-is-already-catalyzing-disinformation-how-long-until-chatbots-manipulate-us-directly> accessed 8 July 2024.

[274] Sarah Kreps, R Miles McCain, and Miles Brundage, 'All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation' (2022) 9(1) Journal of Experimental Political Science 104 <https://doi.org/10.1017/XPS.2020.37> and Goldstein and others (n 262).

[275] Advancements in deep learning (DL) rapidly started transforming the approach to media creation across communication, entertainment, and artistic domains. A particularly notable example of this progress was 'deepfakes', a term that combines 'deep learning' (DL) and 'fake'. This concept, explained Millière, emerged in 2017, originating from a Reddit user who devised a DL-based method to substitute an actor's face with that of a celebrity in pornographic videos. Since its inception, 'deepfake' has been broadly applied to videos where faces are digitally altered using DL algorithms and, more generally, to any DL-based manipulation of sound, image, and video. These technologies are significantly altering media creation, showcasing remarkable capabilities as well as posing substantial potential risks. Raphaël Millière, 'Deep Learning and Synthetic Media' (2022) 200 *Synthese* 231 <https://doi.org/10.1007/s11229-022-03739-2>.

[276] Ashley Gold and Sara Fischer, 'Chatbots trigger next misinformation nightmare' (Axios, 21 February 2023) <www.axios.com/2023/02/21/chatbots-misinformation-nightmare-chatgpt-ai> accessed 12 July 2024.

that were previously identified by investigative journalists. They asked GPT-3 to create new articles based on them and presented them to a group of people. The results showed that GPT-3-generated propaganda was highly persuasives and sometimes as effective as human-generated content. Almost half of the respondents agreed or strongly agreed with the thesis statement. In their view, 'this suggests that propagandists could use GPT-3 to generate persuasive articles with minimal human effort, by using existing articles on unrelated topics to guide GPT-3 about the style and length of new articles'.[277]

That is why they considered that generative models required robust detection methods to combat their potential misuse as they could automate disinformation.[278] For that purpose, they identified different *Transformer-based detection algorithms* that can help distinguish between human-generated and computer-generated texts, emphasising the importance of developing robust methods to combat digital disinformation effectively.[279]

Conversely, Simon and others argued that the 'current concerns about the effects of generative AI on the misinformation landscape are overblown'.[280] They assess arguments from communication studies, cognitive science, and political science, concluding that the current concerns about generative AI's role in increasing the quantity, quality, and personalisation of misinformation are speculative and unsupported by evidence. The article suggests that while generative AI may facilitate the creation of misinformation, the actual impact on misinformation diffusion and public belief is likely to be limited due to existing media consumption patterns and the public's selective trust in information sources. The authors call for a more nuanced, evidence-based discussion

[277] Goldstein and others, (n 271) 3.
[278] ibid 5.
[279] Harald Stiff and Fredrik Johansson, 'Detecting Computer-Generated Disinformation' (2021) 13 International Journal of Data Science and Analytics 363 <https://doi.org/10.1007/s41060-021-00299-5>.
[280] Felix M Simon, Sacha Altay, and Hugo Mercier, 'Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown' (HKS Misinformation Review, 2023) <https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/> accessed 12 July 2024.

on the issue, emphasising the importance of strengthening existing media and information institutions. Instead, they emphasise the importance of strengthening current media and information institutions to mitigate any potential negative effects of generative AI.

5. Case Study: Assessing the implications of GenAI for the 2024 general elections in Mexico

The 2024 electoral process was considered to be the largest[281] in Mexico's history due to the number of positions to be elected. With a population of 126 million people,[282] the country has a nominal list (Lista Nominal, LN) of 96 million voters from an Electoral Registry (Padrón Electoral, PN) of 98 million registrations.[283] The difference between these lists is that the PN contains the basic information of the Mexican population that applied for a voting credential, while the LN is the list of inhabitants with a valid voting credential who can cast their vote on election day.[284]

[281] Megan Janetsky, 'Mexico is About to Have Its Biggest Election Ever. Here's What to Know' (Associated Press, 1 March 2024) <www.apnews.com/article/mexico-elections-2024-what-to-know-d104184b02bf5bcf9e08f570a5ba37e2> accessed 4 June 2024.

[282] Instituto Nacional de Estadística y Geografía (INEGI), 'INEGI [National Institute of Statistics and Geography]' <www.inegi.org.mx/default.html> accessed 28 June 2024.

[283] Instituto Nacional Electoral (INE), 'Estadísticas de la Lista Nominal y Padrón Electoral [Statistics of the Nominal List and Electoral Register]' <https://portal.ine.mx/credencial/estadisticas-lista-nominal-padron-electoral/> accessed 20 May 2024.

[284] Instituto Nacional Electoral (INE), 'Padrón Electoral y Lista Nominal de Electores [Electoral Register and Nominal List of Voters]' <https://ine.mx/padron-electoral-lista-nominal-electores/> accessed 20 May 2024.

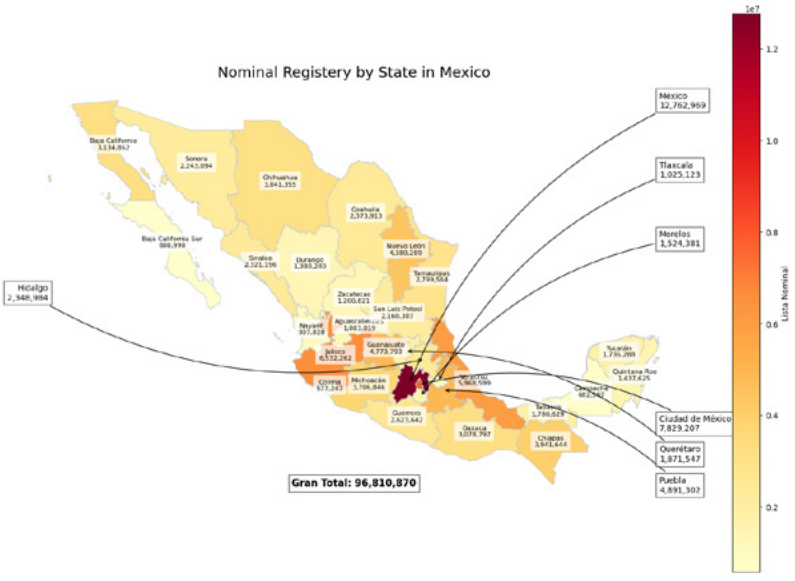

Figure 2. Nominal Registery by State in Mexico
Source: Own elaboration with data from the INE

According to the National Electoral Institute[285] (INE), the 2024 Mexican election involved over 20,000 public positions across federal, state, and local levels. At the federal level, this included the Presidency, 128 Senate seats, and 500 Federal Deputies. State and local positions comprised 8 Governorships, one Head of Government for Mexico City, 1,098 Local Deputies, and numerous municipal positions including 1,802 Municipal Presidencies (mayors), 1,975 Syndics, and 14,560 Councillors. This extensive network of elected positions demonstrates Mexico's complex representative system of democracy at all levels of government.

At the municipal level, Mexico's local governance structure is reflected in 1,802 Municipal Presidencies (mayors), supported by 1,975 Syndics, 431 auxiliary positions, and 14,560 Councillors.[286] This extensive local representation illustrates the country's

[285] Instituto Nacional Electoral (INE), *Numeralia del Proceso Electoral Federal 2023-2024* (2023) <https://repositoriodocumental.ine.mx/xmlui/bitstream/handle/123456789/153578/Numeralia-PEF-2023-2024.pdf> accessed 22 June 2024.
[286] ibid.

transition from an authoritarian system to a democratic one, as noted by Valdés Zurita: 'A political pluralism was installed in society... it went from a hegemonic party system to a plural and competitive one'.[287]

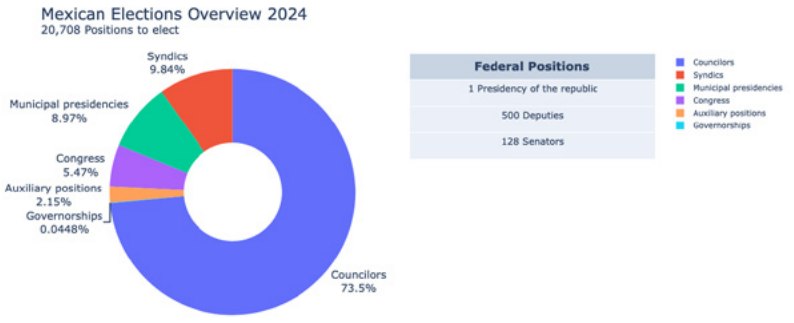

Figure 3. Mexican Election Overview 2024
Source: Own elaboration with data from the INE

This extensive network of elected positions demonstrates Mexico's complex representative democracy at all levels.[288] The system is characterised by 'juxtaposed governments',[289] where territorial units with varying degrees of power coexist. This geo-electoral complexity,[290] combined with the country's geographical characteristics and digital divide, impacts the spread of (dis)

[287] Leonardo Valdés Zurita, 'El sistema electoral mexicano: equidad en la competencia, inequidad en la representación [The Mexican Electoral System: Fairness in Competition, Inequity in Representation]' (2021) 20(21) Elecciones 15-42 16.

[288] Ignacio Daniel Torres Rodríguez and Carlos Enrique Ahuactzin Martínez, 'Democracy and Electoral Reforms in Mexico' (2019) 4(11) Derecho Global. Estudios sobre Derecho y Justicia 143-162 158 <http://www.derechoglobal.cucsh.udg.mx/index.php/DG/article/view/186/245> accessed 22 June 2024.

[289] Alain de Remes, 'Gobiernos yuxtapuestos en México' [2017] 483-Texto del artículo-2082-1-10-20170411 246 <https://revistas.onpe.gob.pe/index.php/elecciones/article/view/110/104> accessed 12 June 2024.

[290] M David Álvarez Hernández, Shaní Eneida Álvarez Hernández and Miguel Álvarez Texocotitla, 'La complejidad del sistema geoelectoral mexicano a nivel municipal [The Complexity of the Mexican Geoelectoral System at the Municipal Level]' (2022) 66 Revista Mexicana de Ciencias Políticas y Sociales 167 <https://dialnet.unirioja.es/servlet/articulo?codigo=8434447> accessed 20 May 2024.

information. As of 2023, whilst 97 million people (81.2% of the population aged six and over) were connected to the internet, approximately 22.46 million remained unconnected, creating a nuanced landscape for information dissemination.[291]

5.1 The Digital Landscape in Mexico

Whilst there is no direct correlation between internet access and disinformation spread in Mexico's electoral process, notable disparities exist. Rural areas,[292] comprising 20% of the population, have only 66% internet users compared to 85.2% in urban areas.[293] Similarly, only 39.5% of lower socio-economic households with ICT equipment are connected, versus 93.5% in higher strata.[294]

The Reuters Institute's Digital News Report indicates that 80% of Mexican respondents use online platforms as primary news sources,[295] with Facebook (56%) and YouTube (39%) being prominent.[296] However, trust in news remains low, with only 36% trusting news in general and 41% trusting personally consumed news.

Internet adoption across socio-economic strata presents an intriguing perspective on potential disinformation diffusion. In lower strata, economic constraints limit internet adoption (around 50% in low and lower-middle strata), potentially creating a barrier to online disinformation campaigns. Conversely, in higher strata, 50.9% of non-adopters cite lack of interest, possibly making them less susceptible to digital disinformation.

---

[291] Instituto Nacional de Estadística y Geografía (INEGI), 'Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares 2023 [National Survey on the Availability and Use of Information Technologies in Households 2023]' (INEGI, 2024) [Hereinafter ENDUTIH] <www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENDUTIH/ENDUTIH_23.pdf> accessed 28 June 2024.

[292] Rural Areas in Mexico are defined by less than 2500 inhabitants. Isidro Soloaga and others, 'Lo rural y lo urbano en México Una nueva caracterización a partir de estadísticas nacionales [The rural and the urban in Mexico A new characterisation based on from national statistics]' Economic Commission for Latin America and the Caribbean (ECLAC, 2022) 14 <https://repositorio.cepal.org/server/api/core/bitstreams/27f4bef7-e9f0-4d61-8baa-7bd1fdc26675/content> accessed 30 June 2024.

[293] INEGI (n 291).

[294] ibid.

[295] Nic Newman and others, *Digital News Report 2023* (Reuters Institute for the Study of Journalism, 2023) 121 <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf> accessed 8 July 2024.

[296] ibid 121.

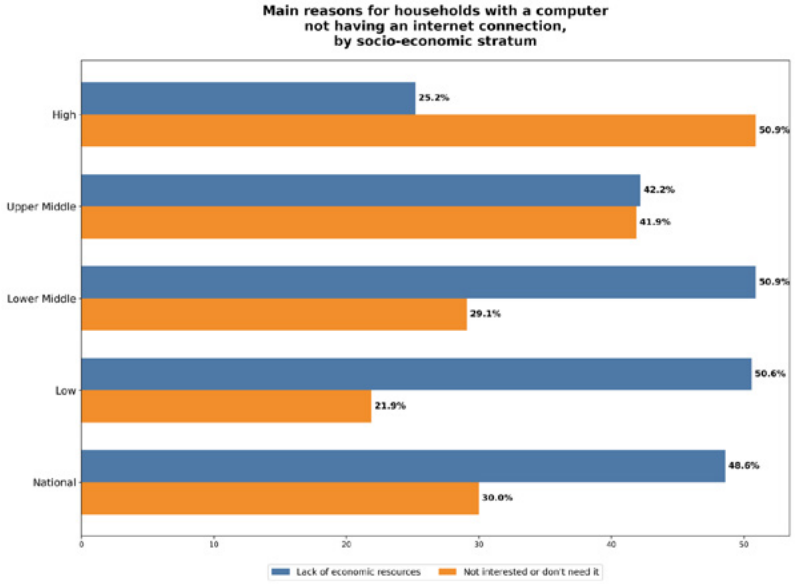

Figure 4. Main reasons for households with a computer not having internet connection by socio-economic stratum
Source: own elaboration with data from ENDUTIH 2023.

The upper-middle stratum, with near-equal distribution between economic and motivational factors for non-adoption, represents a critical transition point, potentially creating a more discerning audience.[297] Nationally, 48.6% of households without internet cite economic reasons, suggesting limited exposure to online content and potentially slowing disinformation spread. However, it is important to note that while this digital divide may temporarily impede the spread of online disinformation, it also presents challenges for digital literacy and access to accurate

[297] INEGI (n 291).

information.[298] The lack of internet access or engagement does not inherently protect against disinformation spread through other channels and may, in fact, leave certain populations more vulnerable to offline forms of disinformation.[299]

The varied landscape of internet adoption in Mexico, as illustrated by this data, suggests a potential retardation in the spread of digital disinformation strategies and particularly the one created with AI. It underscores the need for comprehensive strategies that not only increase internet accessibility but also promote digital literacy and critical thinking skills across all socio-economic strata.[300] Such an approach would aim to create a connected and resilient population against disinformation, regardless of its source or medium of transmission.

---

[298] Claudia Blanca González Calleros and others, 'Addressing the Digital Divide with Educational Systems in Mexico: Challenges and Opportunities' in Łukasz Tomczyk, Francisco David Guillén-Gámez, Juan Ruiz-Palmero, and Alhassan Habibi (eds), *From Digital Divide to Digital Inclusion* (Springer, Singapore 2023) <https://doi.org/10.1007/978-981-99-7645-4_16>.

[299] Elena Gadjanova, Gabrielle Lynch, and Ghadafi Saibu, 'Misinformation Across Digital Divides: Theory and Evidence From Northern Ghana' (2022) 121 African Affairs 161 <https://doi.org/10.1093/afraf/adac009>.

[300] "[...] the education sector is being influenced by the dynamics of the information society. The integration of technology in teaching and learning processes is fostering digital literacy and equipping students with the necessary skills for the digital age. The demand for digital skills is on the rise, and individuals with technological proficiency have a competitive advantage in the job market. The digitalization of industries and the emergence of remote work options are transforming traditional work models and creating new opportunities for economic growth. skills for the digital age.' Claudia Blanca González Calleros (n 296).

