Information Warfare

How Emerging Technologies Threaten Europe and Taiwan

Edited by

Niklas Swanström

& Yi-Chieh Chen

Special Paper | November 2025

Institute for Security & Development Policy

Information Warfare:

How Emerging Technologies Threaten Europe and Taiwan

Edited By

Niklas Swanström & Yi-Chieh Chen

Special Paper

November 2025

Institute for Security & Development Policy

"Information Warfare: How Emerging Technologies Threaten Europe and Taiwan" is a Special Paper published by the Institute for Security and Development Policy. The Institute is based in Stockholm, Sweden, and cooperates closely with research centers worldwide. The Institute serves a large and diverse community of analysts, scholars, policy-watchers, business leaders, and journalists. It is at the forefront of research on issues of conflict, security, and development. Through its applied research, publications, research cooperation, public lectures, and seminars, it functions as a focal point for academic, policy, and public discussion.

ISBN: 978-91-978-91-88551-75-7

Distributed in Europe by:

Institute for Security and Development Policy
Västra Finnbodavägen 2, 131 30 Stockholm-Nacka, Sweden
Tel. +46-841056953; Fax. +46-86403370
Email: info@isdp.eu

Editorial correspondence should be directed to the address provided above (preferably by email).

Cover Photo: Deemerwha studio / Shutterstock

Contents

List of Abbreviations

3PFC	Third-Party Fact-Checking Program
AI	Artificial Intelligence
AIGC	AI-generated Content
CCP	Chinese Communist Party
CNN	Convolutional Neural Network
DL	Deep Learning
DoS	Denial of Service
DPP	Democratic Progressive Party
DSA	Digital Services Act
EC	European Commission
EDMO	European Digital Media Observatory
EFCSN	European Fact-Checking Standards Network
EU	European Union
FDA	Food and Drug Administration
FTC	Federal Trade Commission
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GEC	Global Engagement Center
GRU	Gated Recurrent Unit
IFCN	International Fact-Checking Network
INDSR	Institute for National Defense and Security Research
IRA	Internet Research Agency
LLM	Large Language Model
MIL	Media and Information Literacy

NGO	Non-Governmental Organization
NIST	National Institute for Standards and Technology
NLP	Natural Language Processing
NSA	National Security Agency
OSINT	Open Source Intelligence
PQC	Post-Quantum Cryptography
PRC	People's Republic of China
QKD	Quantum Key Distribution
Qubits	Quantum Bits
RAG	Retrieval-Augmented Generation
RNN	Recurrent Neural Network
RSF	Reporters Without Borders
RvNN	Recursive Neural Network
SWOT	Strengths, Weaknesses, Opportunities, and Threats
TFC	Taiwan FactCheck Center
TSMC	Taiwan Semiconductor Manufacturing Company
UN	United Nations
U.S.	United States
VLOP	Very Large Online Platform
VPN	Virtual Private Network
XAI	eXplainable AI

List of Contributors

Niklas Swanström is the Director of the Institute for Security and Development Policy, and one of its co-founders. He is a Fellow at the Foreign Policy Institute of the Paul H. Nitze School of Advanced International Studies (SAIS) and a Senior Associate Research Fellow at the Italian Institute for International Political Studies (ISPI). His main areas of expertise are conflict prevention, conflict management and regional cooperation; Chinese foreign policy and security in Northeast Asia; the Belt and Road Initiative, traditional and non-traditional security threats and its effect on regional and national security as well as negotiations. His focus is mainly on Northeast Asia, Central Asia and Southeast Asia.

Yi-Chieh Chen (陳奕傑) is a Project Manager and Junior Research Fellow at the Institute for Security and Development Policy's Stockholm Taiwan Center (STC). She is also part of the Stockholm Center for Research and Innovation Security (SCRIS). She holds a Bachelor's degree in Arabic Language and Culture from National Chengchi University in Taiwan and a Master's degree in Global Studies from Gothenburg University in Sweden. Yi-Chieh Chen has a broad interest in East Asian affairs, soft power, and technology. Her research focus includes Taiwan-Europe relations, cross-strait relations, sports diplomacy, and the semiconductor industry.

Ming-Hung Wang (王銘宏) is an Associate Professor in the Department of Computer Science and Information Engineering at National Chung Cheng University, Taiwan. He received his Ph.D. degree in Electrical Engineering from National Taiwan University in 2017. He was the recipient of the Young Faculty Award from National Chung Cheng University in 2024. Dr. Wang serves as the leader of the Digital Society and Security Lab (DIGI-SSL), focusing on research topics in computational social science and network security. His research interests span information manipulation, spam

detection, as well as various social media security issues.

Hsiu-Ling Chu (朱琇伶) received her Master's degree in Information Engineering at Feng Chia University in 2010. She is currently pursuing her Ph.D. degree in the Department of Information Engineering and Computer Science, Feng Chia University, Taichung City, Taiwan. Her research interests include data science and information security.

Wei-Bin Lee (李維斌) serves as the CEO of HonHai Research Institute, where he also directs the Information Security Research Center, and as CISO of HonHai Technology Group (Foxconn). He earned his Ph.D. from National Chung Cheng University in 1997. Lee is a professor in the Department of IECS at Feng Chia University and was a visiting scholar at both Carnegie Mellon University (USA) and the University of British Columbia (Canada). He previously held key positions at Taipei City Government, Taipei Fubon Bank, Fubon Financial Holdings, and Feng Chia University. His expertise covers network security, cryptography, digital rights management, and privacy and security governance.

Chih-hsing Ho (何之行), LLM (Columbia), JSM (Stanford), Ph.D. in Law (London School of Economics), is an Associate Professor and Associate Research Fellow at the Institute of European and American Studies (IEAS), Academia Sinica, Taiwan. Her research focuses on the intersection of law, data science, and biomedicine, with particular emphasis on emerging technologies such as big data and artificial intelligence. Ho is the founder and coordinator of the AI Governance Laboratory at IEAS and serves on Academia Sinica's Advisory Committee on the Risks of Generative AI. She is also a member of the Advisory Committee on Digital Economy Development under Taiwan's Ministry of Digital Affairs. Her work has appeared in leading international journals, including Nature Genetics, Journal of Law, Information and Science, and Computer Law & Security Review.

Vera Schmitt is a computer scientist and the founder and head of the XplaiNLP Research Group at Technical University Berlin. Her research

focuses on Natural Language Processing (NLP) for high-stakes decision-making, particularly in areas such as disinformation detection and medical decision-support. Together with her team, she develops core NLP approaches that promote the robust and transparent deployment of AI systems, with a strong emphasis on explainable AI and human-computer interaction. Schmitt holds a Ph.D. in Computer Science from TU Berlin and is an active member of initiatives including AI4Media, AI-Grid, and the Center for European Research in Trusted AI (CERTAIN).

Maya Sobchuk is a researcher within the Research Center for Advanced Science and Technology based at the University of Tokyo and a Non-Resident Fellow in the Artificial Intelligence and Global Governance Programme at the Global Governance Institute in Brussels.