5.2 The Political Landscape in Mexico

Mexico's political landscape has transformed significantly in recent decades.[301] The Institutional Revolutionary Party (PRI) dominated for more than seventy years until 2000, when the right-wing National Action Party (PAN) won the presidency. Key events in this transition included the 1994 Zapatista uprising,[302] the PRI losing congressional majority in 1997,[303] and the Democratic Revolutionary Party (PRD) winning Mexico City's mayoralty.[304]

The PAN governed from 2000 to 2012, marked by controversial elections in 2006 and a violent 'War on Drugs'.[305] The PRI briefly returned to power in 2012. In 2018, Andrés Manuel López Obrador's leftist National Regeneration Movement (Morena) won the presidency, promising a 'Fourth Transformation'.[306] Obra-

[301] Roderic Ai Camp, 'Democratizing Mexican Politics, 1982–2012' (2015) Oxford Research Encyclopedias, Latin American History <https://doi.org/10.1093/acrefore/9780199366439.013.12>.

[302] Gemma van der Haar, 'The Zapatista Uprising and the Struggle for Indigenous Autonomy' (2004) 76 Revista Europea de Estudios Latinoamericanos y Del Caribe / European Review of Latin American and Caribbean Studies 99 <http://www.jstor.org/stable/25676074> accessed 28 June 2024.

[303] María Amparo Casar, 'Los gobiernos sin mayoría en México: 1997-2006' (2008) 15(2) Política y gobierno 221 <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-20372008000200001&lng=es&tlng=es> accessed 28 June 2024.

[304] Mexico's democratic transition is closely linked to the development of its electoral institutions. This began with the creation of a more autonomous body, the Federal Electoral Institute (IFE), in 1990, following the 1988 elections marred by accusations of electoral fraud. Further progress was made in 2014 with reforms that transformed the IFE into the National Electoral Institute (INE), aiming to standardise electoral processes at both federal and local levels. INE, 'Historia del Instituto Federal Electoral [History of the Federal Electoral Institute]' (Instituto Nacional Electoral) <https://portalanterior.ine.mx/archivos3/portal/historico/contenido/menuitem.cdd858023b32d5b7787e6910d08600a0/> accessed 28 June 2024. For studies in relation to the elections in 1988 see the work of Cantú, who uses AI to find the alteration of vote tallies. See Francisco Cantú, 'The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election' (2019) 113 American Political Science Review 710 <https://doi.org/10.1017/S0003055419000285>.

[305] Daniel Shailer, 'The official count of disappeared people in Mexico could be an underestimate, say UN and advocates' (AP News, 4 October 2023) <https://apnews.com/article/mexico-missing-disappearances-united-nations-147b08e445c715fe0ee487a5b0787288> accessed 12 June 2024.

[306] Washington Post, 'López Obrador wins Mexican presidency, becoming first leftist to govern in decades' (1 July 2018) <www.washingtonpost.com/world/mexicans-head-to-polls-to-choose-a-new-president/2018/07/01/517e8670-7a2a-11e8-ac4e-421ef7165923_story.html> accessed 28 June 2024.

dor's government, while popular, faced criticism for eroding democratic institutions.[307] His administration was characterised by attempts to dissolve autonomous bodies,[308] stigmatisation of the press,[309] and efforts to transform the judiciary system.

The 2024 election occurred in a context of escalating violence. According to the 'Voting Between Bullets' report, attacks on elected officials, candidates, and party members surged from 311 (2006-2012) to approximately 836 (2018-2023).[310] Organised crime increasingly influenced local politics, often dictating electoral outcomes through violence.[311]

The election featured three main presidential candidates: Claudia Sheinbaum (Morena), Xóchitl Gálvez (Strength and Heart for Mexico opposition coalition including PRI, PAN and PRD parties), and Jorge Álvarez Maynez (Citizens' Movement, MC) and reflected the fragmentation of the traditional party system, and the rise of new political forces,[312] all against a backdrop of increasing violence and organised crime influence in politics. There was a continuation of a deeply transformed political landscape. In response to Morena's growing dominance,[313] traditional parties (PRI, PAN, PRD) formed a 'mishmash of conservative, centrist, and

---

307 Adriana Garcia and Javier Martin-Reyes, 'Guardians of Democracy: Battling for the Rule of Law in Mexico' (Stanford Law School, 24 October 2023) <https://law.stanford.edu/2023/10/24/guardians-of-democracy-battling-for-the-rule-of-law-in-mexico/> accessed 28 June 2024.

308 Elias Camhaji, 'López Obrador va por la eliminación de siete órganos autónomos y entes reguladores [López Obrador is going for the elimination of seven autonomous bodies and regulatory entities]' El País (6 February 2024) <https://elpais.com/mexico/2024-02-06/lopez-obrador-va-por-la-eliminacion-de-siete-organos-autonomos-y-entes-reguladores.html> accessed 28 June 2024.

309 Proceso, 'La estigmatización de los medios, signo de AMLO, según la CIDH y Artículo 19 [The stigmatization of the media, a hallmark of AMLO, according to the IACHR and Article 19]' (Proceso, 8 June 2020) <www.proceso.com.mx/reportajes/2020/6/8/la-estigmatizacion-de-los-medios-signo-de-amlo-segun-la-cidh-articulo-19-244156.html> accessed 28 June 2024.

310 Animal Político, Data Cívica, México Evalúa, *Democracia Vulnerada: El Crimen Organizado en las Elecciones y la Administración Pública en México [Democracy Undermined: Organized Crime in Elections and Public Administration in Mexico]* (2024) <https://votar-entre-balas.datacivica.org/reportes> accessed 22 June 2024.

311 ibid.

312 AP News, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef1f5e9ee242d495570578d5ed> accessed 28 June 2024.

313 Pablo Ferri, 'The rise of Morena: In one decade, the political party has wiped the PRI from Mexico's electoral map' El País (Madrid, 4 June 2023) <https://english.elpais.com/international/2023-06-04/the-rise-of-morena-in-one-decade-the-political-party-has-wiped-the-pri-from-mexicos-electoral-map.html> accessed 28 June 2024.

progressive'[314] coalition called 'Strength and Heart for Mexico'.[315] Simultaneously, the Citizens' Movement (MC) emerged as a significant left-wing force, capitalising on the shifting political dynamics.[316] The election also saw the consolidation of adaptive 'satellite parties',[317] born from the fragmentation of larger political entities. This tripartite development — a united opposition, a rising alternative left, and adapting smaller parties — underscored the ongoing reconfiguration of Mexico's political system, as it grappled with MORENA's increasing power and the challenges of maintaining democratic integrity amidst rising violence and organised crime influence.

[314] AP News, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef1f5e9ee242d495570578d5ed> accessed 28 June 2024.

[315] AP News, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef1f5e9ee242d495570578d5ed> accessed 28 June 2024.

[316] Elías Camhaji, 'Movimiento Ciudadano, a las puertas de la elección más importante de su historia' El País (14 January 2024) <https://elpais.com/mexico/elecciones-mexicanas/2024-01-14/movimiento-ciudadano-a-las-puertas-de-la-eleccion-mas-importante-de-su-historia.html> accessed 28 June 2024.

[317] José Gil Olmos, 'Morena: hegemonía y partidos satélite' ('Morena: Hegemony and Satellite Parties') (Proceso, 24 June 2024) <www.proceso.com.mx/opinion/2024/6/24/morena-hegemonia-partidos-satelite-331506.html> accessed 10 July 2024.

5.3 The disinformation landscape in Mexico: an analysis of recent reports and investigations

In Mexico, journalistic and academic investigations have brought to light the involvement of in-country actions and foreign actors' attempts to disseminate disinformation in relation to political discourse and electoral processes in this 2024 election[318] and previously,[319]. The following analysis synthesises findings from various sources to provide an overview of the current landscape of online manipulation and disinformation in the Mexican context.

Efforts to interfere, influence, manipulate, or distort the online debate have deep roots in Mexico's recent history.[320] These tactics can be traced back to the use of 'Peñabots,' the trolls and bots associated with former President Enrique Peña Nieto. According to news reports, emerging during the 2012 election, these bots became a widespread strategy to manipulate online conversations and silence dissent and critics of his government.[321] This trend has evolved into more recent political campaigns that rely heavily on digital marketing and public relations (PR) companies. These

[318] Arturo Daen, 'También en la oposición: canales de YouTube desinforman sobre AMLO y Sheinbaum, pero elogian a Xóchitl Gálvez [Also in opposition: YouTube channels misinform about AMLO and Sheinbaum, but praise Xóchitl Gálvez]' (Animal Político, 22 February 2024) <www.animalpolitico.com/verificacion-de-hechos/te-explico/canales-desinforman-morena-amlo-xochitl> accessed 8 July 2024 and Arturo Daen, Samedi Aguirre, and Siboney Flores, '"Liga de Guerreros": Red de cuentas en Twitter que respalda a Taboada y Xóchitl desinforma y usa violencia política [League of Warriors': Network of Twitter accounts backing Taboada and Xóchitl misinforms and uses political violence]' (Animal Político, 30 October 2023) <https://animalpolitico.com/verificacion-de-hechos/te-explico/liga-de-guerreros-desinformacion-taboada-xochitl> accessed 8 July 2024.

[319] Freedom House, 'Mexico: Freedom on the Net 2023' (2023) <https://freedomhouse.org/country/mexico/freedom-net/2023#footnoteref6_2a11ufc> accessed 8 July 2024.

[320] Signa_Lab, 'PRI Edomex II: Estrategias de Influencia [PRI Edomex II: Influence Strategies]' Signalab (12 September 2020) <https://signalab.mx/2020/09/08/pri-edomex-ii-estrategias-influencia/> accessed 28 June 2024.

[321] JM Porup, 'How Mexican Twitter Bots Shut Down Dissent' (Vice, 24 August 2015) <www.vice.com/en/article/z4maww/how-mexican-twitter-bots-shut-down-dissent> accessed 16 June 2024.

campaigns span the entire political spectrum in Mexico, with various candidates spending millions of pesos and employing such tactics to influence public opinion, spread disinformation or manage their online presence.[322]

Investigations have also revealed coordinated efforts to shape online conversations, such as the #RedAMLOVE network supporting President López Obrador and smear campaigns against his opponents. In 2019, a report from Signa Lab,[323] an interdisciplinary laboratory from the Instituto Tecnológico de Estudios Superiores de Occidente (ITESO), observed the use of X (former Twitter) by supporters of Mexican President Andrés Manuel López Obrador (AMLO), particularly through the #RedAMLOVE network. Other campaigns have amplified support for López Obrador while attacking opponents, such as a smear campaign against Supreme Court President Norma Lucía Piña Hernández in March 2023 and the discovery of 'Red Brolan,' a network of pro-government YouTube channels, in April 2023.[324]

The López Obrador administration has been accused of using public resources to influence online discourse. This includes the use of InfodemiaMX,[325] a publicly funded 'fact-checking' initiative, which purportedly fact-checks media claims but has been

[322] Juan Gómez, 'Máynez: La maquinaria detrás del fenómeno Millones en estrategia digital y giras por el país [Máynez: The Machinery Behind the Phenomenon Millions in Digital Strategy and Country Tours]' Fábrica de Periodismo (22 May 2024) <https://fabricadeperiodismo.com/reportajes/maynez-maquinaria-electoral-millones-campana/> accessed 28 June 2024 and Antonio López Cruz, 'Popularidad de MC en redes, ¿artificial? [Popularity of MC on social media, artificial?]' El Universal (1 January 2024) <www.eluniversal.com.mx/elecciones/popularidad-de-mc-en-redes-artificial/> accessed 16 June 2024.

[323] Signalab, 'Signalab' (Signalab) <https://signalab.mx/> accessed 16 June 2024.

[324] Animal Político, 'Red Brolan: youtubers afines a AMLO difunden desinformación política y contra periodistas [Brolan Network: AMLO-friendly youtubers spread political and anti-journalist disinformation]' Animal Político (19 April 2023) <www.animalpolitico.com/verificacion-de-hechos/te-explico/brolan-difunde-desinformacion-youtube> accessed 28 June 2024.

[325] The InfodemiaMX platform, ostensibly a fact-checking initiative coordinated by the Mexican Public Broadcasting System and financed with public funds, has reportedly been used to present biased or false information on behalf of the López Obrador government and his MORENA party. In one example from August 2022, an InfodemiaMX program broadcast on social media was used to defend the government's position on a train infrastructure project, calling opposition to the project #MentirasEcologicas (#EcologicalLies). InfodemiaMX has its own website and publishes content on YouTube, Facebook, and Twitter. Arturo Daen and Tania L Montalvo, 'Infodemia y Quién es Quién, más propaganda que chequeo con recursos públicos en México [Infodemic and Who's Who, more propaganda than checking with public resources in Mexico]' Animal Político (Centro Latinoamericano de Investigación Periodística, 30 November 2022) <www.elclip.org/infodemia-y-quien-es-quien-propaganda-desinformacion-mexico/> accessed 28 June 2024.

found to propagate misleading information and the controversial 'Who's Who in Lies' segment of the President's morning press conferences.[326] These actions have raised concerns about freedom of expression, press freedom and the spread of misleading information. In February 2022, the Inter-American Commission on Human Rights (IACHR) called for the suspension of the 'Who's Who in Lies' ('Quién es quién en las mentiras' in Spanish) segment from President López Obrador's morning press conferences.[327] The Special Rapporterur highlighted the need for the Mexican government to uphold freedom of expression and protect journalists, noting that Mexico remains one of the most dangerous countries for the press, with numerous attacks and threats against media professionals in recent years.[328]

Foreign actors have also been implicated, with a network of Venezuelan websites discovered supporting López Obrador and spreading disinformation about his rivals.[329] Meanwhile, at the national level, the media outlet Animal Político reported that the consulting firm *Heurística* financed disinformation campaigns against opposition candidates during the 2018 election.[330]

Cross-regional collaborations among journalists have been crucial for tracing and exposing disinformation networks, thereby ensuring the public receives accurate and verified news. For instance, the cross-border investigation 'Mercenarios Digitales',[331]

---

326 Artículo 19, 'Negación [Denial]' (Artículo 19, 2023) <https://articulo19.org/negacion/> accessed 28 June 2024 31 and Karla Velázquez and David Martínez, 'Lo cierto y lo falso del "Quién es quién en las mentiras" [The truth and falsehoods of AMLO's "Who's Who in Lies"]' Verificado (8 July 2021) <https://verificado.com.mx/imprecisiones-y-datos-falsos-en-una-tercera-parte-del-quien-es-quien-en-las-mentiras-de-la-mananera-de-amlo/> accessed 18 June 2024.

327 Proceso, 'Exclusiva: Relator de CIDH pide detener "Quién es quién de las mentiras" por violencia a periodistas [Exclusive: IACHR Rapporteur calls to stop "Who's Who in Lies" due to violence against journalists]' (1 February 2022) <www.proceso.com.mx/nacional/2022/2/1/exclusiva-relator-de-cidh-pide-detener-quien-es-quien-de-las-mentiras-por-violencia-periodistas-280177.html> accessed 28 June 2024.

328 ibid.

329 Lidia Sánchez and Esteban Ponce de Léon, 'Red de sitios venezolanos a cargo de desinformación y propaganda sobre México, El Salvador, España y Perú [Network of Venezuelan sites in charge of disinformation and propaganda on Mexico, El Salvador, Spain and Peru]' Animal Político (27 July 2022) <https://mirror.animalpolitico.com/2022/07/red-de-sitios-venezolanos-desinformacion-propaganda-sobre-mexico-espana/> accessed 28 June 2024.

330 Zedryk Raziel, 'Empresa publicista de campaña de AMLO financió desinformación contra Ricardo Anaya en 2018 [AMLO's PR firm financed misinformation against Ricardo Anaya in 2018]' Animal Político (27 April 2022) <www.animalpolitico.com/2022/04/empresa-publicista-amlo-2018-financio-desinformacion-anaya/> accessed 28 June 2024.