Tzu-Wei Hung (洪子偉) is a Professor and Research Fellow at the Institute of European and American Studies, Academia Sinica, Taiwan. His fields include the philosophy of cognitive science, the philosophy of language, and the philosophy of artificial intelligence. He read philosophy at King's College London and taught at Taipei Medical University. Hung was also a fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford University and received a visiting fellowship from the Harvard-Yenching Institute. Since 2024, he has served as the president of the Taiwan Philosophical Association.

Chen-Ling Hung (洪貞玲) is the Associate Dean of the College of Social Sciences and a professor at the Graduate Institute of Journalism, National Taiwan University. Her research fields include disinformation, citizen journalism, communication law and policy, and global communications. She has worked with media reform organizations for a long time. She is the chairperson of the Taiwan Media Watch Foundation. Between September 2022 and August 2024, she served as the chairperson of the Taiwan Communication Association. Between 2016 and 2020, she served as the Commissioner of the National Communications Commission, the media regulator in Taiwan.

Giovanni Zagni, Ph.D., is a journalist based in Milan, Italy, and the director of the fact-checking projects Pagella Politica and Facta. He is a member of the Executive board of the European Digital Media Observatory (EDMO) and served as the chair of its Task Force on the European Parliamentary Elections 2024. He was a member of the Committee of Experts on the Integrity of Online Information established in 2022 by the Council of Europe and took part in the Monitoring Unit on Disinformation around Covid-19 established by the Italian government in 2020. He is the co-author of three books on fact-checking and disinformation.

Andrea G. Rodríguez is the Chair of the Governing Board at ImpaQT UA, a quantum tech consortium based in Delft (Netherlands). She is also an Associate Researcher at the Centre for European Policy Studies (CEPS), working on the intersection of emerging technologies, European policy, and transatlantic relations. Also, Andrea teaches Cybersecurity at the Catholic University of Lille (France). Before her roles at CEPS and ImpaQT, she was Lead Digital Policy Analyst at the European Policy Centre. She has been part of several advisory groups at different organizations, such as NATO or the European Cybersecurity Forum.

Irène Dubois is the EU Digital Policy Assistant at The Policy Pulse, specializing in digital interdependencies, semiconductors economic security with a regional focus on Europe and Asia. Fluent in French, English, and Mandarin. Her French-Taiwanese background enables her to conduct in-depth research involving Chinese-language sources. Irène holds a Master's degree in Digital Politics and Governance and a Bachelor's degree in International Relations from Lille Catholic University's European School of Political and Social Sciences (ESPOL).

Introduction: Development of New Technology and the Impact of False Information in Europe and Taiwan

Niklas Swanström and Yi-Chieh Chen

The intersection of emerging technologies and disinformation has created unprecedented challenges for democratic societies, particularly in geopolitically sensitive regions like Taiwan and Europe. As artificial intelligence (AI), deepfake technology, computer vision algorithms and social media algorithms become increasingly sophisticated, the landscape of information warfare has transformed dramatically in recent years. This transformation represents not merely an evolution of existing propaganda techniques, but there has been a fundamental shift in how false information is created, disseminated, and consumed internationally.[1] This publication aims to disseminate the ongoing challenges and the reactions from Taiwan and Europe, but more importantly look at the emerging security challenges and how the actions in the future should look.

Why Taiwan and Europe?

Taiwan and Europe represent critical case studies in the global battle against technologically enhanced disinformation for several compelling reasons. Both regions combine advanced digital infrastructure with robust democratic institutions, making them both attractive targets and innovative defenders against information warfare. They share some of the ongoing and future challenges, but at the same time, there is a relevant and interesting distinction between the two actors that could highlight both different strategies and purposes. Finally, both actors are under attack from external threats at an unprecedented level, with increased pressure from China on Taiwan and a combined threat from China and Russia in regard to Europe, with Russia conducting a barbaric invasion of an independent European nation, Ukraine.

Taiwan in the Line of Fire

Taiwan stands at the forefront of this battle against technologically enhanced disinformation due to its unique position in global geopolitics. Situated at the crossroads of great power competition, Taiwan faces extraordinary challenges as a democracy under constant pressure. The island's historical entanglement with China, strategic importance in global semiconductor production, and international trade make it a prime target for sophisticated influence operations.[2] This vulnerability is amplified by Taiwan's exceptional internet penetration rate, around 87 percent as of 2023, and its advanced technological ecosystem that creates both opportunities and vulnerabilities.[3] The population's high digital engagement enables rapid information spread but also increases exposure to coordinated disinformation campaigns.[4]

The cultural and linguistic complexity of Taiwan's information environment creates additional challenges for content moderation and fact-checking systems. The interplay between simplified Chinese, traditional Chinese, Taiwanese, and English is frequently exploited by adversaries to create targeted disinformation campaigns that resonate with different demographic groups. The similarity between simplified Chinese and traditional Chinese increases the possibility of disseminating information. This multilingual environment has become increasingly vulnerable as AI systems have grown more sophisticated. This is a challenge that is shared by Europe, which has an even more complex linguistic environment.

The sophistication of operations targeting Taiwan has grown exponentially with technological innovations. Large Language Models (LLMs) with transformer architecture now generate culturally nuanced content that seamlessly integrates local idioms and cultural references, making detection increasingly difficult. Recent studies indicate a 300 percent increase in AI-generated content targeting Taiwanese social media between 2022 and 2024.[5] The integration of text, images, and video through advanced AI systems creates compelling cross-platform narratives that can rapidly spread through Taiwan's dense social networks. These systems can now generate content that appears authentic even to experienced observers,

and civil society organizations in Taiwan have specialized in detecting fake pictures.

Voice cloning technology has emerged as a particularly concerning threat in Taiwan's political context. Advanced neural voice synthesis can replicate political leaders' voices with almost 100 percent accuracy, enabling the creation of fake audio clips that spread rapidly through popular messaging apps and social media like LINE, Facebook, TikTok, Instagram, and Xiaohongshu.[6] This technology, combined with sophisticated video manipulation capabilities, presents an unprecedented challenge to information integrity during critical periods such as elections.

The ever-changing social media algorithms increase the difficulty of identifying disinformation and misinformation. The algorithms can depoliticize the content that decreases users' alertness to false information. Algorithms also create an echo chamber where the users believe the surrounding world is similar to the information they are subjected to. This situation creates an environment where the users neglect the possibility of cultural penetration.[7] The similarities in language and part of culture between Taiwan and China amplify the risks that algorithms can bring.

Europe's Distinct Challenges

The European context presents a different but equally critical set of challenges in the fight against technologically enhanced disinformation. The EU's 24 official languages and myriads of minority languages create unique vulnerabilities that emerging technologies exploit in unprecedented ways. LLMs trained on massive multilingual datasets now enable rapid production of culturally adapted disinformation across all EU languages. These systems can automatically adapt content to reflect regional dialects and cultural nuances, while maintaining coordinated narrative structures across borders.

The EU's layered governance structure creates challenges for implementing unified responses to technological threats. However, this complexity has also driven innovative policy solutions. The Digital Services Act (DSA) represents

the world's most comprehensive attempt to regulate digital platforms, while the AI Act sets global standards for AI governance. These two acts are EU regulations adopted in 2022 and 2024, respectively. DSA aims to "prevent illegal and harmful activities online and the spread of disinformation"[8] which complements the AI Act's purpose of regulating the use of AI systems.[9] These cross-border cooperation mechanisms enable rapid response to emerging threats, though coordination remains an ongoing challenge.

The Russian invasion of Ukraine has highlighted how modern technologies can amplify disinformation's impact on European security. Synthetic media generation has evolved to include real-time deepfake capabilities, enabling fake live streams and manipulated video conferences. Advanced computer vision algorithms can now alter satellite imagery and drone footage, creating false evidence of military activities or civilian casualties. The openness and transparency of the European societies have been precisely the vulnerabilities exploited by Russian information warfare.

The Democratization of Information Spread and Falsification

Perhaps most concerning is how new technologies have democratized the creation and distribution of disinformation. The implementation of edge computing and 5G networks has fundamentally altered the dynamics of disinformation spread. These technological advancements have accelerated the dissemination of false information by enabling real-time content generation and distribution at the network edge. The reduced latency and increased processing capabilities allow disinformation campaigns to rapidly adapt to current events and target specific geographic areas with unprecedented precision. This technological infrastructure has created new vulnerabilities in both European and Taiwanese information ecosystems, as threat actors can now deploy sophisticated campaigns with minimal delay between content creation and distribution.[10]

Automated bot networks have become increasingly sophisticated, leveraging reinforcement learning to optimize their influence operations. These systems can

analyze engagement patterns and adapt their behavior in real-time, maximizing the spread of false information across social networks. The integration of emotion AI and sentiment analysis tools has enabled these systems to calibrate disinformation for maximum psychological impact, targeting specific emotional vulnerabilities within different demographic groups.[11]

The emergence of sophisticated GPT-style models has similarly revolutionized text-based disinformation. GPT-style refers to Generative Pretrained Transformer, which is a type of LLM with transformer architecture. These systems can craft persuasive content in multiple languages with minimal human intervention, adapting tone and style to match target audiences. The ability to generate culturally nuanced content at scale has made it increasingly difficult for readers to distinguish between authentic and artificial narratives. Neural rendering engines have further complicated the situation by enabling the creation of photorealistic 3D environments for fake news videos, providing convincing backdrops for fabricated events.[12]

Commercial AI tools now offer sophisticated capabilities that were once limited to state actors, fundamentally altering the landscape of information warfare. Stable Diffusion and DALL-E-like systems have transformed image manipulation, enabling users to generate or modify visual content that supports false narratives with remarkable authenticity. These systems have dramatically lowered the technical barriers to creating convincing fake imagery, making visual disinformation increasingly prevalent across social media platforms.[13]

Quantum computing looms as a significant threat multiplier in data security and cyber defense. While still in development, quantum computers could potentially break current encryption standards, compromising digital signatures and authentication systems that help verify content authenticity. It also has the potential to accelerate the dissemination of false information and amplify deepfake technologies. The EU's investment in post-quantum cryptography reflects this growing concern about future technological vulnerabilities.

Amid rapid technological developments, the creation and dissemination of false information, facilitated by the democratization of disinformation tools, pose formidable challenges that both Taiwan and the EU are struggling to overcome.

Responses and Countermeasures

The response to these technological challenges has necessarily been multifaceted, with Taiwan and Europe developing distinct but complementary approaches. Taiwan has emerged as a global leader in developing innovative approaches to combat technology-enabled disinformation. The island's comprehensive digital literacy programs have reached over 85 percent of the adult population, creating a more resilient citizenry capable of identifying and resisting false information.[14] To share its experiences and further develop its capacity to combat the challenges of disinformation and foreign information manipulation, Taiwan has co-hosted several international media literacy workshops under the Global Cooperation and Training Framework (GCTF).[15] Taiwan's civil society initiatives have pioneered new models of citizen engagement in fact-checking and verification, creating distributed networks of digital actors that collaboratively monitor information and enhance the public trust in information ecosystem.[16] The resources are listed on the website of the Ministry of Justice Investigation Bureau and include the Taiwan FactCheck Center initiated by non-governmental organizations and LINE, the dominant communication app in Taiwan.[17]

The European Union (EU) has taken a more regulatory approach, introducing comprehensive legislation like the DSA and the AI Act to increase platform accountability and transparency. This regulatory framework establishes clear obligations for digital platforms while creating mechanisms for coordinated responses to disinformation campaigns. The EU's approach emphasizes the importance of systematic oversight and corporate responsibility in maintaining information integrity.[18] In 2022, the European Fact-Checking Standards Network (EFCSN) was launched by European fact-checking organizations. It aims to "promote the highest standards of fact-checking and promote media literacy for the public benefit."[19] With civil society-

initiated EFCSN, the EU's regulatory-based approach is complemented by civil society. Such a combination can form a sophisticated network with both top-down and bottom-up approaches to regulate and restrain the impact of false information. Technical countermeasures continue to evolve in parallel with emerging threats. Advanced forensic tools employing machine learning algorithms have shown increasing success in detecting synthetic media, while distributed ledger technologies create immutable records of content provenance. Digital watermarking and content authentication protocols are being standardized across platforms, creating more robust verification systems for authentic content. However, these technical solutions face ongoing challenges as adversarial AI systems become more sophisticated in evading detection.[20]

Future Research Imperatives

The rapid evolution of technology continues to outpace defensive measures, creating urgent needs for future research. Cross-platform disinformation dynamics require particular attention, as content increasingly flows between different social media environments, morphing and adapting to platform-specific constraints. The cultural similarity and the divergent political stance on cross-strait relations between Taiwan and China complicate Taiwan's challenges of combatting disinformation under the technological boom. Taiwan is not only simply facing the situation of false information spreading within society but also working on securing its political entity as a democratic self-governed island. The unique characteristics of Taiwan's social media ecosystem, where platforms like Instagram, Facebook, and TikTok, coexist with Western, traditional Chinese and simplified Chinese content intertwined with diverse political views and cultural similarities about Taiwan and China, provide valuable insights into these complex dynamics.

The development of quantum-resistant authentication systems has become increasingly critical as quantum computing capabilities advance. Research must focus on practical implementations of post-quantum cryptography for content verification, particularly in high-stakes contexts like election security and public health communications. The EU's investments in this area reflect

growing recognition of the need for future-proof security solutions.

Understanding the cognitive impact of AI-generated disinformation compared to traditional false content represents another crucial research direction. Studies must examine how exposure to deepfakes and synthetic media affects trust in authentic content and institutions over time. The European context, with its diverse populations and varying levels of digital literacy, provides an ideal environment for such research.

International Cooperation and Strategic Implications

The technological arms race in disinformation demands unprecedented levels of international cooperation. Neither Taiwan nor Europe can effectively address these challenges in isolation, nor in a simple bilateral partnership. The complex interplay of technological capabilities, geopolitical tensions, and information ecosystems requires a coordinated global response that transcends traditional diplomatic and technological boundaries.