331 El Clip, 'Mercenarios digitales [Digital Mercenaries]' El Clip (2023) <www.elclip.org/mercenarios-digitales/> accessed 28 June 2024.

coordinated by CLIP, uncovered Neurona's involvement in propaganda and disinformation campaigns. This investigation revealed Neurona's promotion of Mexican presidential aspirant Adán Augusto López and other Morena political figures, as well as its support for leftist political candidates across various Latin American countries.[332]

Civil society organisations have played a crucial role in exposing disinformation networks. For instance, the Red en Defensa de los Derechos Digitales (Network in Defence of Digital Rights, R3D) revealed the existence of a covert Cyberspace Operations Centre within the Ministry of National Defence, allegedly engaged in surveillance and manipulation of online discourse.[333] The COC purportedly employs inauthentic bots to influence public opinion, raising critical issues about the state's role in digital rights violations and the human rights implications of such actions. These findings, according to R3D, underscore the need for increased transparency and accountability within state operations concerning digital communications and public information management.[334]

[332] Eduardo Buendía and others, 'Neurona: la fábrica de engaño para las izquierdas en América Latina [Neurona: The Deception Factory for the Left in Latin America]' (El Clip, 2023) <www.elclip.org/neurona-la-fabrica-de-engano-para-las-izquierdas-en-america-latina/> accessed 22 June 2024.

[333] R3D, 'Ejército de Bots: Las operaciones militares para monitorear las críticas en redes sociales y manipular la conversación digital [Army of Bots: Military operations to monitor social media critics and manipulate digital conversation]' (R3D, 27 February 2024) <https://r3d.mx/2024/02/27/ejercito-de-bots-las-operaciones-militares-para-monitorear-las-criticas-en-redes-sociales-y-manipular-la-conversacion-digital/> accessed 28 June 2024.

[334] WIRED, 'El ejército mexicano tiene un centro secreto para monitorear opositores y operar bots [The Mexican Army has a secret center to monitor opponents and operate bots]' WIRED (2024) <https://es.wired.com/articulos/el-ejercito-mexicano-tiene-un-centro-secreto-para-monitorear-opositores-y-operar-bots> accessed 28 June 2024.

Meta has also taken action against Coordinated Inauthentic Behaviour (CIB) in Mexico, removing thousands of accounts and pages connected to political strategy firms.[335] According to Meta's 2022 report on enforcement against CIB, Mexico ranked third globally, following Russia and Iran, regarding detected networks involved in disseminating disinformation.[336]

This review highlights the extensive efforts by state and non-state actors to manipulate online discourse and influence public opinion. Social media platforms like Meta have taken actions such as removing inauthentic accounts and increasing political advertising transparency, but these measures have been insufficient. Recently, Meta announced the closure of CrowdTangle, a tool for monitoring political disinformation and hate speech, prompting the Mozilla Foundation to urge Meta to keep it operational through January 2025.[337] This call emphasises the need for effective real-time transparency tools to monitor the online ecosystem, particularly during the 2024 election year.

In the disinformation landscape, fact-checking initiatives play a crucial role in the information ecosystem in Mexico. Moreno-Gil and others have found that 'In a time characterized by profound challenges, fact-checking interventions represent an alternative destination in journalism that seeks to combat the spread of disinformation, educate citizens, and contribute to restoring the credibility of journalism'.[338] Noteworthy contributions in Mexico come from El Sabueso,[339] a project by the media outlet Animal

[335] Meta, 'Recapitulamos nuestras acciones contra el comportamiento inauténtico coordinado en 2022 [We recap our actions against coordinated inauthentic behavior in 2022]' (Meta, 20 December 2022) <https://about.fb.com/ltam/news/2022/12/recapitulamos-nuestras-acciones-contra-el-comportamiento-inautentico-coordinado-en-2022/> accessed 28 June 2024.

[336] Meta, 'June 2021 Coordinated Inauthentic Behavior Report' (Meta, June 2021) <https://about.fb.com/wp-content/uploads/2021/07/June-2021-CIB-Report-Final.pdf> accessed 28 June 2024.

[337] Mozilla Foundation, 'Open Letter to Meta: Support CrowdTangle Through 2024 and Maintain CrowdTangle Approach' (Mozilla Foundation, 2023) <https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/> accessed 28 June 2024.

[338] Victoria Moreno-Gil and others, 'Fact-Checking Interventions as Counteroffensives to Disinformation Growth: Standards, Values, and Practices in Latin America and Spain' (2021) 9(1) Media and Communication 251 <https://doi.org/10.17645/mac.v9i1.3443>.

[339] Animal Político, 'Verificación de Hechos' (Animal Político, 2024) <www.animalpolitico.com/verificacion-de-hechos> accessed 28 June 2024.

Político, Reuters Verification[340] and AP Fact Check.[341] El Sabueso and Animal Político represent an example of debunking disinformation and journalistic investigations. Other cases of fact-checking can be found between 2019 and 2021. The Google News Initiative provided support for at least three journalism and fact-checking projects in Mexico, specifically CódigoPostal.com, Periodistas de a Pie, and Verificado MX.[342] The National Electoral Institute (INE) has also been proactive, launching the #Certeza2024[343] initiative to monitor and address false information during electoral periods.[344]

For the 2023-2024 Federal Electoral Process, INE representatives collaborated with Meta's Public Policy team to enhance 'election integrity'.[345] This collaboration included the implementation of various tools and mechanisms to prevent digital gender-based political violence and combat disinformation. Meta's preparations included launching the 'Inés' chatbot on WhatsApp,[346] establishing a network of independent fact-checkers, operating an Election Operations Centre, emphasising transparency in political adverts, and labelling AI-generated content. TikTok committed to platform

340 Reuters, 'Fact Check en Español' (Reuters) <www.reuters.com/fact-check/espanol/> accessed 28 June 2024.

341 AP News, 'AP Fact Check' (AP News) <https://apnews.com/ap-fact-check> accessed 28 June 2024.

342 Google News Initiative, 'Proyectos seleccionados [Selected Projects]' (Google News Initiative, 2018) <https://20190322t164649-dot-gweb-news-initiative.appspot.com/intl/es/innovation-challenges/funding/latin-america/> accessed 14 June 2024.

343 Instituto Nacional Electoral, '#Certeza 2024 - Central Electoral' (Central Electoral, 2024) <https://centralelectoral.ine.mx/certeza/> accessed 14 June 2024.

344 Instituto Nacional Electoral, 'Metodología Certeza 2024' (Central Electoral, 2024) <https://centralelectoral.ine.mx/wp-content/uploads/2024/03/Metodologi%CC%81a_Certeza-2024.pdf> accessed 14 June 2024.

345 Instituto Nacional Electoral, 'Se reúnen Consejeras y Consejeros del INE con representantes de Meta [INE Councilors meet with Meta representatives]' Central Electoral (19 September 2023) <https://centralelectoral.ine.mx/2023/09/19/se-reunen-consejeras-y-consejeros-del-ine-con-representantes-de-meta/> accessed 28 June 2024.

346 Anna Lagos, 'Meta se prepara para las elecciones del próximo 2 de junio en México con estas acciones [Meta prepares for the upcoming June 2 elections in Mexico with these actions]' WIRED (16 April 2024) <https://es.wired.com/articulos/meta-se-prepara-para-las-elecciones-del-proximo-2-de-junio-en-mexico-con-estas-acciones> accessed 28 June 2024.

integrity by updating its Community Guidelines, launching an Electoral Guide with INE, partnering with fact-checking organisations, and implementing stricter policies for political accounts, including prohibiting paid political adverts and monetisation.[347]

5.4 GenAI in Mexican elections: from disinformation to democratic discourse

Generative AI in the context of elections extends beyond generating disinformation content; it also plays a crucial role in other forms of expression within political debates. This technology can introduce and amplify various viewpoints, contributing to the free flow of information and enriching political discourse. This section explores the use of GenAI during the Mexican elections and some instances of disinformation, considering the broader implications for information flow and democratic processes.

The cases presented in the following section were documented during the electoral process,[348] which spanned the following periods: 1) 'pre-campaigns' from November 20, 2023 to January 18, 2024; 2) 'Inter-campaigns' from January 19 to February 29, 2024; 3) 'Campaigns' from March 1 to May 29, 2024; 4) reflection time[349] from May 29 to June 2, 2024; 5) Election Day on June 2, 2024, and 6) Election counts from June 5 to 9, 2024.

---

[347] TikTok, 'TikTok refuerza su compromiso con la integridad de la plataforma con iniciativas clave de cara a las próximas elecciones en México [TikTok reinforces its commitment to platform integrity with key initiatives for the upcoming elections in Mexico]' TikTok Newsroom (9 May 2024) <https://newsroom.tiktok.com/es-latam/elecciones-mexico-2024> accessed 28 June 2024.

[348] Instituto Nacional Electoral (n 344) 1.

[349] INE, 'Concluyen campañas electorales e inicia periodo de reflexión [Electoral campaigns conclude and reflection period begins]' (Central Electoral, 29 May 2024) <https://centralelectoral.ine.mx/2024/05/29/concluyen-campanas-electorales-e-inicia-periodo-de-reflexion/> accessed 14 June 2024. Article 251 from LGIPE establishes that it is 'prohibited to publish or disseminate by any means, the results of opinion polls or surveys whose purpose is to make known the electoral preferences of citizens, and those who do so shall be subject to the penalties applicable to those who commit any of the offences foreseen and sanctioned in the General Law on Electoral Offences'. LGIPE Art 251 Numeral 6. Although candidates and political parties were not allowed to publish any information during this period, the Partido Verde Ecológista de México (Green Ecologist Party of Mexico, PVEM) has repeatedly used a network of influencers to promote its image in past elections. See Diana Soto, 'El Partido Verde lo volvió a hacer: recibe promoción de influencers y modelos, pese a prohibición en intercampañas [The Green Party did it again: receives promotion from influencers and models despite inter-campaign prohibition]' (Animal Político, 29 May 2024) <https://animalpolitico.com/verificacion-de-hechos/fact-checking/influencers-y-modelos-publican-mensajes-a-favor-del-partido-verde> accessed 29 June 2024.

AI-generated content appeared in three distinct categories: i) scams and fraudulent content; ii) disinformation aimed at manipulating public perception, and iii) other content, including satire and protected expressions in line with the Interamerican human rights system.[350]

AI-generated disinformation was observed during the election cycle.[351] These ranged from manipulated videos depicting candidates endorsing fraudulent investment schemes[352] to doctored audio recordings spreading false narratives about political figures.[353] The National Securities Market Commission[354] (CNMV)

[350] The jurisprudence of the Inter-American Human Rights System (IAHRS) has established three categories of specially protected speech, recognising their fundamental role in strengthening democracy and the full exercise of human rights. Firstly, the protection of political speech and speech on matters of public interest, considering that this type of expression is essential for the formation of an informed public opinion and for the participation of individuals in democratic processes and public affairs. Secondly, enhanced protection for speech about public officials in the exercise of their functions and about candidates for public office, on the understanding that scrutiny of the actions of those who hold or aspire to positions of power is crucial for transparency and accountability. Finally, special protection for speech that constitutes an element of the identity or personal dignity of the speaker, thus recognising the importance of freedom of expression for individual development and personal autonomy. This differentiated protection seeks to ensure that these types of speech, which are fundamental to public debate and personal fulfilment, are not unduly restricted, thus promoting a more open, pluralistic and democratic society. Comisión Interamericana de Derechos Humanos (CIDH), 'Marco Jurídico Interamericano del Derecho a la Libertad de Expresión' ('Inter-American Legal Framework on the Right to Freedom of Expression') (Organización de los Estados Americanos, 2009) parr 32-56, Disponible en: <www.oas.org/es/cidh/expresion/docs/publicaciones/MARCO%20JURIDICO%20INTERAMERICANO%20DEL%20DERECHO%20A%20LA%20LIBERTAD%20DE%20EXPRESION%20ESP%20FINAL%20portada.doc.pdf> accessed 10 June 2024.

[351] Antonio, '¡QUE SE HAGA VIRAL! Si el PAN obtiene menos de 7 millones de votos, no podrán llegar al Congreso los plurinominales: - Cabeza de Vaca - Ricardo Anaya - Marko Cortés Mendoza [MAKE IT GO VIRAL! If the PAN gets less than 7 million votes, the proportional representatives: - Cabeza de Vaca - Ricardo Anaya - Marko Cortés Mendoza]' (X, 14 April 2024) <https://x.com/JTonyy35/status/1779626643629150662> accessed 29 June 2024

[352] Andrés Martínez, 'Fraude en venta de acciones de Pemex emplea imágenes y voz falsas de Claudia Sheinbaum [Fraud in Pemex stock sale uses fake images and voice of Claudia Sheinbaum]' Infobae, 2 January 2024 <https://shorturl.at/izEOX> accessed 29 June 2024; Oscar Nogueda Romero, 'Claudia Sheinbaum no llamó a invertir en petróleo, el video fue manipulado [Claudia Sheinbaum did not call for investment in oil, video was manipulated]' Animal Político, 1 December 2023 <https://animalpolitico.com/verificacion-de-hechos/desinformacion/sheinbaum-invertir-petroleo-falso> accessed 29 June 2024.

[353] vagagu, 'las traiciones del bienestar... [the betrayals of well-being]' (X, 3 October 2023) <https://x.com/vagagu/status/1709041132523463096> accessed 29 June 2024.

[354] EP, 'La CNMV alerta del fraude de Quantum AI por usar imágenes de famosos para publicitarse en redes sociales [The CNMV alerts about the fraud of Quantum AI for using images of celebrities to advertise on social networks]' (El País, 12 December 2023) <https://elpais.com/economia/2023-12-12/la-cnmv-alerta-del-fraude-de-quantum-ai-por-usar-imagenes-de-famosos-para-publicitarse-en-redes-sociales.html> accessed 29 June 2024.

even issued warnings about allegedly AI-powered platforms using celebrity images without authorisation to promote dubious investment opportunities. The morning press conference of President López Obrador also alerted people about the use of deepfakes to scam people.[355]

The sophistication of image and video manipulation techniques was particularly notable. Campaign teams and supporters alike shared altered images and out-of-context videos,[356] demonstrating the potential for even official accounts to inadvertently disseminate manipulated content.[357] Following the second presidential debate, Claudia Sheinbaum's team shared a visibly altered image on her official X (formerly Twitter) account, notably depicting her with six fingers in one hand. In another instance, a manipulated image falsely showed Xóchitl Gálvez waving the Mexican flag upside down, shared by a Morena supporter account with over 229,000 followers.[358] These incidents highlighted the increasing difficulty in distinguishing between genuine and fabricated content.

The spread of disinformation took various forms, including the creation of false endorsements from well-known companies. A fabricated advertisement purportedly from the poultry company Bachoco[359] endorsing the opposition candidate Xóchitl Gálvez circulated widely on social media. Similarly, images fabricated to impersonate coffee shop Starbucks promotional materials[360] fea-

[355] Jenaro Villamil, '#ConferenciaMañanera. Alertan sobre las Deepfakes que usan herramientas de inteligencia artificial para estafar a usuarios de plataformas digitales [Morning Conference. Warning about Deepfakes using artificial intelligence tools to scam digital platform users]' (X, 1 May 2024) <https://x.com/jenarovillamil/status/1785665951578407183> accessed 29 June 2024.

[356] Maximoam_, '#Exclusiva "YO TENÍA TODAS LAS NARCOTIENDITAS"... [#Exclusive "I HAD ALL THE DRUG STORES"...]' (X, 22 May 2024) <https://x.com/maximoam_/status/1793337455824802018> accessed 29 June 2024.

[357] Xochitl Galvez, 'Con razón no te salen las cuentas, Sheinbaum. En tu foto sales con 6 dedos en cada mano.'[No wonder you don't figure it out, Sheinbaum. In your photo you appear with 6 fingers on each hand] (X, 29 April 2024) <https://x.com/XochitlGalvez/status/1784822900672893328> accessed 12 May 2024

[358] Rest of World Staff, 'A manipulated Mexican flag' Rest of World, 13 May 2024 <https://restofworld.org/2024/elections-ai-tracker/#/manipulated-mexico-flag> accessed 29 June 2024.