Taiwan's unique position at the intersection of technological innovation and geopolitical tension makes it a critical node in understanding global disinformation dynamics. Its experiences with persistent information warfare from state and non-state actors provide invaluable insights into sophisticated technological manipulation strategies. The island's resilience has become a global model for developing adaptive digital defense mechanisms that combine technological solutions with societal resistance.

European approaches complement Taiwan's strategies through their emphasis on regulatory frameworks and multinational coordination. The EU's comprehensive legislative efforts, particularly the DSA and emerging AI regulations, represent the most sophisticated attempt to create systemic governance for digital information environments. These regulatory approaches provide a potential template for other democratic societies seeking to balance technological innovation with information integrity.

Addressing the complex challenges of technological disinformation

requires a multidimensional approach that integrates technological, educational, and policy interventions. Future technological development must prioritize ethical considerations and vulnerability mitigation. This involves creating AI systems with inherent verification mechanisms, developing more sophisticated synthetic media detection tools, and establishing international standards for content authentication. The goal is not to restrict technological innovation but to create more robust and trustworthy digital environments.

Digital literacy must become a core component of educational curricula at all educational levels. This goes beyond teaching technical skills to developing critical thinking capabilities that enable individuals to navigate complex information environments. Both Taiwan and Europe have demonstrated the potential of comprehensive digital education programs in building societal resilience against disinformation, but more needs to be done to stay ahead of the development.

The rapid advancement of AI and synthetic media technologies raises profound ethical questions. Researchers, policymakers, and technology developers must engage in ongoing dialogue about the potential societal impacts of these technologies. Transparency, accountability, and human-centric design must be prioritized to prevent the potential weaponization of advanced technological capabilities.

Navigating Technological Uncertainty

The experiences of Taiwan and Europe offer crucial insights into the complex relationship between technological innovation and information integrity. These regions demonstrate both the vulnerabilities and the potential resilience of democratic societies in the face of sophisticated technological challenges. Their strategies reveal that effective defense against disinformation requires a holistic approach combining technological innovation, robust governance, societal education, and international cooperation.

As AI, quantum computing, and synthetic media technologies continue

to advance, the battle for information integrity will become increasingly sophisticated. The ability to adapt, innovate, and maintain human agency in increasingly complex technological environments will be critical to preserving democratic discourse and societal trust.

The journey is not about creating impenetrable technological barriers but about developing more nuanced, adaptive, and resilient information ecosystems that can withstand sophisticated technological manipulations while preserving the fundamental values of openness, creativity, and democratic engagement. This publication aims at providing another step in the long journey towards a more secure and innovative global environment.

Endnotes

1 E. Broda and J. Strömbäck, "Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review," *Annals of the International Communication Association 48*, no. 2 (2024): 139–66, doi:10.1080/23808985.2024.2323736.

2 N. Swanström and F. B. Månsson, "The Convergence of Disinformation: Examining Russia and China's Partnership in the Digital Age," Institute for Security and Development Policy, 2024, https://www.isdp.eu/publication/the-convergence-of-disinformation-examining-russia-and-chinas-partnership-in-the-digital-age/.

3 Ministry of Digital Affairs (Taiwan) and United Market Research, "112 年數位發展調查報告及摘要," April 2023, https://www-api.moda.gov.tw/File/Get/moda/zh-tw/yg2U5UZOHhvWvtm.

4 Doublethink Lab, "Artificial Multiverse: Foregin Information Manipulation and Interference in Taiwan's 2024 National Elections," August 13, 2024, https://medium.com/doublethinklab/artificial-multiverse-foreign-information-manipulation-and-interference-in-taiwans-2024-national-f3e22ac95fe7.

5 Ibid.

6 F. Efthymiou, C. Hildebrand, E. de Bellis, and W. H. Hampton, "The Power of AI-Generated Voices: How Digital Vocal Tract Length Shapes Product Congruency and Ad Performance," *Journal of Interactive Marketing 59*, no. 2 (2024): 117-134, https://doi.org/10.1177/10949968231194905.

7 潘姿羽 and 吳家豪, "小紅書TikTok在台灣1 / 中國小紅書披流行外衣 精準演算法讓台灣年輕人不設防," *Central News Agency*, August 12, 2024, https://www.cna.com.tw/news/aipl/202408120047.aspx.

8 European Commission, "The Digital Service Act," https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en (accessed June 9, 2025).

9 European Union, "Regulation - EU - 2024/1689 - EN - EUR-Lex," https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689 (accessed June 9, 2025).

10 W. S. Admass, Y. Y. Munay, and A. A. Diro, "Cyber Security: State of the art, challenges and future directions," *Cyber Security and Applications 2*, 2024, https://doi.org/10.1016/j.csa.2023.100031.

11 T. Chutia and N. Baruah, "A review on emotion detection by using deep learning techniques," *Artif Intell Rev 57*, no. 203 (2024), https://doi.org/10.1007/s10462-024-10831-1.

12 S. K. Sharma, A. AlEnizi, M. Kumar, O. Alfarraj, and M. Alowaidi, "Detection of real-time deep fakes and face forgery in video conferencing employing generative adversarial networks," *Heliyon 10*, no. 17 (2024), https://doi.org/10.1016/j.heliyon.2024.e37163.

13 M. Shahbazi and D. Bunker, "Social media trust: Fighting misinformation in the time of crisis," *International Journal of Information Management 77*, (2024), https://doi.org/10.1016/j.ijinfomgt.2024.102780.

14 Pek Hua Lim, "Social Media Statisitics for Taiwan," Meltwater, September 25, 2024, https://www.meltwater.com/en/blog/social-media-statistics-taiwan.

15 Global Cooperation& Training Framework, "Media Literacy," https://www.gctf.tw/en/issues6_0.htm (accessed June 9, 2025).

16 C. Hung, W. Fu, C. Liu, and H. Tsai, "AI Disinformation Attacks and Taiwan's Responses during the 2024 Presidential Election," Thomson Foundation and Taiwan Communication Association, 2024, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf.

17 Ministry of Justice Investigation Bureau (Taiwan), "假訊息查證參考資訊," https://www.mjib.gov.tw/EditPage/?PageID=adf9b60f-98af-4b65-b996-4f74145a4cd0 (accessed June 9, 2025).

18 European Commission, "Corporate sustainability and responsibility," 2024, https://single-market-economy.ec.europa.eu/industry/sustainability/corporate-sustainability-and-responsibility_en.

19 European Fact-Checking Standards Network (EFCSN), "Statutes," https://efcsn.com/statutes (accessed June 9, 2025).

20 I. Jada and T. Mayayise, "The impact of artificial intelligence on organizational cyber security: An outcome of systematic literature review," *Data and Information Management 8*, no. 2 (2024), https://doi.org/10.1016/j.dim.2023.100063.

1. A Computational Lens on Information Warfare in Taiwan

Ming-Hung Wang, Hsiu-Ling Chu and Wei-Bin Lee

Introduction

In an era where a single smartphone notification can ignite outrage, sway opinions, or even shape electoral outcomes, the concept of warfare has transcended traditional battlefields. No longer confined to missiles, armies, or territorial disputes, conflict has evolved into a new, intangible realm: **the war of information**. This form of warfare is subtle, psychological, and deeply manipulative, and its impact is particularly pronounced and dangerous in Taiwan, a region on the front lines of such threats.

Taiwan occupies a critical position amidst global geopolitical tension. With its thriving democratic society, advanced technology sector, and pivotal role in the Indo-Pacific, Taiwan stands as both a symbol of liberal democratic values and a contested arena in the struggle against authoritarian influence. While conventional military threats are well known, **information warfare** poses a quieter but equally existential threat to Taiwan's political stability, social unity, and national sovereignty.

Information warfare can be defined as the strategic use of disinformation, propaganda, and psychological operations to influence public perception, disrupt civil trust, and destabilize societies. Unlike traditional warfare, it does not require guns or soldiers—only narratives, memes, bots, and platforms. The weapon is disinformation, the battlefield is public discourse, and the casualty is truth.

Emerging technologies, especially artificial intelligence (AI), machine learning, and big data analytics, have rapidly changed the scope and scale of these operations. What used to be manual and slow—writing propaganda,

forging documents, spreading rumors by word of mouth—has now become **automated, scalable, and lightning-fast**, capable of reaching millions in a matter of seconds.

These tools are often deployed by organized, well-funded groups, who target Taiwan's elections, referendum campaigns, social debates, and international standing. These campaigns exploit vulnerabilities in human psychology, social media design, and cultural narratives to create confusion, provoke anger, and fragment communities.

This chapter delves into how these technological innovations are being used to wage information warfare against Taiwan and, importantly, what we can do to defend itself. Through the lens of computational social science, we explore the methods researchers and engineers are using to detect, understand, and counteract this evolving threat.

The Nature of the Threat

Taiwan's unique geopolitical situation, positioned precariously between global superpowers, renders it a focal point for sophisticated information attacks, often serving as both a testing ground for hybrid warfare strategies and a symbolic battleground in the contest for regional hegemony. Foreign actors exploit Taiwan's democratic openness, leveraging sharp power—as conceptualized by Joseph Nye[1]—to manipulate and distort information environments, aiming to destabilize its political system and influence its domestic and international standing. Disinformation campaigns in Taiwan can be categorized into several distinct strategies, each with profound implications for its political stability and societal cohesion:

1. **Election Manipulation:** During elections and referendums, disinformation campaigns aim to sway public opinion, discredit candidates, and suppress voter turnout, undermining the legitimacy of electoral outcomes. Such tactics align with theories of electoral interference, where external actors disrupt the democratic process to favor specific outcomes, weakening Taiwan's democratic institutions

and eroding public faith in the electoral system.

2. **Social Division:** False narratives are deliberately crafted to exacerbate societal cleavages, deepening divides along ethnic, generational, and ideological lines. This strategy mirrors the classical tactic of "divide and rule," fragmenting societies to reduce their collective resilience.[2] In Taiwan, such narratives amplify tensions between ethnic groups (e.g., indigenous communities versus Han Chinese descendants) or generational cohorts (e.g., older versus younger generations), fostering polarization and hindering social unity. This dynamic is further intensified by the echo chamber effect, where groups exhibit designed and guided behavioral alignment, with strengthened in-group cohesion amplifying biases against out-groups, and as group sizes grow, the proliferation of incompatible, chaotic perspectives may ultimately lead to societal collapse. Social identity and conflict theories highlight how constructed narratives intensify in-group/out-group dynamics, while emphasizing that dialogue between groups could serve as a potential strategy for mutual understanding and reconciliation.[3]

3. **International Isolation:** Propaganda campaigns target Taiwan's sovereignty and legitimacy within international discourse, aiming to isolate it diplomatically. These efforts involve spreading narratives that question Taiwan's status as a sovereign entity, employing a strategy of de-legitimation to challenge its global recognition. By eroding Taiwan's international support, such disinformation reinforces the "One China" narrative, pressuring global actors to distance themselves from Taiwan and weakening its position in international organizations like the United Nations or the World Health Organization.

4. **Public Panic or Distrust:** Disinformation surrounding public health crises, military activities, or foreign diplomacy is designed to incite fear and erode trust in governmental institutions. For instance, fake news about fabricated military threats or health emergencies can trigger widespread panic, a process akin to securitization, where issues are framed as

existential threats to justify public fear.[4] This undermines the state's legitimacy by portraying it as incapable of managing crises, fracturing the social contract—a foundational element ensuring governance through mutual trust and accountability—between the government and its citizens.

The potency of these threats is amplified by their evolved mode of delivery in the digital age. Unlike the fundamental methods of the past—such as low-quality broadcasts—modern disinformation campaigns are **visually sophisticated**, **emotionally resonant**, and **algorithmically optimized** for virality. These campaigns harness advanced multimedia techniques to create compelling content that exploits psychological biases like confirmation bias or fear-based decision-making, while platform algorithms amplify their reach and impact. This reflects the broader paradigm of information warfare, where the control and manipulation of information flows become a critical battleground in modern geopolitics. In Taiwan's context, such tactics not only threaten domestic stability but also challenge its democratic resilience, making the development of robust countermeasures a pressing priority for safeguarding its political integrity and societal harmony.[5]

Emerging Technologies: Tools of Modern Propaganda

The dissemination of false information has long been a tactic in conflicts, dating back centuries, but the advent of emerging technologies has dramatically amplified the scale, sophistication, and impact of contemporary disinformation campaigns. These technologies not only enhance the believability of deceptive content but also pose significant challenges for detection and mitigation, as they exploit the digital ecosystem's interconnectedness and speed. In Taiwan, where political tensions and democratic openness make it a prime target for information warfare, these tools are particularly potent, enabling malicious actors to manipulate public perception with unprecedented precision and reach. The following outlines the core technological innovations driving modern propaganda and their specific applications in disinformation campaigns:

1. AI-Generated Content

AI has revolutionized the creation of deceptive media, producing highly believable content that blurs the line between reality and fabrication. AI-driven tools can generate photorealistic images, fake videos (commonly known as deepfakes), coherent text, and even synthetic voices that mimic real individuals with striking accuracy. In the context of disinformation, these capabilities are weaponized to:

- Fabricate news reports or interviews that appear authentic, misleading audiences into believing false narratives.
- Impersonate public figures, such as politicians or celebrities, to spread misleading statements or endorsements that influence public opinion.
- Create emotionally resonant but entirely fictional stories, often designed to provoke fear, anger, or division, particularly during sensitive political moments like elections or referendums in Taiwan.

2. Automated Bot Networks

Automated bot networks, consisting of programmed accounts that mimic human behavior, have become a cornerstone of digital propaganda. These bots interact with social media content to artificially inflate visibility, create the illusion of widespread consensus, or drown out opposing viewpoints. In Taiwan, bot armies are deployed with alarming efficiency to:

- Amplify disinformation by rapidly sharing false content across platforms, ensuring it reaches a wide audience before it can be debunked.
- Attack celebrities and politicians through coordinated harassment campaigns, undermining their credibility and sowing distrust among their supporters.
- Manipulate trending topics by flooding platforms with specific hashtags or keywords, shaping public discourse and steering attention toward fabricated narratives.