[359] Reuters Fact Check, "Verificación: Bachoco no publicó cartel en apoyo a Xóchitl Gálvez [Verification: Bachoco did not publish a poster in support of Xóchitl Gálvez]", October 5, 2023, <www.reuters.com/fact-check/espanol/NBESMKHW7RJE5EHFBW-GOVMNCCI-2023-10-10/> accessed 29 June 2024.

[360] Sociedad Civil México, <https://x.com/SocCivilMx/status/1769850004548559314?s=20> accessed 29 June 2024.

turing the hashtag #Xóchitl2024 gained significant traction online. These incidents prompted swift responses from the implicated businesses,[361] underscoring the reputational risks posed by AI-generated false content.

In response to these challenges, a coalition of organisations and institutions played crucial roles in combating disinformation.[362] Fact-checking organisations and the candidates themselves worked tirelessly to debunk false claims and maintain the information integrity of the electoral process.[363]

However, not all AI-generated content was false information, disseminated intentionally to cause serious social harm or to manipulate the elections. A significant portion of users leveraged Generative AI for creating humorous[364] and satirical content,[365] including memes and videos[366] that garnered millions of views. This showcased the technology's potential for enhancing political engagement and discourse, allowing people to explore other forms for their right to freedom of expression. For example, AI-generated Balenciaga memes87 portrayed presidential candidates, other political figures, and the current president López Obrador. This viral trend garnered 6 million views on TikTok.[367] Other us-

[361] Bachoco, "Mensaje importante a toda la comunidad: [Important message to the entire community:], Facebook, Sepmtember 22, 2023, <www.facebook.com/BachocoMX/posts/681164897379600> accessed 29 June 2024.

[362] Samedi Aguirre, "'Las máquinas aprenden': Inteligencia Artificial evoluciona y puede usarse para engañar, pero no todo está perdido ['Machines learn': Artificial Intelligence evolves and can be used to deceive, but all is not lost]," April 23, 2023, <www.animalpolitico.com/verificacion-de-hechos/te-explico/inteligencia-artificial-evoluciona-desinformacion-herramientas> accessed 8 April 2024; @elsabuesoap, El Sabueso de Animal Político, TikTok, April 6, 2023, <www.tiktok.com/@elsabuesoap/video/7218968175268859142> accessed 5 May 2024.

[363] Sara Pantoja, 'Sheinbaum alerta sobre video falso hecho con inteligencia artificial: "es mi voz, pero es un fraude"' ('Sheinbaum warns about fake video made with artificial intelligence: "it's my voice, but it's a fraud"') (Proceso, 25 January 2024) <www.proceso.com.mx/nacional/politica/2024/1/25/sheinbaum-alerta-sobre-video-falso-hecho-con-inteligencia-artificial-es-mi-voz-pero-es-un-fraude-322795.html> accessed 10 June 2024.

[364] ElFranky_, 'Unas caguamas bien heladas al que lo hizo [A few ice-cold beers to the one who did it]' (X, 27 March 2024) <https://x.com/ElFranky_/status/1773039123499995226> accessed 5 May 2024.

[365] spikolsmaniac, 'Hola Twitter... [Hello Twitter...]' (X, 1 June 2024) https://x.com/spikolsmaniac/status/1796736418091237677> accessed 29 June 2024.

[366] Nación321, '"¡Awilsoooon!"... [From storm to tropical storm, Anaya 'castaway' survives AMLO's term and sends support to Xóchitl Gálvez from exile and AI]' (X, 31 August 2023) <https://x.com/Nacion321/status/1697300718213018066> accessed 29 June 2024.

[367] @noporsuave, 'Elecciones presidenciales en Balenciaga [Presidential elections in Balenciaga]' (TikTok, 29 May 2024) <www.tiktok.com/@noporsuave/video/7374464596154813701> accessed 29 June 2024.

ers, for example, created a video[368] depicting the three presidential candidates dancing to a trending song by Colombian singers *Karol G* and *Shakira* after one of the three presidential debates. The video has more than 19 million views and almost 19 thousand comments. It was generated with Viggle.ai, 'a video-3d foundation model that [allows you to] animate any character image with text prompts'.[369] This tool can be deployed in Discord,[370] and according to the company, it has a community of 2 million creators who use it to 'make fun and creative videos'.[371] In the comment section, someone asked, 'How will we explain the 2024 elections in the history books?' while another person inquired, 'Chat, is this real?' to which someone responded, 'Yes, at the end of the debate, they always record for their TikTok, only this time they showed off'.[372] The former presidential candidate, Jorge Álvarez Maynez, shared the video on his TikTok account.

During the pre-campaign period, candidate Xóchitl Gálvez, capitalising on the positive reception of several videos portraying her as a strong presidential candidate,[373] decided to incorporate GenAI into her communication strategy. In a video created with this technology, she announced, 'I want to tell you that from today we will implement a special spokesperson with the use of artificial intelligence [...] we have decided to officialize this approach. Generative AI makes time and resources efficient, it is very quick

[368] oscarwildones, 'Tercer Debate Ine [Third Debate Ine]' (TikTok, 20 May 2024) <www.tiktok.com/@oscarwildones/video/7370923051560602886> accessed 29 June 2024.

[369] Viggle, 'Viggle AI' <www.viggle.ai/> accessed 29 June 2024.

[370] 'It's like Zoom, but more flexible and fun. It's Slack, but without that feeling that your boss is always checking your online status. Facebook, without an algorithm that prioritizes the types of posts that turned your aunt into a racist'. Jaina Grey, 'How to Use Discord: A Beginner's Guide' (WIRED, 3 June 2024) <www.wired.com/story/how-to-use-discord/> accessed 8 July 2024.

[371] ibid.

[372] oscarwildones (n 368) 'comment section'.

[373] Helen VL, '#IXOCHITL: Construyendo el Futuro de México [#IXOCHITL: Building the Future of Mexico]' (X, 2 July 2023) <https://x.com/HelenVL6/status/1675511645869842433> accessed 5 May 2024 and Grupo Fórmula, 'Simpatizantes de Xóchitl Gálvez crearon un video con IA luego de las declaraciones de AMLO sobre que ella sería la candidata de la oposición [Supporters of Xóchitl Gálvez created an AI video following AMLO's statements about her being the opposition candidate]' (X, 4 July 2023) <https://x.com/Radio_Formula/status/1676279234992414731> accessed 5 May 2024.

to produce material, and it is very cheap. From today on, iXóchitl will be one of the spokespersons for the pre-candidate, with my endorsement, supervision, and approval of the messages, which will be permanently hosted on her social networks'.[374]

A Facebook video shows a young woman with fair skin, reddish-blonde hair, and blue eyes, claiming to be an AI named Luna. The post states that Luna was asked who to vote for in the upcoming June 2 elections. In the video, Luna says, 'To all the young people of Mexico, you are the hope that the Morena party will not continue to govern Mexico. If you are thinking of voting for Morena, let me tell you that you are making a big mistake'.[375] A comment questions the video's authenticity and includes a screenshot of ChatGPT advising voters to research candidates before deciding.[376]

During the observation period, it was possible to identify users in TikTok posting videos to teach how to use GenAI tools to clone voice or create deepfakes,[377] while also alerting about using AI to scam people. 'You have to be very careful. The benefits of Artificial Intelligence are obvious, but with these benefits come **certain responsibilities** and **care** that we must employ to avoid falling into this type of advertising done with voice cloning and AI lip-syncing'.[378]

---

[374] Xóchitl Gálvez Ruiz, ‹Les presento a mi nueva vocería de inteligencia artificial. Una herramienta pionera e innovadora que únicamente será oficial a través de mis redes sociales: iXóchitl› (‹I present to you my new artificial intelligence spokesperson. A pioneering and innovative tool that will only be official through my social networks: iXóchitl') (X, 17 December 2023) <https://x.com/XochitlGalvez/status/1736511528130478348> accessed 12 May 2024.

[375] Jose O Trejo, 'Usando la tecnología (inteligencia artificial) se le preguntó a un robot por quién votar en las próximas elecciones del 2 de junio, y vean lo qué contestó en base a hechos [Using technology (artificial intelligence) a robot was asked who to vote for in the upcoming June 2 elections, and see what it answered based on facts]' (Facebook, 19 May 2024) <www.facebook.com/jose.o.trejo/posts/usando-la-tecnolog%C3%ADa-inteligencia-artificial-se-le-pregunt%C3%B3-a-un-robot-por-qui%C3%A9n/8116051695072480/> accessed 29 June 2024.

[376] ibid.

[377] @edsonstartingnet, 'Nuevo e increíble método para hacer tus deeps... [New and amazing method for making your deeps...]' (TikTok, 8 March 2024) <www.tiktok.com/@edsonstartingnet/video/7344128059370458373> accessed 29 June 2024.

[378] edsonstartingnet, 'Video' (TikTok, 6 January 2023) <www.tiktok.com/@edsonstartingnet/video/7320759295358881029> accessed 29 June 2024.

Despite some media narratives that 'deepfakes' and 'fake news' won in the voting,[379] generative AI opens up exciting new possibilities for political engagement and freedom of expression during elections. This technology gave users the chance to participate in the electoral process by creating engaging, satirical, and expressive content, enriching democratic discourse. Also, candidates used AI to connect with voters in alternative ways, and integrated GenAI into their campaign.

[379] Christopher Calderón, 'Elecciones México 2024: Deep Fakes y Fake News Ganan en las Votaciones (Mexico Elections 2024: Deep Fakes and Fake News Win in the Votes)' (El Financiero, 4 June 2024, 3:00 am) <www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/> accessed 13 July 2024.

6. Conclusions

This thesis has explored the complex relationship between GenAI, freedom of expression, and the principles underlying the integrity of democratic processes in the disinformation age, using the Mexican electoral landscape as a case study. In the crossroads of technological advancement and democratic principles, the impact of Generative AI on election integrity remains a subject of intense debate. While some researchers warn of its potential to amplify disinformation and manipulate public opinion, others argue that these concerns may be overstated. After conducting this research, it can be concluded that the informational landscape underwent an alteration due to the presence of GenAI on social media. Whether it was significant, the conclusion might not be completely compelling. The information ecosystem is a complex and intricate network, and challenging to analyse effectively. Yoachim Benkler has stated that, 'it is equally critical not to confound whether a phenomenon is observable and whether it actually has an impact'.[380] However, this research allowed it to be affirmed that the digital divide may have a contentious effect on the spread of disinformation, though further research is required to understand its full impact. This election demonstrated that GenAI, when used on social media platforms, can stimulate public participation in political debates and increase involvement in public affairs. It can also serve as a vehicle to interfere with the right to know, to confuse and deceive.

The election served as a real-world laboratory, revealing both the challenges and opportunities presented by AI in democratic processes. On the one hand, it highlighted the relevance of counter narratives and fact-checking initiatives against sophisticated disinformation campaigns and AI-enabled scams. On the other, it underscored AI's capacity to democratise content creation, enabling a broader spectrum of voices to participate in political dialogue through memes, videos, and other creative content. This dichotomy emphasises the complex role of AI in modern elections.

[380] Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press 2018).

The LLM exercise reveals critical insights into the use of generative artificial intelligence in the context of information access. It underscores the importance of recognising the limitations of these systems, whether due to structural constraints, training data quality, or inherent biases. Furthermore, it highlights the ongoing necessity for evaluating these chatbots from a human rights perspective, particularly in assessing their potential negative impacts on individuals and society, both when used as a falsehood factory or to provide factual content. Crucially, the exercise emphasises the responsibility of companies developing and deploying these technologies to implement robust due diligence mechanisms and establish effective remediation processes. These measures are essential for mitigating human rights impacts and addressing any adverse effects. As LLMs continue to evolve and integrate into various aspects of information dissemination, maintaining a vigilant and proactive approach to their development and use will be paramount in safeguarding human rights whilst harnessing the potential benefits of this transformative technology.

This is why the UN Guiding Principles on Business and Human Rights are relevant in the context of GenAI and elections. It highlights the responsibility of technology companies, particularly social media platforms and AI developers, to conduct human rights due diligence and provide effective remedies for adverse impacts of their technologies. As shown in section three, there is space to continue adopting different mechanisms to provide remedies in a context where tension exists between GenAI's potential to democratise information access and its capacity to amplify dominant voices and potentially manipulate public opinion. Companies can also play a significant role in being proactive in identifying possible human rights impacts on their platforms. For example, by dismantling attempts to spread disinformation, or by introducing changes to promote greater transparency in the functioning of their content moderation governance systems. This may include the participation of more researchers to better understand the information dynamics and the functioning of AI internal systems of social media platforms.

The evolution of AI, from its conceptual inception to the current era of GenAI, has brought about a paradigm shift in how scientific research is transforming alongside how information is created, disseminated, and consumed. This transformation has profound implications for the exercise of human rights, particularly

freedom of expression and the right to participate in public affairs. The thesis has underscored the importance of viewing these technological advancements through a human rights lens, ensuring that the development and deployment of AI systems align with principles of human dignity and respect. The international human rights framework, including key instruments such as the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights, as well as other soft law instruments like General Comments and the vast jurisprudence from regional courts, provides a robust foundation for addressing the challenges posed by GenAI in electoral contexts. It recognises the relevance to protect the free flow of information and a robust political debate, as well as the equality for every person to participate in public affairs. The research has highlighted the critical need to balance the protection of freedom of expression with efforts to combat disinformation, emphasising that any restrictions on these rights must meet the 'three-part test' of legality, necessity, and proportionality.

The dual nature of GenAI as both a tool for expanding creative expression and a potential capability for spreading disinformation requires a careful approach to promote multistakeholder participation in the governance of AI systems, as well as solid and effective regulation from states that fosters innovation whilst safeguarding against undermining the very heart of democracies. To address this, it is crucial for stakeholders—including voters, candidates, media organisations, and regulatory bodies—to develop a multidimensional understanding of AI's capabilities and limitations. This understanding is essential, among other things, for fostering an informed electorate capable of distinguishing between harmful disinformation and legitimate political expression in the age of artificial intelligence. Moving forward, striking this balance will be key to ensuring the integrity of democratic processes while leveraging the benefits of AI technology.

International human rights organisations have also noted the role of multiple stakeholders in combating disinformation and preserving electoral and informational integrity. Fact-checking organisations, civil society groups, electoral authorities, and social media platforms have played vital roles in identifying and countering false information. However, the persistence and sophistication of disinformation campaigns highlight the need for

more robust and coordinated efforts. This underscores the importance of a comprehensive strategy that involves all relevant parties working together to address the complex challenges posed by disinformation.