3. Sentiment Analysis and Microtargeting

Advancements in sentiment analysis and microtargeting enable disinformation operators to craft highly personalized messages that exploit individual biases, fears, and emotional triggers. By analyzing users' online behavior—such as

their browsing history, social media interactions, and emotional responses—actors can tailor content to resonate with specific audiences, making it more persuasive and harder to detect. This technology is often used to target vulnerable demographics, such as undecided voters or groups with strong ideological leanings, amplifying divisive narratives. For instance, tailored messages might exploit fears of economic instability or national security threats, subtly influencing voter behavior during critical political events like elections or referendums.

4. Multimodal Disinformation

Modern disinformation campaigns have evolved beyond simple text-based falsehoods, increasingly relying on multimodal formats that integrate images, videos, memes, and audio to create more engaging and convincing content. This blending of media types makes disinformation not only more compelling but also more resistant to traditional fact-checking methods, as visual and auditory elements are harder to verify than text alone. Visual content, in particular, holds immense power in emotionally charged political environments, where it can evoke strong emotional responses that override rational scrutiny. In Taiwan, the use of memes and propaganda videos has surged during key political moments, such as election campaigns or referendums, where emotionally charged visuals are deployed to sway public sentiment, deepen societal divides, or undermine trust in democratic institutions. For example, a single viral meme or video can rapidly spread narratives that question governmental legitimacy or incite fear, leveraging the emotional immediacy of visual media to amplify its impact.[6]

Decoding Political Messaging: Deep Learning for Multimodal Fake News Detection

The primary challenges in detecting fake news stem from the variability of its formats, the demand for large annotated datasets, and the difficulty of identifying subtle manipulations. Traditional methods struggle to address the complexities posed by multimodal content, which encompasses text, images, and videos. However, deep learning (DL) models, with their ability to integrate multimodal data, demonstrate significant potential in combating

fake news. For instance, by employing convolutional neural networks (CNNs) and transformers, these models can analyze and classify multimodal data, capturing intricate patterns across different data types and thereby substantially improving detection accuracy.[7]

CNNs excel at analyzing visual data, employing convolutional filters to effectively extract spatial features for classifying images or detecting objects. By integrating models within the neural network ecosystem—such as recurrent neural networks (RNNs), gated recurrent units (GRUs), and recursive neural networks (RvNNs) for processing sequential or hierarchical data, generative adversarial networks (GANs) for creative tasks, and transformers for capturing dynamic contextual shifts—this collaboration expands the research scope of multimodal data.

These models can also be applied to infer relationships between promotional materials and political activities, such as analyzing the proportion of text to images in propaganda forms, the strategic messaging conveyed through text in images, the differences in compositional objects reflecting various political stances, and the strategic significance of color schemes in political party interactions.[8] Beyond facilitating a deeper understanding of specific events, this approach extends foundational methodologies for studying propaganda strategies, while also providing a structured research framework for computational social science.

Mapping Political Influence: AI for Stance Detection

A more advanced frontier in disinformation analysis is **stance detection**—understanding the ideological leanings of online content and the communities who spread it.

In Taiwan, Facebook fan pages often play crucial roles in political communication. Using **social network analysis**, researchers can:

- Analyze user interactions (likes, shares, comments).
- Map networks of pages, users, and themes.
- Detect shifts in sentiment or surges in polarizing content.

These methods use unsupervised machine learning and graph mining to visualize ideological clusters, revealing not only where disinformation is coming from but how it spreads. It helps policymakers and civil society intervene early when polarizing content begins gaining traction.[9]

The Defense: Taiwan's Growing Arsenal of Countermeasures

Taiwan, confronting persistent threats from disinformation campaigns, has adopted a proactive and forward-thinking approach to safeguard its democratic society, leveraging cutting-edge **technology** alongside **policy**, **education**, and **international cooperation**. By integrating innovative tools with community engagement and global partnerships, Taiwan is building a resilient framework to protect its information ecosystem from malicious interference, positioning itself at the forefront of combating digital disinformation. The following outlines the key pillars of Taiwan's defense strategy and their implementation:

1. Fact-Checking Networks

Taiwan has cultivated a robust ecosystem of independent fact-checking organizations, including platforms such as the **Taiwan FactCheck Center**,[10] **MyGoPen**,[11] and **Cofact**,[12] which play a pivotal role in combating viral falsehoods.[13] These groups employ a combination of expert analysis and AI-driven tools to swiftly verify information, and refute it through clarification mechanisms such as expert opinions or narrative bulletins. For instance, during election periods, these networks collaborate with local media to provide real-time corrections, reducing the harm of disinformation to society. Their efforts are further supported by community reporting mechanisms, where citizens can flag suspicious content, creating a participatory model of truth-seeking that enhances public trust in verified information.

2. AI-Driven Monitoring Tools

New tools are being developed that scan social media in real-time, flag suspicious posts, and detect coordinated inauthentic behavior. These use AI models trained on local datasets for better accuracy. To enhance their

effectiveness, these tools are designed to adapt to the unique linguistic and cultural nuances of Taiwan's online environment, ensuring that they can identify disinformation tactics that might otherwise go unnoticed in a globalized context. For instance, integrating propaganda template multimodal detection model, incorporate comprehensive analysis of text, images, physical objects and other elements into the system.[14] These models are trained on datasets that include local slang, political terminology, and historical references specific to Taiwan, allowing them to better distinguish between legitimate discourse and manipulative content. Additionally, these designs could incorporate human-in-loop interactions in real-time feedback mechanisms to refine the models' accuracy over time, creating a dynamic system that evolves with emerging threats. This approach not only improves detection capabilities but also enables proactive intervention, helping to curb the spread of false information during critical moments such as elections or public health crises.

3. Public Education

Recognizing that technology alone cannot combat disinformation, organizations in Taiwan have launched extensive digital literacy campaigns across schools, universities, and public institutions to empower its citizens with critical thinking skills. These initiatives teach individuals how to spot false information, question the credibility of sources, and critically evaluate online content before sharing it. Supporting channels include public activities, social platforms and media programs, focusing on common disinformation tactics, such as emotionally charged fake news or manipulated images. By fostering a culture of media literacy, Taiwan aims to build a more discerning public that can resist manipulative narratives, particularly during politically sensitive periods like elections, where disinformation campaigns are most prevalent.[15]

4. International Collaboration

Taiwan actively collaborates with democratic allies, research institutions, and international non-governmental organizations to strengthen its defenses against disinformation. Through these partnerships, Taiwan shares best

practices, exchanges data on emerging threats, and learns from global experiences in countering information warfare. Additionally, Taiwan works closely with major tech platforms like Meta and Google to improve response times to false content, advocating for faster removal of harmful posts and greater transparency in algorithmic processes. These collaborative efforts not only enhance Taiwan's technical capabilities but also reinforce its position within a global network of democracies committed to protecting the integrity of information spaces.[16]

Looking Ahead: The Role of Next-Generation AI

Looking ahead, future research will build upon these findings to develop more sophisticated analytical models. Specifically, by incorporating modern advanced methodologies such as large language models (LLMs) and retrieval-augmented generation (RAG), we aim to expand the dimensionality of data analysis, thereby enhancing both the interpretability and predictive accuracy of disinformation detection frameworks.

These methodological innovations will enable us to better capture the contextual nuances of political discourse across multiple modalities and temporal dimensions. The integration of LLMs will facilitate deep semantic understanding of textual content beyond surface-level features, while RAG systems will support evidence-based reasoning through dynamic knowledge retrieval from verified information repositories. For instance, in the coordination process between a LLM and RAG, the retriever initially identifies and selects relevant documents from the knowledge base based on the user's query. Subsequently, the LLM generator leverages these retrieved results as contextual input to produce coherent text. This integrated approach effectively mitigates the risk of hallucinations while enhancing the factual accuracy of the output.[17]

This multi-faceted approach not only promises to improve technical performance metrics but also strengthens the ecological validity of computational social science research in real-world information environments. Furthermore, we envision developing transferable frameworks that can adapt to emerging disinformation tactics and cross-cultural contexts, ultimately

contributing to more resilient digital information ecosystems.

Reflection on the risks of using generative AI should be incorporated into relevant developments, including reliability, over-dependence, privacy and security, impact on scholarly publishing, and concerns about employment in generative AI.[18]

Conclusion: A Digital Shield for Democracy

The war for Taiwan's future may never be fought with tanks or aircraft—but it is already being fought on phones, feeds, and screens. Disinformation is a quiet threat, but a dangerous one, eroding trust, poisoning debate, and weakening national identity from within. This cognitive battlefield, however, is intricately linked to the physical realm, as the vulnerability of critical infrastructure underpins the effectiveness of information warfare. For an island nation like Taiwan, the reliance on undersea cables as a single point of failure poses a strategic risk, amplifying the potential impact of cyberattacks or sabotage that could sever digital connectivity. By integrating the physical and cognitive battlefields, we see how Taiwan's unique geopolitical role—caught between global powers and heavily dependent on digital networks—makes it particularly susceptible to hybrid threats that exploit both infrastructure weaknesses and societal divisions, threatening its democratic resilience on multiple fronts.

Emerging technologies have empowered the spread of false information, but they can also be our best tools for fighting back. Taiwan's experience offers a blueprint for other nations facing similar challenges: invest in research, build resilient digital systems, educate the public, and never underestimate the power of truth in the face of deception. The construction of resilient digital systems must necessarily correspond to the resistance of vulnerable critical infrastructure. Strengthening external connectivity effectiveness through alternative solutions such as satellite communications will enhance the ability of the digital ecosystem to resist hybrid threats.

The volume of information disseminated through social media platforms

is vast and complex, rendering traditional human-driven or quantitative observation methods insufficient for effective analysis. This chapter has focused on the detection of disinformation, highlighting the potential of integrating and advancing computational technologies to enhance the depth and efficiency of fact-checking processes. These advancements aim to mitigate the societal impact of false information by offering robust, data-driven solutions.

As we move forward, the battle against disinformation will not be won by technology alone, but by the values we embed within it—transparency, accountability, and democracy.

Endnotes

1 J. Nye, "How sharp power threatens soft power," *Foreign Affairs*, January 24, 2018.

2 G. H. Karlsen, "Divide and rule: ten lessons about Russian political influence activities in Europe," *Palgrave Communications 5*, no. 1 (2019): 1–14.

3 A. Bankole, "Bridging Divides: Navigating Political Polarization In Taiwan's Democratic Process," *International Journal Of History And Political Sciences 4*, no. 4 (2024): 1–4.

4 M. Mälksoo, "Memory must be defended': Beyond the politics of mnemonical security," *Security Dialogue 46*, no. 3 (2015): 221–237.

5 K. C. Kurata, "Behind Every Good Lie is a Grain of Truth: Deriving Identity-Based Demand for Disinformation in Moldova and Taiwan Using GIS Applications," *Journal of Information Policy 14* (2024): 35–103.

6 A. B. López, J. Pastor-Galindo, and J. A. Ruipérez-Valiente, "Frameworks, modeling and simulations of misinformation and disinformation: a systematic literature review," *arXiv preprint arXiv:2406.09343*, 2024.

7 M. Nasser, N. I. Arshad, A. Ali, H. Alhussian, F. Saeed, A. Da'u, and I. Nafea, "A systematic review of multimodal fake news detection on social media using deep learning models," *Results in Engineering 26*, 104752, 2025.

8 M. H. Wang, W. Y. Chang, K. H. Kuo, and K. Y. Tsai, "Analyzing image-based political propaganda in referendum campaigns: from elements to strategies," *EPJ Data Science 12*, no. 1 (2023): 29.

9 K. H. Kuo, M. H. Wang, H. Y. Kao, and Y. C. Dai, "Advancing stance detection of political fan pages: A multimodal approach," in *Companion Proceedings of the ACM Web Conference 2024*, (May, 2024): 702–705.

10 The Taiwan FactCheck Center (https://en.tfc-taiwan.org.tw) is the first organization in the Chinese-speaking and East Asian region to receive International Fact-Checking Network (IFCN) accreditation.

11 The MyGoPen (https://www.mygopen.com) is an important and valuable fact-checking and media literacy promotion organization. It plays an important role in the information environment and public education promotion of Taiwan society.

12 The Cofact (https://cofacts.tw) is a suspicious information verification platform that checks facts through mass collaboration and chatbots.

13 W. S. Li, Y. C. Lu, W. K. Hsiao, Y. Y. Tseng, and M. H. Wang, "DRM-SN: Detecting Reused Multimedia Content on Social Networks," in *2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, (August, 2024): 169–175.

14 M. H. Wang and Y. I Chen, "Beyond Text: Detecting Image Propaganda on Online Social Networks," *IEEE Transactions on Sustainable Computing* 10, no. 1 (2025): 120–131.

15 The Taiwan FactCheck School (https://school.tfc-taiwan.org.tw) integrates all the educational resources of Taiwan FactCheck Center to enhance the public's recognition ability.

16 A brief webpage of the Department of International Cooperation of the Ministry of

Digital Affairs (MODA) (https://moda.gov.tw/en/aboutus/functions/444) and a news page in partnership with Google, LINE, and Meta (https://moda.gov.tw/en/press/press-releases/13876).

17 F. Mumuni and A. Mumuni, "Explainable artificial intelligence (XAI): from inherent explainability to large language models," *arXiv preprint arXiv:2501.09967*, 2025.

18 K. B. Wagman, M. T. Dearing, and M. Chetty, "Generative AI Uses and Risks for Knowledge Workers in a Science Organization," in *2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*, (April 2025): 1-17.

2. Reconfiguring AI Governance for Sovereign AI: A Comparative Analysis of Global Approaches

Chih-hsing Ho

Introduction

The global competition for artificial intelligence (AI) dominance has extended well beyond technological innovation, becoming closely intertwined with geopolitics, economic strategy, and national security. As large language models (LLMs) and generative AI technologies advance at a rapid pace, AI has become a critical asset, shaping political and economic dynamics worldwide. In this new era of AI-driven competition, the concept of Sovereign AI has risen to prominence. For many nations, attaining sovereignty over AI governance extends beyond boosting national competitiveness; it represents a critical strategic priority linked to data security, information control, and maintaining socio-political stability. This chapter explores the concept of Sovereign AI, analyzes strategic approaches to AI governance from a comparative perspective, and evaluates the challenges and opportunities Taiwan faces in the evolving AI geopolitical landscape. By examining different AI governance models, this analysis offers insights into how governments can navigate the delicate balance between technological innovation and regulatory frameworks.