Bibliography

Books and articles

Adam M with C Hocquard, 'Artificial intelligence, democracy and elections' (Members' Research Service, PE 751.478, October 2023)

Adorno TW and M Horkheimer. *Dialectic of Enlightenment* (1997)

Álvarez Hernández MD, SE Álvarez Hernández and M Álvarez Texocotitla, 'La complejidad del sistema geoelectoral mexicano a nivel municipal [The Complexity of the Mexican Geoelectoral System at the Municipal Level]' (2022) 66 Revista Mexicana de Ciencias Políticas y Sociales 167 <https://dialnet.unirioja.es/servlet/articulo?codigo=8434447> accessed 20 May 2024

Auma Simiyu M, 'Freedom of Expression and African Elections: Mitigating the Insidious Effect of Emerging Approaches to Addressing the False News Threat' (2022) 22(1) African Human Rights Law Journal <https://hdl.handle.net/10520/ejc-ju_ahrlj_v22_n1_a5> accessed 13 July 2024

Baglayan B, 'A Study on Potential Human Rights Due Diligence Legislation in Luxembourg' (2021) <https://orbilu.uni.lu/bitstream/10993/48683/1/Baglayan_Study_HRDD.pdf> accessed 8 July 2024

Bail CA and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization' (2018) 115(37) PNAS 9216, 9221 <https://doi.org/10.1073/pnas.1804840115>

Bebiak E, 'Human Rights Due Diligence: The European Union's Approach to Ensuring Respect for Human Rights in Business' (European Master's Degree in Human Rights and Democratisation, Adam Mickiewicz University 2018/2019) < http://dx.doi.org/10.25330/1936>

Belli L and M Wisniak, What's in an Algorithm? Empowering Users Through Nutrition Labels for Social Media Recommender Systems, 23-06 Knight First Amend Inst (Aug 22, 2023), <https://knightcolumbia.org/content/whats-in-an-algorithm-empowering-users-through-nutrition-labels-for-social-media-recommender-systems>

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

Bommasani R and others, 'On the Opportunities and Risks of Foundation Models' (2021) <https://arxiv.org/pdf/2108.07258.pdf> accessed 5 April 2023

Bontcheva K, Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities (February 2024) <https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf> accessed 6 June 2024

Botero Arcila B and R Griffin, 'The influence of social media on elections and political debate' (Policy Department for Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, PE 743.400, April 2023) <www.europarl.europa.eu/RegData/etudes/STUD/2023/743400/IPOL_STU(2023)743400_EN.pdf> accessed 8 May 2024 78

Bowman SR, "Eight Things to Know about Large Language Models" (2023) International Conference on Machine Learning <https://arxiv.org/pdf/2304.00612> accessed 30 June 2024

Brynjolfsson E, The Turing Trap: The Promise and Peril of Human-Like Artificial Intelligence (American Academy of Arts & Sciences, 2022) <www.amacad.org/publication/turing-trap-promise-peril-human-artificial-intelligence> accessed 8 July 2024

Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency (2018) 81:1–15 <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf> accessed 8 July 2024

Cadwalladr C and E Graham-Harrison, 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach' (The Guardian, 17 March 2018) <www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> accessed 12 June 2024.

Cantú F, 'The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election' (2019) 113 American Political Science Review 710 <https://doi.org/10.1017/S0003055419000285>

Carsten Stahl B and D Eke, 'The ethics of ChatGPT – Exploring the ethical issues of an emerging technology' (2024) 74 International Journal of Information Management 102700 <https://doi.org/10.1016/j.ijinfomgt.2023.102700>

Casar MA, 'Los gobiernos sin mayoría en México: 1997-2006' [Governments without Majority in Mexico: 1997–2006] (2008) 15(2) Política y gobierno 221 <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-20372008000200001&lng=es&tlng=es> accessed 28 June 2024

Comisión Global sobre Elecciones, Democracia y Seguridad, 'Profundizando la democracia: Una estrategia para mejorar la integridad electoral en el mundo [Deepening Democracy: A Strategy to Improve Electoral Integrity Worldwide]' (Septiembre 2012) <www.idea.int/sites/default/files/publications/profundizando-la-democracia.pdf> accessed 12 June 2024

Crawford K, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial* Intelligence (Yale University Press 2021)

De Remes A, 'Gobiernos yuxtapuestos en México' [2017] 483-Texto del artículo-2082-1-10-20170411 246 <https://revistas.onpe.gob.pe/index.php/elecciones/article/view/110/104> accessed 12 June 2024.

Espinoza A and CA Piña-García, 'Propaganda and Manipulation in Mexico: A Programmed, Coordinated and Manipulative "Pink" Campaign' (2023) 4(2) Journalism and Media 578 <https://doi.org/10.3390/journalmedia4020037>

Freedom House, 'Mexico: Freedom on the Net 2023' (2023) <https://freedomhouse.org/country/mexico/freedom-net/2023#footnoteref6_2a11ufc> accessed 8 July 2024

Gadjanova E, G Lynch, and G Saibu, 'Misinformation Across Digital Divides: Theory and Evidence From Northern Ghana' (2022) 121 African Affairs 161 <https://doi.org/10.1093/afraf/adac009> accessed 28 June 2024

Garcia A and J Martin-Reyes, 'Guardians of Democracy: Battling for the Rule of Law in Mexico' (Stanford Law School, 24 October 2023) <https://law.stanford.edu/2023/10/24/guardians-of-democracy-battling-for-the-rule-of-law-in-mexico/> accessed 28 June 2024

Garnett HA and TS James, 'Cyber Elections in the Digital Age: Threats and Opportunities of Technology for Electoral Integrity' (2020) 19(2) Election Law Journal <www.liebertpub.com/doi/full/10.1089/elj.2020.0633>

Goldstein JA and others, 'How persuasive is AI-generated propaganda?' (2024) 3(2) PNAS Nexus pgae034 <https://doi.org/10.1093/pnasnexus/pgae034>

Goldstein JA, G Sastry, M Musser, R DiResta, M Gentzel, and K Sedova, 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations' (2023) arXiv:2301.04246 <https://arxiv.org/abs/2301.04246> accessed 4 May 2024

González Calleros CB and others, 'Addressing the Digital Divide with Educational Systems in Mexico: Challenges and Opportunities' in Łukasz Tomczyk, Francisco David Guillén-Gámez, Juan Ruiz-Palmero, and Alhassan Habibi (eds), From Digital Divide to Digital Inclusion (Springer, Singapore 2023) <https://doi.org/10.1007/978-981-99-7645-4_16>

Guidetti A, Artificial Intelligence as General Purpose Technology: An Empirical and Applied Analysis of its Perception (Master's Thesis, Università della Valle d'Aosta - Université de la Vallée d'Aoste 2020), p.1 <https://univda.unitesi.cineca.it/bitstream/20.500.14084/428/1/ETI_104_Guidetti_André.pdf> accessed 6 May 2024

Hameleers M, 'Disinformation as a Context-Bound Phenomenon: Toward a Conceptual Clarification Integrating Actors, Intentions and Techniques of Creation and Dissemination' (2022) <https://doi.org/10.1093/ct/qtac021>

Harrison J, 'Establishing a Meaningful Human Rights Due Diligence Process for Corporations: Learning from Experience of Human Rights Impact Assessment' (2013) 31(2) Impact Assessment and Project Appraisal 107, 117 <http://dx.doi.org/10.1080/14615517.2013.774718>

Heikkilä M, 'How to opt out of Meta's AI training' (MIT Technology Review, 14 June 2024) <www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/> accessed 29 June 2024

Heikkilä M. 'Five Things You Need to Know About the EU's New AI Act' (MIT Technology Review, 11 December 2023) <www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/> accessed 12 July 2024

Hendrycks D, M Mazeika, and T Woodside, 'An Overview of Catastrophic AI Risks' (arXiv.org, 21 June 2023) <https://arxiv.org/abs/2306.12001> accessed 5 May 2024

Hirvonen N and others, 'Artificial intelligence in the information ecosystem: Affordances for everyday information seeking' (2023) Journal of the Association for Information Science and Technology <https://doi.org/10.1002/asi.24860>

Jabotinsky HY and R Sarel, 'Co-Authoring with an AI? Ethical Dilemmas and Artificial Intelligence' (2023) SSRN. <https://ssrn.com/abstract=4303959> accessed 4 May 2024

Judge EF and AM Korhani, 'Disinformation, Digital Information Equality, and Electoral Integrity' (Forthcoming in Election Law Journal, 24 February 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518800> accessed 13 July 2024

Kasneci E and others, 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education' (2023) 103 Learning and Individual Differences 102274 <https://doi.org/10.1016/j.lindif.2023.102274>

Keller D, 'Five Big Problems with Canada's Proposed Regulatory Framework for "Harmful Online Content', (Tech Policy Press, August 31, 2021) <https://techpolicy.press/five-big-problems-with-canadas-proposed-regulatory-framework-for-harmful-online-content/?&utm_source=newsletter&utm_medium=email&utm_campaign=automation_and_the_plight_of_the_worker&utm_term=2021-09-05> accessed 29 June 2024

Kertysova K, 'Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered' (2018) 29(1-4) Security and Human Rights 55 <https://doi.org/10.1163/18750230-02901005>

Keyes R, *The Post-Truth Era: Dishonesty and Deception in Contemporary Life* (Macmillan + ORM 2004) <https://books.google.pl/books?id=f0Kvm3KObXoC&redir_esc=y> accessed 30 June 2024

Kielland T, 'Embracing AI in Journalism — The News Carousel' (Medium, 8 November 2023) <https://pub.towardsai.net/embracing-ai-in-journalism-the-news-carousel-dd6b170ce376> accessed 5 May 2024

Klonick K, 'The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression', Yale Law Journal, Vol. 129, No. 2418, 2020, (June 30, 2020) 2418 SSRN: <https://ssrn.com/abstract=3639234> accessed 29 June 2024

Kreps S and D Kriner, 'How AI Threatens Democracy' (2023) 34(4) Journal of Democracy 122 <www.journalofdemocracy.org/articles/how-ai-threatens-democracy/> accessed 12 May 2024

Kreps S, RM McCain, and M Brundage, 'All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation' (2022) 9(1) Journal of Experimental Political Science 104 <https://doi.org/10.1017/XPS.2020.37>

Lance Bennett W and S Livingston (eds), 'The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States' (2021) Cambridge University Press. <https://doi.org/10.1017/9781108914628>

Laufer B and H Nissenbaum, 'Algorithmic Displacement of Social Trust' (Knight First Amendment Institute, 29 November 2023) <https://knightcolumbia.org/content/algorithmic-displacement-of-social-trust> accessed 12 May 2024

Lesch M and N Reiners, 'Informal Human Rights Lawmaking: How Treaty Bodies Use General Comments to Develop International Law' (2023) 12(2) Global Constitutionalism 378-401 <www.cambridge.org/core/journals/global-constitutionalism/article/informal-human-rights-lawmaking-how-treaty-bodies-use-general-comments-to-develop-international-law/7D1E7EF25889DDD944D8FB2691AA36A7> accessed 4 June 2024

Lorenz P, K Perset and J Berryhill, 'Initial Policy Considerations for Generative Artificial Intelligence' (OECD Artificial Intelligence Papers, No 1, 18 September 2023) <https://doi.org/10.1787/fae2d1e6-en>

Malek A, 'Criminal Courts' Artificial Intelligence: The Way it Reinforces Bias and Discrimination' (2022) 2 AI Ethics 233 <https://doi.org/10.1007/s43681-022-00137-9>

McCarthy J and others, 'Dartmouth Summer Research Project on Artificial Intelligence' (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> accessed 8 July 2024

Melvin Kranzberg postulated in 1986, that technology is neither inherently good nor bad, nor is it neutral; its effects are dependent on the social, environmental, and historical context in which it is embedded. His postulations are also known as the Kranzberg's Laws. Kranzberg M, 'Technology and History: 'Kranzberg's Laws'' (1986) 27(3) Technology and Culture 544-560

Millière R, 'Deep Learning and Synthetic Media' (2022) 200 Synthese 231 <https://doi.org/10.1007/s11229-022-03739-2>

Mitchell M, *Artificial Intelligence: A Guide for Thinking Humans* (Penguin Books Limited 2019)

Monsiváis-Carrillo A, 'Populismo, desinformación e integridad electoral en México' (2023) 22(25) Revista Elecciones 151 <https://dx.doi.org/10.53557/elecciones.2023.v22n25.05>

Moreno-Gil V and others, 'Fact-Checking Interventions as Counteroffensives to Disinformation Growth: Standards, Values, and Practices in Latin America and Spain' (2021) 9(1) Media and Communication 251 <https://doi.org/10.17645/mac.v9i1.3443>

Murgia M, *Code Dependent: Living in the Shadow of AI* (Pan Macmillan 2024)

Newman N and others, Digital News Report 2023 (Reuters Institute for the Study of Journalism, 2023) <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf> accessed 8 July 2024

Norris P, 'Are There Global Norms and Universal Standards of Electoral Integrity and Malpractice? Comparing Public and Expert Perceptions' (Harvard University, John F Kennedy School of Government, Faculty Research Working Paper Series RWP12-010, 2012) <https://dash.harvard.edu/bitstream/handle/1/8506826/RWP12-010_Norris.pdf> accessed 12 June 2024

O'Neil C *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016)

Perea González A, 'Justicia predictiva: una solución al 'pleito masa' [Predictive Justice: A Solution to Mass Litigation]' (Cinco Días, 16 March 2021) <https://cincodias.elpais.com/cincodias/2021/03/16/legal/1615930000_454211.html> accessed 9 June 2024.

Polarization Lab, 'Current Research' (Polarization Lab at Duke University, no date) <www.polarizationlab.com/current-research> accessed 31 May 2024.

Roderic Ai Camp, 'Democratizing Mexican Politics, 1982–2012' (2015) Oxford Research Encyclopedias, Latin American History <https://doi.org/10.1093/acrefore/9780199366439.013.12>

Rotman D, 'How to Solve AI's Inequality Problem' (MIT Technology Review, 19 April 2022) <www.technologyreview.com/2022/04/19/1049378/ai-inequality-problem/> accessed 8 July 2024

Rotolo D, D Hicks and B Martin, 'What Is an Emerging Technology?' accessed 12 July 2024 <https://doi.org/10.1016/j.respol.2015.06.006>

Sargeant H and others, *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual* (Organization for Security and Co-operation in Europe, 17 March 2022) <www.osce.org/representative-on-freedom-of-media/510332> accessed 8 July 2024

Schuilenburg M and R Peeters (eds), The Algorithmic Society: Technology, Power, and Knowledge (1st edn, Routledge 2020) <https://doi.org/10.4324/9780429261404>

Shailer D, 'The official count of disappeared people in Mexico could be an underestimate, say UN and advocates' (AP News, 4 October 2023) <https://apnews.com/article/mexico-missing-disappearances-united-nations-147b08e445c715fe0ee487a5b0787288> accessed 12 June 2024

Sherman J, 'Human Rights Due Diligence and Corporate Governance' (29 April 2022) Human Rights Due Diligence for Lawyers, American Bar Association, forthcoming, available at SSRN: <https://ssrn.com/abstract=3862624> accessed 8 July 2024 or <http://dx.doi.org/10.2139/ssrn.3862624>

Simon FM, S Altay, and H Mercier, 'Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown' (HKS Misinformation Review, 2023) <https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/> accessed 12 July 2024

Singh P, M Patidar, and L Vig, 'Translating Across Cultures: LLMs for Intralingual Cultural Adaptation' (20 June 2024) arXiv:2406.14504 <https://arxiv.org/abs/2406.14504> accessed 12 June 2024

Smuha N and others, 'We Are Not Ready for Manipulative AI – Urgent Need for Action' (Euractiv, 2023) <https://lirias.kuleuven.be/handle/20.500.12942/717687> accessed 22 June 2024

Soloaga I and others, 'Lo rural y lo urbano en México Una nueva caracterización a partir de estadísticas nacionales [The rural and the urban in Mexico A new characterisation based on from national statistics]' Economic Commission for Latin America and the Caribbean (ECLAC, 2022) <https://repositorio.cepal.org/server/api/core/bitstreams/27f4bef7-e9f0-4d61-8baa-7bd1fdc26675/content> accessed 30 June 2024

Stiff H and F Johansson, 'Detecting Computer-Generated Disinformation' (2021) 13 International Journal of Data Science and Analytics 363 <https://doi.org/10.1007/s41060-021-00299-5>

Suleyman M and M Bhaskar, *The Coming Wave: Technology, Power and the 21st Century's Dilemma* (Crown 2023)

Tamkin A and others, 'Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models' (2021) arXiv preprint arXiv:2102.02503 <https://arxiv.org/abs/2102.02503>

Tandoc Jr EC and others, 'Defining 'Fake News'' (2018) 6(2) Digital Journalism 137, <https://doi.org/10.1080/21670811.2017.1360143>

Torres Rodríguez ID and CE Ahuactzin Martínez, 'Democracy and Electoral Reforms in Mexico' (2019) 4(11) Derecho Global. Estudios sobre Derecho y Justicia 143-162 <http://www.derechoglobal.cucsh.udg.mx/index.php/DG/article/view/186/245> accessed 22 June 2024 158

Townsen Hicks M, J Humphries, and J Slater, 'ChatGPT is Bullshit' (2024) 26 Ethics Inf Technol 38 <https://doi.org/10.1007/s10676-024-09775-5>

Valdés Zurita L, 'El sistema electoral mexicano: equidad en la competencia, inequidad en la representación [The Mexican Electoral System: Fairness in Competition, Inequity in Representation]' (2021) 20(21) Elecciones 15-42 16

Valenzuela S and others, 'Social Media and Belief in Misinformation in Mexico: A Case of Maximal Panic, Minimal Effects?' (2022) 29(3) The International Journal of Press/Politics 8 <https://doi.org/10.1177/19401612221088988>

Van der Haar G, 'The Zapatista Uprising and the Struggle for Indigenous Autonomy' (2004) 76 Revista Europea de Estudios Latinoamericanos y Del Caribe / European Review of Latin American and Caribbean Studies 99 <http://www.jstor.org/stable/25676074> accessed 28 June 2024

Vosoughi S, D Roy, and S Aral, 'The Spread of True and False News Online' (2018) 359 Science 1146. Available at <www.science.org/doi/10.1126/science.aap9559>

Wack M, D Linvill, and P Warren, 'Old Despots, New Tricks - An AI-Empowered Pro-Kagame/RPF Coordinated Influence Network on X' (Clemson University, June 2024) <https://tigerprints.clemson.edu/mfh_reports/5> accessed 22 June 2024

Wardle C and H Derakhshan, Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making (Council of Europe report DGI(2017)09, 2017) 20 <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 22 June 2024

Xu D, S Fan, and M Kankanhalli, 'Combating Misinformation in the Era of Generative AI Models' (Proceedings of the 31st ACM International Conference on Multimedia, October 29-November 3, 2023, Ottawa, ON, Canada) <https://doi.org/10.1145/3581783.3612704>

Zuiderveen Borgesius F, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making' (Council of Europe, 2018) <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> accessed 5 May 2024

Official Documents

Comisión Interamericana de Derechos Humanos (CIDH), 'Marco Jurídico Interamericano del Derecho a la Libertad de Expresión' ('Inter-American Legal Framework on the Right to Freedom of Expression') (Organización de los Estados Americanos, 2009) parr 32-56, <www.oas.org/es/cidh/expresion/docs/publicaciones/MARCO%20JURIDICO%20INTERAMERICANO%20DEL%20DERECHO%20A%20LA%20LIBERTAD%20DE%20EXPRESION%20ESP%20FINAL%20portada.doc.pdf> accessed 10 June 2024

Council of Europe Committee of Ministers, 'Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' (Adopted by the Committee of Ministers on 17 May 2024 at the 133rd Session of the Ministers' Deputies) <https://search.coe.int/cm?i=0900001680afb11f> accessed 8 July 2024

Council of Europe, 'The Convention in 1950 - The European Convention on Human Rights' (Council of Europe, 1950) <www.coe.int/en/web/human-rights-convention/the-convention-in-1950> accessed 22 June 2024

European Union Agency for Fundamental Rights, Bias in Algorithms: Artificial Intelligence and Discrimination (Vienna, 2022) <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf> accessed 8 July 2024

House of Commons Digital, Culture, Media and Sport Committee, Disinformation and 'Fake News': Final Report: Government Response to the Committee's Eighth Report of Session 2017–19 Seventh Special Report of Session 2017–19 (Ordered by the House of Commons to be printed 8 May 2019) <https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/2184/2184.pdf> accessed 8 July 2024

Human Rights Committee, 'General Comment No 25 on Article 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service' (1996) CCPR/C/21/Rev1/Add.7 [Hereinafter CCPR/C/21/Rev1/Add.7] <https://undocs.org/Home/Mobile?FinalSymbol=CCPR%2FC%2F21%2FRev.1%2FAdd.7&Language=E&DeviceType=Desktop&LangRequested=False> accessed 22 June 2024

–, 'General Comment No 34 on Article 19: Freedoms of Opinion and Expression' (12 September 2011) [Hereinafter CCPR/C/GC/34] <https://undocs.org/Home/Mobile?FinalSymbol=a%2FHRC%2F26%2F30&Language=E&DeviceType=Desktop&LangRequested=False> accessed 22 June 2024

INE, '#Certeza 2024 - Central Electoral' (Central Electoral, 2024) <https://centralelectoral.ine.mx/certeza/> accessed 14 June 2024

–, 'Historia del Instituto Federal Electoral [History of the Federal Electoral Institute]' (Instituto Nacional Electoral) <https://portalanterior.ine.mx/archivos3/portal/historico/contenido/menuitem.cdd858023b32d5b7787e6910d08600a0/> accessed 28 June 2024

INE, 'Concluyen campañas electorales e inicia periodo de reflexión [Electoral campaigns conclude and reflection period begins]' (Central Electoral, 29 May 2024) <https://centralelectoral.ine.mx/2024/05/29/concluyen-campanas-electorales-e-inicia-periodo-de-reflexion/> accessed 14 June 2024

--, 'Estadísticas de la Lista Nominal y Padrón Electoral [Statistics of the Nominal List and Electoral Register]' <https://portal.ine.mx/credencial/estadisticas-lista-nominal-padron-electoral/> accessed 20 May 2024

--, 'Metodología Certeza 2024' (Central Electoral, 2024) <https://centralelectoral.ine.mx/wp-content/uploads/2024/03/Metodologi%CC%81a_Certeza-2024.pdf> accessed 14 June 2024

--, Numeralia del Proceso Electoral Federal 2023-2024 (2023) <https://repositoriodocumental.ine.mx/xmlui/bitstream/handle/123456789/153578/Numeralia-PEF-2023-2024.pdf> accessed 22 June 2024

--, 'Padrón Electoral y Lista Nominal de Electores [Electoral Register and Nominal List of Voters]' <https://ine.mx/padron-electoral-lista-nominal-electores/> accessed 20 May 2024

--, 'Se reúnen Consejeras y Consejeros del INE con representantes de Meta [INE Councilors meet with Meta representatives]' Central Electoral (19 September 2023) <https://centralelectoral.ine.mx/2023/09/19/se-reunen-consejeras-y-consejeros-del-ine-con-representantes-de-meta/> accessed 28 June 2024

Instituto Nacional de Estadística y Geografía (INEGI), 'Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares 2023 [National Survey on the Availability and Use of Information Technologies in Households 2023]' (INEGI, 2024) [Hereinafter ENDUTIH] <www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENDUTIH/ENDUTIH_23.pdf> accessed 28 June 2024

--, 'INEGI [National Institute of Statistics and Geography]' <www.inegi.org.mx/default.html> accessed 28 June 2024

OECD, 'Explanatory memorandum on the updated OECD definition of an AI system' (OECD Artificial Intelligence Papers No. 8, OECD Publishing, Paris 2024) <https://doi.org/10.1787/623da898-en>

--, 'The State of Implementation of the OECD AI Principles Four Years On' (OECD Artificial Intelligence Papers, No. 3, OECD Publishing, Paris, 2023) <https://doi.org/10.1787/835641c9-en>

Office of the High Commissioner for Human Rights, 'The Corporate Responsibility to Respect Human Rights: An Interpretive Guide' (United Nations 2012) <www.ohchr.org/sites/default/files/Documents/Publications/HR.PUB.12.2_En.pdf> accessed 17 June 2024

Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, Department of Electoral Cooperation and Observation, and Department of International Law of the General Secretariat of the Organization of American States, Guide to Guarantee Freedom of Expression Regarding Deliberate Disinformation in Electoral Contexts (OAS Official Records OEA/Ser.D/XV.22, OEA/Ser.G CP/CAJP/INF.652/19, 2019) 13 <www.oas.org/en/iachr/expression/publications/Guia_Desinformacion_VF%20ENG.pdf> accessed 8 July 2024

Organization of American States, 'American Convention on Human Rights' (22 November 1969) <https://cidh.oas.org/Basicos/English/Basic3.American%20Convention.htm> accessed 7 June 2024

RELE-CIDH, 'La CIDH advierte un punto de inflexión de la libertad de expresión en internet y convoca a diálogo en la región [IACHR warns of a turning point for freedom of expression on the internet and calls for dialogue in the region]' (5 February 2021) <www.oas.org/es/CIDH/jsForm/?File=/es/cidh/prensa/comunicados/2021/026.asp> accessed 29 June 2024.

The Royal Society, 'Science in the Age of AI: How Artificial Intelligence is Changing the Nature and Method of Scientific Research' (May 2024) <https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf> accessed 12 June 2024

UN AI Advisory Body, 'Interim Report: Governing AI for Humanity' (2023) <www.un.org/en/ai-advisory-body> accessed 12 June 2024 11

UNESCO, "Your Opinion Doesn't Matter, Anyway": Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI (2023, UNESCO) <https://unesdoc.unesco.org/ark:/48223/pf0000387483> accessed 22 June 2024

UNESCO, Elections in Digital Times: A Guide for Electoral Practitioners (UNESCO 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000387339> accessed 22 June 2024

--, Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multi-Stakeholder Approach (2023, UNESCO) <https://unesdoc.unesco.org/ark:/48223/pf0000387339> accessed 22 June 2024

United Nations, 'Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' [hereinafter UNGPs] (New York and Geneva, 2011) <www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf> accessed 23 June 2024

--, 'Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests' (2020) <https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F44%2F24&Language=E&DeviceType=Desktop&LangRequested=False> accessed 5 May 2024

--, 'Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms' (June 2023) <www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf> accessed 10 July 2024

--, 'Urgent Action Needed over Artificial Intelligence Risks to Human Rights' (UN News, 17 September 2021) <https://news.un.org/en/story/2021/09/1099972> accessed 5 May 2024

--, 'Vienna Declaration and Programme of Action' (25 June 1993) <www.ohchr.org/en/instruments-mechanisms/instruments/vienna-declaration-and-programme-action> accessed 12 June 2024

United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Cooperation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. Joint declaration on freedom of expression and 'fake news', disinformation and propaganda. (2017). OSCE. <www.osce.org/fom/302796> accessed 30 June 2024

United Nations General Assembly, 'International Covenant on Civil and Political Rights' (OHCHR, 16 December 1966) <www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights> accessed 13 April 2024

--, 'Universal Declaration of Human Rights' (UN, 10 December 1948) <www.un.org/en/about-us/universal-declaration-of-human-rights> accessed 13 April 2024

United Nations Human Rights Council, 'Disinformation and Freedom of Opinion and Expression' (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, 13 April 2021) UN Doc A/HRC/47/25 [Hereinafter A/HRC/47/25] <www.undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F47%2F25&Language=E&DeviceType=Desktop&LangRequested=False> accessed 12 June 2024

--, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27 <www.ohchr.org/sites/default/files/Documents/Issues/Opinion/A-HRC-17-27.pdf> accessed 12 June 2024

--, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' (2 July 2014) UN Doc A/HRC/26/30 [hereinafter A/HRC/26/30] para 7 3 <https://undocs.org/Home/Mobile?FinalSymbol=a%2FHRC%2F26%2F30&Language=E&DeviceType=Desktop&LangRequested=False> accessed 12 June 2024.

United Nations Human Rights Office of the High Commissioner, 'Access to Remedy and the Technology Sector: Basic Concepts and Principles' (2023) <www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf> accessed 10 July 2024

--, 'B-Tech Project' (OHCHR and Business and Human Rights) <www.ohchr.org/en/business-and-human-rights/b-tech-project> accessed 8 July 2024

--, 'Designing and Implementing Effective Company-Based Grievance Mechanisms: A B-Tech Foundational Paper' (2021) <www.ohchr.org/Documents/Issues/Business/B-Tech/access-to-remedy-company-based-grievance-mechanisms.pdf> accessed 10 July 2024

United Nations Human Rights Office of the High Commissioner, Taxonomy of Human Rights Risks Connected to Generative AI: Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI (2024) <www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf> accessed 8 July 2024

United Nations Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media and OAS Special Rapporteur on Freedom of Expression, 'Joint Declaration on Freedom of Expression and Elections in the Digital Age' (30 April 2020) <www.osce.org/files/f/documents/9/8/451150_0.pdf> accessed 30 June 2024

US Department of Justice, Report On The Investigation Into Russian Interference In The 2016 Presidential Election Volume I of II (Special Counsel Robert S Mueller, III, Washington, DC, March 2019) <www.justice.gov/d9/report.pdf> accessed 8 July 2024 14-19

US Senate Select Committee on Intelligence, 'Russian Active Measures Campaigns and Interference in the 2016 US Election, Volume II: Russia's Use of Social Media with Additional Views' (2020) <www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf> accessed 12 June 2024

World Health Organization, Ethics and Governance of Artificial Intelligence for Health (WHO, 2021) <www.who.int/publications/i/item/9789240029200> accessed 8 July 2024

Legislation

Directive (EU) 2024/1760 of the European Parliament and of the Council of 13 June 2024 on corporate sustainability due diligence and amending Directive (EU) 2019/1937 and Regulation (EU) 2023/2859 [2024] OJ L1760/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401760>

European Parliament, 'EU AI Act: First Regulation on Artificial Intelligence' (European Parliament, 1 June 2023) <www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> accessed 7 May 2024

Case Law

Bowman v United Kingdom App no 24839/94 (ECHR, 19 February 1998)

Cheltsova v Russia App no 44294/06 (ECHR, 13 June 2017)

Compulsory Membership of Journalists, Advisory Opinion OC-5/85 (IACHR, 13 November 1985)

Delfi AS v Estonia App no 64569/09 (ECHR, 16 June 2015)

Długołęcki v Poland App no 23806/03 (ECHR, 24 February 2009)

Féret v Belgium App no 15615/07 (ECHR, 16 July 2009, Section II)

Kwiecień v Poland App no 51744/99 (ECHR, 9 January 2007)

López Lone y otros vs Honduras (IACHR, 5 October 2015) Serie C No 302

Marina Koktish v Belarus' (3 November 2014) UN Doc CCPR/C/111/D/1985/2010

Mathieu-Mohin and Clerfayt v Belgium, App no 9267/81 (ECHR, 2 March 1987)

Orlovskaya Iskra v Russia App no 42911/08 (ECHR, 21 February 2017)

Ricardo Canese vs Paraguay (IACHR, 31 August 2004) Series C No 111

Savva Terentyev v Russia App no 10692/09 (ECHR 28 August 2018)

Internet Sources

Aguirre S, "'Las máquinas aprenden': Inteligencia Artificial evoluciona y puede usarse para engañar, pero no todo está perdido ['Machines learn': Artificial Intelligence evolves and can be used to deceive, but all is not lost]," April 23, 2023, <www.animalpolitico.com/verificacion-de-hechos/te-explico/inteligencia-artificial-evoluciona-desinformacion-herramientas> accessed 8 April 2024

Al Jazeera, 'EU launches disinformation probe against social media giant Meta' (Al Jazeera, 30 April 2024) <www.aljazeera.com/news/2024/4/30/eu-opens-probe-against-social-media-giant-meta-over-disinformation> accessed 8 May 2024

Animal Político, Data Cívica, México Evalúa, Democracia Vulnerada: El Crimen Organizado en las Elecciones y la Administración Pública en México [Democracy Undermined: Organized Crime in Elections and Public Administration in Mexico] (2024) <https://votar-entre-balas.datacivica.org/reportes> accessed 22 June 2024.

--, 'Red Brolan: youtubers afines a AMLO difunden desinformación política y contra periodistas [Brolan Network: AMLO-friendly youtubers spread political and anti-journalist disinformation]' Animal Político (19 April 2023) <www.animalpolitico.com/verificacion-de-hechos/te-explico/brolan-difunde-desinformacion-youtube> accessed 28 June 2024.

--, TikTok, April 6, 2023, <www.tiktok.com/@elsabuesoap/video/7218968175268859142> accessed 5 May 2024

--, 'Verificación de Hechos' (Animal Político, 2024) <www.animalpolitico.com/verificacion-de-hechos> accessed 28 June 2024.

Antonio, '¡QUE SE HAGA VIRAL! Si el PAN obtiene menos de 7 millones de votos, no podrán llegar al Congreso los plurinominales: - Cabeza de Vaca - Ricardo Anaya - Marko Cortés Mendoza [MAKE IT GO VIRAL! 📢 If the PAN gets less than 7 million votes, the proportional representatives: - Cabeza de Vaca - Ricardo Anaya - Marko Cortés Mendoza]' (X, 14 April 2024) <https://x.com/JTonyy35/status/1779626643629150662> accessed 29 June 2024

AP News, 'AP Fact Check' (AP News) <https://apnews.com/ap-fact-check> accessed 28 June 2024.

AP News, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef-1f5e9ee242d495570578d5ed> accessed 28 June 2024.

--, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef1f5e9ee242d-495570578d5ed> accessed 28 June 2024.

--, 'Mexican senator seeks to break male dominance and become first female president' (AP News, 27 June 2023) <https://apnews.com/article/mexico-politics-elections-2024-xochitl-galvez-nominee-8df70cef1f5e9ee242d-495570578d5ed> accessed 28 June 2024

AP, 'CEO Zuckerberg apologizes for Facebook's privacy failures' (Breitbart, April 10, 2018) <www.breitbart.com/news/ceo-zuckerberg-apologizes-for-facebooks-privacy-failures/> accessed 29 June 2024

APC, "APC policy explainer: Platform responsibility and accountability", Association for Progressive Communications, November 1, 2020 <www.apc.org/en/pubs/apc-policy-explainer-platform-responsibility-and-accountability> accessed 29 June 2024

Applio, 'Models' (Applio, 2024) <https://applio.org/models> accessed 30 June 2024

ARTICLE 19, '(Des)información oficial y comunicación social' [(Official Disinformation and Social Communication)] (Artículo 19, 14 March 2023) <https://articulo19.org/desinformacion-oficial-y-comunicacion-social/> accessed 8 July 2024

--, 'Malaysia: Repeal "Fake News" Emergency Ordinance' (ARTICLE 19, 15 March 2021) <www.article19.org/resources/malaysia-fake-news-ordinance/> accessed 8 July 2024

--, 'Negación [Denial]' (Artículo 19, 2023) <https://articulo19.org/negacion/> accessed 28 June 2024

ARTICLE 19, 'Senegal: “Fake news” and disinformation laws threaten freedom of expression' (ARTICLE 19, 17 January 2024) <www.article19.org/resources/senegal-fake-news-and-disinformation-laws-threaten-freedom-of-expression/> accessed 8 July 2024

--, “#InternetBajoAtaque: La regulación de las redes sociales como mecanismo de control”, Febrero 2021, <https://articulo19.org/wp-content/uploads/2021/02/Article19_2021-PosicionamientoInternet_v3.pdf> accessed 8 July 2024

Ashley G and Sara F, 'Chatbots trigger next misinformation nightmare' (Axios, 21 February 2023) <www.axios.com/2023/02/21/chatbots-misinformation-nightmare-chatgpt-ai> accessed 12 July 2024

Bachoco, “Mensaje importante a toda la comunidad: [Important message to the entire community:], Facebook, Sepmtember 22, 2023, <www.facebook.com/BachocoMX/posts/681164897379600> accessed 29 June 2024

Bai Y and others, 'Constitutional AI: Harmlessness from AI Feedback' (arXiv, 15 December 2022) <https://arxiv.org/abs/2212.08073> accessed 30 June 2024

Bedingfield W, 'A Chatbot Encouraged Him to Kill the Queen. It’s Just the Beginning' (WIRED, 18 October 2023) <www.wired.com/story/chatbot-kill-the-queen-eliza-effect/?redirectURL=%2Fstory%2Fchatbot-kill-the-queen-eliza-effect%2F> accessed 30 June 2024

Bickert M, 'Our Approach to Labeling AI-Generated Content and Manipulated Media' (Meta, 17 April 2024) <https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/> accessed 29 June 2024

Buendía E and others, 'Neurona: La fábrica de engaño para las izquierdas en América Latina [Neurona: The deception factory for the left in Latin America]' El Clip (31 July 2023) <www.elclip.org/neurona-la-fabrica-de-engano-para-las-izquierdas-en-america-latina/> accessed 28 June 2024

Calderón C, 'Elecciones México 2024: Deep Fakes y Fake News Ganan en las Votaciones [Mexico Elections 2024: Deep Fakes and Fake News Win in the Votes]' (El Financiero, 4 June 2024, 3:00 am) <www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/> accessed 13 July 2024

Camhaji E, 'López Obrador va por la eliminación de siete órganos autónomos y entes reguladores [López Obrador is going for the elimination of seven autonomous bodies and regulatory entities]' El País (6 February 2024) <https://elpais.com/mexico/2024-02-06/lopez-obrador-va-por-la-eliminacion-de-siete-organos-autonomos-y-entes-reguladores.html> accessed 28 June 2024

--, 'Movimiento Ciudadano, a las puertas de la elección más importante de su historia' El País (14 January 2024) <https://elpais.com/mexico/elecciones-mexicanas/2024-01-14/movimiento-ciudadano-a-las-puertas-de-la-eleccion-mas-importante-de-su-historia.html> accessed 28 June 2024

Canales MP and I Barber, 'What Would a Human Rights-Based Approach to AI Governance Look Like?' (Global Partners Digital, 19 September 2023) <www.gp-digital.org/what-would-a-human-rights-based-approach-to-ai-governance-look-like/> accessed 8 July 2024

Center for Countering Digital Hate, 'Fake Image Factories: How AI Image Generators Threaten Election Integrity' (Center for Countering Digital Hate, 6 March 2024) <https://counterhate.com/research/fake-image-factories/> accessed 8 July 2024.

Coqui, 'XTTS-v2' (Hugging Face, 2024) <https://huggingface.co/coqui/XTTS-v2> accessed 30 June 2024

Daen A and TL Montalvo, ‘Infodemia y Quién es Quién, más propaganda que chequeo con recursos públicos en México [Infodemic and Who's Who, more propaganda than checking with public resources in Mexico]’ Animal Político (Centro Latinoamericano de Investigación Periodística, 30 November 2022) <www.elclip.org/infodemia-y-quien-es-quien-propaganda-desinformacion-mexico/> accessed 28 June 2024

Daen A 'También en la oposición: canales de YouTube desinforman sobre AMLO y Sheinbaum, pero elogian a Xóchitl Gálvez [Also in opposition: YouTube channels misinform about AMLO and Sheinbaum, but praise Xóchitl Gálvez]' (Animal Político, 22 February 2024) <www.animalpolitico.com/verificacion-de-hechos/te-explico/canales-desinforman-morena-amlo-xochitl> accessed 8 July 2024

--, S Aguirre, and S Flores, '"Liga de Guerreros": Red de cuentas en Twitter que respalda a Taboada y Xóchitl desinforma y usa violencia política [League of Warriors': Network of Twitter accounts backing Taboada and Xóchitl misinforms and uses political violence]' (Animal Político, 30 October 2023) <https://animalpolitico.com/verificacion-de-hechos/te-explico/liga-de-guerreros-desinformacion-taboada-xochitl> accessed 8 July 2024

DW, 'Brasil: Jair Bolsonaro prohíbe a redes sociales quitar contenidos [Brazil: Jair Bolsonaro prohibits social networks from removing content]' (7 September 2021) <www.dw.com/es/brasil-jair-bolsonaro-proh%C3%ADbe-a-redes-sociales-quitar-contenidos/a-59105724> accessed 29 June 2024

edsonstartingnet, 'Nuevo e increíble método para hacer tus deeps... [New and amazing method for making your deeps...]' (TikTok, 8 March 2024) <www.tiktok.com/@edsonstartingnet/video/7344128059370458373> accessed 29 June 2024

--, 'Video' (TikTok, 6 January 2023) <www.tiktok.com/@edsonstartingnet/video/7320759295358881029> accessed 29 June 2024

El Clip, 'Mercenarios digitales [Digital Mercenaries]' El Clip (2023) <www.elclip.org/mercenarios-digitales/> accessed 28 June 2024

ElFranky_, 'Unas caguamas bien heladas al que lo hizo [A few ice-cold beers to the one who did it]' (X, 27 March 2024) <https://x.com/ElFranky_/status/1773039123499995226> accessed 5 May 2024

EP, 'La CNMV alerta del fraude de Quantum AI por usar imágenes de famosos para publicitarse en redes sociales [The CNMV alerts about the fraud of Quantum AI for using images of celebrities to advertise on social networks]' (El País, 12 December 2023) <https://elpais.com/economia/2023-12-12/la-cnmv-alerta-del-fraude-de-quantum-ai-por-usar-imagenes-de-famosos-para-publicitarse-en-redes-sociales.html> accessed 29 June 2024

Ferri P, 'The rise of Morena: In one decade, the political party has wiped the PRI from Mexico's electoral map' El País (Madrid, 4 June 2023) <https://english.elpais.com/international/2023-06-04/the-rise-of-morena-in-one-decade-the-political-party-has-wiped-the-pri-from-mexicos-electoral-map.html> accessed 28 June 2024

FlexOS, 'Generative AI Top 150: The World's Most Used AI Tools' (29 January 2024) <www.flexos.work/learn/generative-ai-top-150> accessed 30 June 2024

Galvez X, 'Con razón no te salen las cuentas, Sheinbaum. En tu foto sales con 6 dedos en cada mano [No wonder you don't figure it out, Sheinbaum. In your photo you appear with 6 fingers on each hand]' (X, 29 April 2024) <https://x.com/XochitlGalvez/status/1784822900672893328> accessed 12 May 2024

--, 'Les presento a mi nueva vocería de inteligencia artificial. Una herramienta pionera e innovadora que únicamente será oficial a través de mis redes sociales: iXóchitl ['I present to you my new artificial intelligence spokesperson. A pioneering and innovative tool that will only be official through my social networks: iXóchitl]' (X, 17 December 2023) <https://x.com/XochitlGalvez/status/1736511528130478348> accessed 12 May 2024

Gómez J, 'Máynez: La maquinaria detrás del fenómeno Millones en estrategia digital y giras por el país [Máynez: The Machinery Behind the Phenomenon Millions in Digital Strategy and Country Tours]' Fábrica de Periodismo (22 May 2024) <https://fabricadeperiodismo.com/reportajes/maynez-maquinaria-electoral-millones-campana/> accessed 28 June 2024

Google News Initiative, 'Proyectos seleccionados [Selected Projects]' (Google News Initiative, 2018) <https://20190322t164649-dot-gweb-news-initiative.appspot.com/intl/es/innovation-challenges/funding/latin-america/> accessed 14 June 2024

Grey J, 'How to Use Discord: A Beginner's Guide' (WIRED, 3 June 2024) <www.wired.com/story/how-to-use-discord/> accessed 8 July 2024

Grupo Fórmula, 'Simpatizantes de Xóchitl Gálvez crearon un video con IA luego de las declaraciones de AMLO sobre que ella sería la candidata de la oposición [Supporters of Xóchitl Gálvez created an AI video following AMLO's statements about her being the opposition candidate]' (X, 4 July 2023) <https://x.com/Radio_Formula/status/1676279234992414731> accessed 5 May 2024

Hamilton M and P Ugwudike, 'A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up' (The Conversation, 26 July 2023) <https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594> accessed 8 July 2024

Harbath K, 'Different Approaches to Counting Elections' (Anchor Change, 2022) <https://anchorchange.substack.com/p/different-approaches-to-counting> accessed 5 May 2024

Harris DE and L Norden, 'Meta's AI Watermarking Plan Is Flimsy, at Best' (IEEE Spectrum, 4 March 2024) <https://spectrum.ieee.org/meta-ai-watermarks> accessed 10 July 2024

Heaven WD, 'Geoffrey Hinton tells us why he's now scared of the tech he helped build' (MIT Technology Review, 2 May 2023) <www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/> accessed 30 June 2024

Helen VL, '#IXOCHITL: Construyendo el Futuro de México [#IXOCHITL: Building the Future of Mexico]' (X, 2 July 2023) <https://x.com/HelenVL6/status/1675511645869842433> accessed 5 May 2024

Hoover A, 'An Eating Disorder Chatbot Is Suspended for Giving Harmful Advice' (Wired, 1 June 2023) <www.wired.com/story/tessa-chatbot-suspended/> accessed 12 May 2024

Hsu T, SA Thompson, and SLee Myers, 'Election Disinformation 2024' (The New York Times, 9 January 2024) <www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html> accessed 5 May 2024

IBM, 'AI Inference' (IBM, 2024) <www.ibm.com/think/topics/ai-inference> accessed 13 June 2024

Internet Society, 'Gobernanza de Internet [Internet Governance]' (30 October 2015) <www.internetsociety.org/es/policybriefs/internet-governance/> accessed 8 May 2024

Janetsky M, 'Mexico is About to Have Its Biggest Election Ever. Here's What to Know' (Associated Press, 1 March 2024) <www.apnews.com/article/mexico-elections-2024-what-to-know-d104184b02bf5bcf9e08f570a5ba37e2> accessed 4 June 2024

Kaminska I, 'Cambridge Analytica Probe Finds No Evidence it Misused Data to Influence Brexit' (Financial Times, 7 October 2020) <www.ft.com/content/aa235c45-76fb-46fd-83da-0bdf0946de2d> accessed 22 June 2024

Lagos A, 'Meta se prepara para las elecciones del próximo 2 de junio en México con estas acciones [Meta prepares for the upcoming June 2 elections in Mexico with these actions]' WIRED (16 April 2024) <https://es.wired.com/articulos/meta-se-prepara-para-las-elecciones-del-proximo-2-de-junio-en-mexico-con-estas-acciones> accessed 28 June 2024

Lajka A, 'New AI Voice-Cloning Tools "Add Fuel" to Misinformation Fire' (AP News, 11 February 2023) <https://apnews.com/article/technology-science-fires-artificial-intelligence-misinformation-26cabd20dcacbd68c8f38610fec39f5b>, accessed 5 May 2024

Leibowicz C, 'Why watermarking AI-generated content won't guarantee trust online' (MIT Technology Review, 9 August 2023) <www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/> accessed 10 July 2024

López Cruz A, 'Popularidad de MC en redes, ¿artificial? [Popularity of MC on social media, artificial?]' El Universal (1 January 2024) <www.eluniversal.com.mx/elecciones/popularidad-de-mc-en-redes-artificial/> accessed 16 June 2024

Lovens PF, 'Sans ces Conversations avec le Chatbot Eliza, Mon Mari Serait Toujours Là' (La Libre Belgique, 28 March 2023, 6:35 am, updated 7:06 am) <www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPDST24/> accessed 13 July 2024

Luma, 'Luma Dream Machine' (Luma, 2024) <https://lumalabs.ai/dream-machine> accessed 30 June 2024

Mac R, 'Mark Zuckerberg Sent An Apology Letter About Myanmar. These NGOs Called It "Grossly Insufficient' (Buzzfeed News, April 9, 2018) <www.buzzfeednews.com/article/ryanmac/mark-zuckerbeg-apology-myanmar-ngos-insufficient> accessed 29 June 2024

Majewski T, 'It's time to retire the term "user"' (MIT Technology Review, 19 April 2024) <www.technologyreview.com/2024/04/19/1090872/ai-users-people-terms/> accessed 30 June 2024.

Malik A, 'OpenAI's ChatGPT Now Has 100 Million Weekly Active Users' (TechCrunch, 6 November 2023, 10:49 am PST) <https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/?guccounter=1> accessed 13 July 2024

Martínez A, 'Fraude en venta de acciones de Pemex emplea imágenes y voz falsas de Claudia Sheinbaum [Fraud in Pemex stock sale uses fake images and voice of Claudia Sheinbaum]' Infobae, 2 January 2024 <https://shorturl.at/izEOX> accessed 29 June 2024

Mashkoor L "Sheikh Jarrah content takedowns reveal pattern of online restrictions in Palestine", May 10, 2021, <www.thenationalnews.com/mena/sheikh-jarrah-content-takedowns-reveal-pattern-of-online-restrictions-in-palestine-1.1220037> accessed 29 June 2024

Maximoam, '#Exclusiva "YO TENÍA TODAS LAS NARCOTIENDITAS"... [#Exclusive "I HAD ALL THE DRUG STORES"...]' (X, 22 May 2024) <https://x.com/maximoam_/status/1793337455824802018> accessed 29 June 2024

Meta, 'June 2021 Coordinated Inauthentic Behavior Report' (Meta, June 2021) <https://about.fb.com/wp-content/uploads/2021/07/June-2021-CIB-Report-Final.pdf> accessed 28 June 2024

--, 'Recapitulamos nuestras acciones contra el comportamiento inauténtico coordinado en 2022 [We recap our actions against coordinated inauthentic behavior in 2022]' (Meta, 20 December 2022) <https://about.fb.com/ltam/news/2022/12/recapitulamos-nuestras-acciones-contra-el-comportamiento-inautentico-coordinado-en-2022/> accessed 28 June 2024

--, 'How Meta uses information for generative AI models' (Meta, 2024) <www.facebook.com/privacy/genai/> accessed 29 June 2024

Meyer-Resende M, A Davis, O Denkovski, and D Allen, 'Are Chatbots Misinforming Us About the European Elections? Yes.' (Democracy Reporting International, 11 April 2024) <https://democracy-reporting.org/en/office/global/publications/chatbot-audit> accessed 8 July 2024

Midjourney, 'About' (Midjourney, 2024) <www.midjourney.com/home> accessed 30 June 2024

Mozilla Foundation, 'Open Letter to Meta: Support CrowdTangle Through 2024 and Maintain CrowdTangle Approach' (Mozilla Foundation, 2023) <https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/> accessed 28 June 2024

Muller J, 'Advanced Generative Summarization Techniques: A deep dive with example code' (Medium, 2024) <https://medium.com/@flux07/advanced-generative-summarization-techniques-939605601fba> accessed 12 June 2024

Musil S, 'Zuckerberg apologizes for data scandal in full-page ads' (CNET, 26 March 2018) <www.cnet.com/news/zuckerberg-apologizes-for-data-scandal-in-full-page-ads/> accessed 29 June 2024

Nación321, '"¡Awilsoooon!"... [From storm to tropical storm, Anaya 'castaway' survives AMLO's term and sends support to Xóchitl Gálvez from exile and AI]' (X, 31 August 2023) <https://x.com/Nacion321/status/1697300718213018066> accessed 29 June 2024

Nogueda Romero O, 'Claudia Sheinbaum no llamó a invertir en petróleo, el video fue manipulado [Claudia Sheinbaum did not call for investment in oil, video was manipulated]' Animal Político, 1 December 2023 <https://animalpolitico.com/verificacion-de-hechos/desinformacion/sheinbaum-invertir-petroleo-falso> accessed 29 June 2024.

noporsuave, 'Elecciones presidenciales en Balenciaga [Presidential elections in Balenciaga]' (TikTok, 29 May 2024) <www.tiktok.com/@noporsuave/video/7374645960154813701> accessed 29 June 2024

Olmos JG, 'Morena: hegemonía y partidos satélite' ('Morena: Hegemony and Satellite Parties') (Proceso, 24 June 2024) <www.proceso.com.mx/opinion/2024/6/24/morena-hegemonia-partidos-satelite-331506.html> accessed 10 July 2024

Open AI, Dall-E, (Open AI), <https://openai.com/index/dall-e-3/> accessed 30 June 2024

--, Sora, Open AI <https://openai.com/index/sora/> accessed 24 June 2024

oscarwildones, 'Tercer Debate Ine [Third Debate Ine]' (TikTok, 20 May 2024) <www.tiktok.com/@oscarwildones/video/7370923051560602886> accessed 29 June 2024

Oversight Board, 'Altered Video of President Biden' (Oversight Board, 2023) <www.oversightboard.com/decision/FB-GW8BY1Y3/> accessed 29 June 2024

Pantoja A, 'Sheinbaum alerta sobre video falso hecho con inteligencia artificial: "es mi voz, pero es un fraude"' ('Sheinbaum warns about fake video made with artificial intelligence: "it's my voice, but it's a fraud"') (Proceso, 25 January 2024) <www.proceso.com.mx/nacional/politica/2024/1/25/sheinbaum-alerta-sobre-video-falso-hecho-con-inteligencia-artificial-es-mi-voz-pero-es-un-fraude-322795.html> accessed 10 June 2024

Pearson J, 'What the RIAA lawsuits against Udio and Suno mean for AI and copyright' (The Verge, 26 June 2024) <www.theverge.com/24186085/riaa-lawsuits-udio-suno-copyright-fair-use-music> accessed 30 June 2024

Porup JM, 'How Mexican Twitter Bots Shut Down Dissent' (Vice, 24 August 2015) <www.vice.com/en/article/z4maww/how-mexican-twitter-bots-shut-down-dissent> accessed 16 June 2024

Proceso, 'La estigmatización de los medios, signo de AMLO, según la CIDH y Artículo 19 [The stigmatization of the media, a hallmark of AMLO, according to the IACHR and Article 19]' (Proceso, 8 June 2020) <www.proceso.com.mx/reportajes/2020/6/8/la-estigmatizacion-de-los-medios-signo-de-amlo-segun-la-cidh-articulo-19-244156.html> accessed 28 June 2024

--, 'Exclusiva: Relator de CIDH pide detener "Quién es quién de las mentiras" por violencia a periodistas [Exclusive: IACHR Rapporteur calls to stop "Who's Who in Lies" due to violence against journalists]' (1 February 2022) <www.proceso.com.mx/nacional/2022/2/1/exclusiva-relator-de-cidh-pide-detener-quien-es-quien-de-las-mentiras-por-violencia-periodistas-280177.html> accessed 28 June 2024

R3D, 'Ejército de Bots: Las operaciones militares para monitorear las críticas en redes sociales y manipular la conversación digital [Army of Bots: Military operations to monitor social media critics and manipulate digital conversation]' (R3D, 27 February 2024) <https://r3d.mx/2024/02/27/ejercito-de-bots-las-operaciones-militares-para-monitorear-las-criticas-en-redes-sociales-y-manipular-la-conversacion-digital/> accessed 28 June 2024

Raziel Z, 'Empresa publicista de campaña de AMLO financió desinformación contra Ricardo Anaya en 2018 [AMLO's PR firm financed misinformation against Ricardo Anaya in 2018]' Animal Político (27 April 2022) <www.animalpolitico.com/2022/04/empresa-publicista-amlo-2018-financio-desinformacion-anaya/> accessed 28 June 2024

Rest of World Staff, 'A manipulated Mexican flag' Rest of World, 13 May 2024 <https://restofworld.org/2024/elections-ai-tracker/#/manipulated-mexico-flag> accessed 29 June 2024

Reuters Fact Check, "Verificación: Bachoco no publicó cartel en apoyo a Xóchitl Gálvez[Verification: Bachoco did not publish a poster in support of Xóchitl Gálvez]", October 5, 2023, <www.reuters.com/fact-check/espanol/NBESMKHW7RJE5EHFBWGOVMNCCI-2023-10-10/> accessed 29 June 2024

Reuters, "Myanmar: UN blames Facebook for spreading hatred of Rohingya", The Guardian, March 13, 2018 <www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya> accessed 29 June 2024

--, 'Fact Check en Español' (Reuters) <www.reuters.com/fact-check/espanol/> accessed 28 June 2024

Rogoff Z, 'Generative AI is Already Catalyzing Disinformation. How Long Until Chatbots Manipulate Us Directly?' (Tech Policy Press, 23 October 2023) <https://techpolicy.press/generative-ai-is-already-catalyzing-disinformation-how-long-until-chatbots-manipulate-us-directly> accessed 8 July 2024

Royal Society, 'Who We Are' (Royal Society, 2024) <https://royalsociety.org/about-us/who-we-are/> accessed 30 June 2024

Sánchez L and E Ponce de Léon, 'Red de sitios venezolanos a cargo de desinformación y propaganda sobre México, El Salvador, España y Perú [Network of Venezuelan sites in charge of disinformation and propaganda on Mexico, El Salvador, Spain and Peru]' Animal Político (27 July 2022) <https://mirror.animalpolitico.com/2022/07/red-de-sitios-venezolanos-desinformacion-propaganda-sobre-mexico-espana/> accessed 28 June 2024

Signa_Lab, 'PRI Edomex II: Estrategias de Influencia [PRI Edomex II: Influence Strategies]' Signalab (12 September 2020) <https://signalab.mx/2020/09/08/pri-edomex-ii-estrategias-influencia/> accessed 28 June 2024

Signalab, 'Signalab' (Signalab) <https://signalab.mx/> accessed 16 June 2024

Singh J, 'Google won't let you use its Gemini AI to answer questions about an upcoming election in your country' (TechCrunch, 12 March 2024) <https://techcrunch.com/2024/03/12/google-gemini-election-related-queries/?guccounter=1> accessed 8 July 2024

Situational Awareness, 'The Decade Ahead' (2024) <https://situational-awareness.ai/from-gpt-4-to-agi/> accessed 7 July 2024.

Sociedad Civil México, <https://x.com/SocCivilMx/status/1769850004548559314?s=20> accessed 29 June 2024

Soto D, 'El Partido Verde lo volvió a hacer: recibe promoción de influencers y modelos, pese a prohibición en intercampañas [The Green Party did it again: receives promotion from influencers and models despite inter-campaign prohibition]' (Animal Político, 29 May 2024) <https://animalpolitico.com/verificacion-de-hechos/fact-checking/influencers-y-modelos-publican-mensajes-a-favor-del-partido-verde> accessed 29 June 2024

spikolsmaniac, 'Hola Twitter... [Hello Twitter...]' (X, 1 June 2024) <https://x.com/spikolsmaniac/status/1796736418091237677> accessed 29 June 2024

Statista, 'Countries where a national election is/was held in 2024' (Statista, 2024) <www.statista.com/chart/31604/countries-where-a-national-election-is-was-held-in-2024/> accessed 5 May 2024

Sullum J, 'The Crusade Against 'Malinformation' Explicitly Targets Inconvenient Truths' (Reason, 22 March 2023) <https://reason.com/2023/03/22/the-crusade-against-malinformation-explicitly-targets-inconvenient-truths/> accessed 30 June 2024

Suno, 'About Suno' (Suno, 2024) <https://suno.com/about> accessed 30 June 2024

Susman-Peña T, 'Why Information Matters: A Foundation for Resilience' (May 2015, Internews) 12 <https://internews.org/wp-content/uploads/legacy/resources/150513-Internews_WhyInformationMatters.pdf> accessed 5 May 2024

Tencent AI Lab, 'V-Express: Conditional Dropout for Progressive Training of Portrait Video Generation' (GitHub, 2024) <https://github.com/tencent-ailab/V-Express> accessed 30 June 2024

TikTok, 'Integrity and Authenticity, Edited Media and AI-Generated Content (AIGC)' (TikTok Community Guidelines, 17 April 2024) <www.tiktok.com/community-guidelines/en/integrity-authenticity#3> accessed 29 June 2024

–, 'TikTok refuerza su compromiso con la integridad de la plataforma con iniciativas clave de cara a las próximas elecciones en México [TikTok reinforces its commitment to platform integrity with key initiatives for the upcoming elections in Mexico]' TikTok Newsroom (9 May 2024) <https://newsroom.tiktok.com/es-latam/elecciones-mexico-2024> accessed 28 June 2024

Trejo JO, 'Usando la tecnología (inteligencia artificial) se le preguntó a un robot por quién votar en las próximas elecciones del 2 de junio, y vean lo qué contestó en base a hechos [Using technology (artificial intelligence) a robot was asked who to vote for in the upcoming June 2 elections, and see what it answered based on facts]' (Facebook, 19 May 2024) <www.facebook.com/jose.o.trejo/posts/usando-la-tecnolog%C3%ADa-inteligencia-artificial-se-le-pregunt%C3%B3-a-un-robot-por-qui%C3%A9n/8116051695072480/> accessed 29 June 2024

Udio, 'Udio | AI Music Generator - Official Website' (Udio, 2024) <www.udio.com/> accessed 30 June 2024

vagagu, 'las traiciones del bienestar... [the betrayals of well-being]' (X, 3 October 2023) <https://x.com/vagagu/status/1709041132523463096> accessed 29 june 2024

Vallance C, 'Artificial intelligence could lead to extinction, experts warn' (BBC News, 30 May 2023) <www.bbc.com/news/uk-65746524> accessed 8 July 2024

Vasa-1 has not been released. Microsoft Research, 'VASA-1' (Microsoft, 2024) <www.microsoft.com/en-us/research/project/vasa-1/> accessed 30 June 2024.

Velázquez K and D Martínez, 'Lo cierto y lo falso del "Quién es quién en las mentiras" [The truth and falsehoods of AMLO's "Who's Who in Lies"]' Verificado (8 July 2021) <https://verificado.com.mx/imprecisiones-y-datos-falsos-en-una-tercera-parte-del-quien-es-quien-en-las-mentiras-de-la-mananera-de-amlo/> accessed 18 June 2024.

Viggle, 'Viggle AI' (Viggle, 2024) <www.viggle.ai/> accessed 30 June 2024

Villamil J, '#ConferenciaMañanera. Alertan sobre las Deepfakes que usan herramientas de inteligencia artificial para estafar a usuarios de plataformas digitales [Morning Conference. Warning about Deepfakes using artificial intelligence tools to scam digital platform users]' (X, 1 May 2024) <https://x.com/jenarovillamil/status/1785665951578407183> accessed 29 June 2024

Vincent J, 'The Lawsuit That Could Rewrite the Rules of AI Copyright' (The Verge, 8 November 2022) <www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data> accessed 6 june 2024

Wardle C, 'Understanding Information Disorder' (First Draft 2019) <https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76708> accessed 30 June 2024

Washington Post, 'López Obrador wins Mexican presidency, becoming first leftist to govern in decades' (1 July 2018) <www.washingtonpost.com/world/mexicans-head-to-polls-to-choose-a-new-president/2018/07/01/517e8670-7a2a-11e8-ac4e-421ef7165923_story.html> accessed 28 June 2024

Westby J, ''The Great Hack': Cambridge Analytica is just the tip of the iceberg' (Amnesty International, 24 July 2019) <www.amnesty.org/en/latest/news/2019/07/the-great-hack-facebook-cambridge-analytica/> accessed 12 June 2024

WIRED, 'El ejército mexicano tiene un centro secreto para monitorear opositores y operar bots [The Mexican Army has a secret center to monitor opponents and operate bots]' WIRED (2024) <https://es.wired.com/articulos/el-ejercito-mexicano-tiene-un-centro-secreto-para-monitorear-opositores-y-operar-bots> accessed 28 June 2024

Global Campus of Human Rights

Europe
South East Europe
Latin America-Caribbean
Asia-Pacific
Caucasus
Arab World
Africa

Global Campus of Human Rights
The Global Campus of Human Rights is a unique network of more than one hundred participating universities around the world, seeking to advance human rights and democracy through regional and global cooperation for education and research. This global network is promoted through eight Regional Programmes which are based in Venice for Europe, in Sarajevo/Bologna for South East Europe, in Yerevan for the Caucasus, in Pretoria for Africa, in Bangkok for Asia-Pacific, in Buenos Aires for Latin America and the Caribbean, in Beirut for the Arab World, and in Bishkek for Central Asia.

The Global Campus Awarded Theses
Every year each regional master's programmes select the best master thesis of the previous academic year that is published online as part of the GC publications. The selected GC master theses cover a range of different international human rights topics and challenges.

The present thesis - ***Voices amplified or silenced? Navigating the impact of generative AI on freedom of expression in Mexican elections*** written by **Vladimir Cortés Roshdestvensky** and supervised by Łukasz Szoszkiewicz, Adam Mickievicz University - was submitted in partial fulfilment of the requirements for the European Master's Programme in Human Rights and Democratisation (EMA), coordinated by Global Campus of Human Rights Headquarters.

GC Headquarters
Monastery of San Nicolò,
Riviera San Nicolò, 26
I-30126 Venice Lido (Italy)

www.gchumanrights.org

This document has been produced with the financial assistance of the European Union and as part of the Global Campus of Human Rights. The contents of this document are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the position of the European Union or of Global Campus of Human Rights.