Sovereign AI in Global Geopolitics

Sovereign AI refers to a nation's ability to independently control the entire lifecycle of AI technologies—from data collection, model training, and algorithm design to the final deployment of AI applications—without relying on foreign enterprises or external technologies.[1] The primary goal of sovereign AI is to safeguard national data sovereignty, maintain the security of critical infrastructure, and ensure autonomous control over information flow. In

recent years, the race for sovereign AI has become a critical battleground in the global technology competition, as nations increasingly recognize AI as a next-generation strategic asset with profound implications across multiple domains. AI is not merely a tool for technological advancement; it is a driving force behind the Fourth Industrial Revolution. By significantly boosting productivity and innovation, AI offers countries with advanced technologies a substantial competitive edge in global markets.

Beyond economic benefits, sovereign AI is also essential to national security. Modern defense systems and cybersecurity strategies rely heavily on AI-driven technologies. Autonomous systems enhance military capabilities, while predictive analytics and machine learning improve threat detection and strategic decision-making. Nations that develop their own AI technologies reduce the risks associated with relying on foreign systems, which could compromise critical defense infrastructures or expose sensitive information to external manipulation. In addition, in an era where information warfare and digital manipulation are growing threats, sovereign AI also plays a vital role in maintaining societal stability. The rise of AI-generated content, such as deepfakes and fake news, has heightened the potential for external entities to influence public opinions and disrupt democratic processes. By governing AI technologies domestically, countries can safeguard their media environments and reduce vulnerabilities to disinformation campaigns, thereby preserving the integrity of their democratic institutions.[2]

The pursuit of sovereign AI has therefore become a new battleground in the global competition among major powers, reshaping the geopolitical landscape. The European Union (EU), the United States (U.S.), and China are making significant investments in developing independent AI capabilities or establishing regulatory frameworks, recognizing that technological sovereignty is not merely about economic competitiveness but also a critical strategic asset in maintaining global influence. In addition, countries that achieve self-reliance in AI technology are better positioned to harness these economic benefits without dependency on external powers. This autonomy reduces vulnerability to external pressures, such as export restrictions

or technological embargoes, which could stymie economic development. For example, the U.S. has imposed export controls on high-performance computing chips, aiming to limit China's AI development capabilities and maintain its own competitive edge. Such measures illustrate how sovereign AI and AI governance are increasingly intertwined.

Diverse Approaches to AI Governance

The current divergence in AI governance frameworks across the globe underscores the difficulty of defining AI governance without considering the underlying values, priorities, and objectives that shape its implementation. AI governance is not merely a set of technical regulations; rather, it is a political, ethical, and economic construct that reflects the interests of different stakeholders. The fundamental question of what AI governance is for and whose interests it serves plays a decisive role in determining the choice of regulatory models, enforcement mechanisms, and ethical considerations.[3] Different governance models emerge depending on whether AI regulation is primarily intended to advance state-led technological development, ensure national security, or safeguard individual rights and democratic values. In authoritarian or state-driven models, AI governance tends to prioritize government control and strategic dominance in AI innovation. This approach can be observed in countries where AI governance is deeply intertwined with state policy, often involving strict surveillance mechanisms, centralized data access, and government-led AI initiatives aimed at enhancing national competitiveness. In such cases, governance frameworks are designed to maximize AI's economic and geopolitical potential, sometimes at the cost of individual freedoms and privacy.

In contrast, rights-based AI governance frameworks, as seen in democratic societies, emphasize ethical AI development, data protection, and individual rights preservation. These frameworks often focus on transparency and accountability, ensuring that AI systems do not reinforce biases or compromise fundamental freedoms. The EU's AI Act is a prominent example of such an approach, where governance mechanisms impose strict compliance requirements on AI developers and users to prevent ethical violations and

protect citizens from AI-related harms. Here, AI governance serves the broader public interest by prioritizing human rights, privacy, and non-discrimination. Meanwhile, market-driven AI governance models, such as those prevalent in the U.S. take a more decentralized approach, where industry self-regulation and sector-specific policies play a significant role. Rather than a single comprehensive AI regulatory framework, the U.S. governance model relies on private sector innovation, voluntary compliance standards, and a patchwork of regulations that vary by industry (e.g., finance, healthcare, defense). In this case, AI governance serves to stimulate innovation, maintain competitive advantage, and balance regulation with technological progress, often giving companies greater autonomy in shaping their own AI policies.

The divergence in AI governance models raises critical questions about global AI policy alignment and the feasibility of establishing universal AI governance standards.[4] As AI technologies continue to advance, international cooperation will be necessary to bridge regulatory gaps, address cross-border AI challenges, and ensure that governance frameworks promote ethical AI development while reflecting diverse geopolitical and cultural contexts. Without recognizing the foundational values and interests shaping AI governance, any attempt to define a singular, global AI governance model will remain incomplete and insufficient to address the complexities of AI's impact on society.

AI Ethics and Accountability: Competing Visions

The governance of AI is deeply shaped by varying ethical values, legal traditions, and societal priorities across different regions. These differences manifest in how countries regulate AI ethics, assign liability for AI-related harms, and enforce accountability mechanisms for AI-driven decisions. While some jurisdictions emphasize human rights, transparency, and fairness, others focus on economic growth, national security, or state control as the guiding principles of AI governance. These divergences create inconsistent regulatory environments, leading to challenges in international AI governance harmonization and cross-border compliance.[5] Ethical AI governance is a cornerstone of regulatory discussions worldwide, yet different

countries prioritize different moral and legal foundations when determining what constitutes responsible AI behavior. In Europe, AI ethics frameworks are deeply rooted in human rights principles. The EU AI Act, for instance, adopts a risk-based classification system, ensuring that AI applications posing significant risks—such as biometric surveillance or algorithmic hiring systems—are subject to strict compliance requirements. Ethical AI in this model emphasizes non-discrimination and explainability, ensuring that AI does not reinforce social biases or infringe upon fundamental rights.

Conversely, in China, AI ethics are closely tied to state interests, prioritizing national security and economic progress. Rather than emphasizing individual rights, AI ethics in China focus on the collective good as defined by the state, with AI governance frameworks mandating alignment with government policies and controlled information dissemination. This approach is evident in AI-driven censorship systems, mass surveillance technologies, and algorithmic content moderation, where ethical considerations are framed through the lens of social harmony and political stability rather than personal freedoms. In the U.S., AI ethics take a more market-driven approach, where corporate responsibility and voluntary compliance play a central role in shaping governance. Unlike the EU's rights-based model or China's state-driven governance, AI ethics in the U.S. are often guided by industry standards and sector-specific oversight. This decentralized approach allows for greater flexibility in AI innovation, but it also raises concerns about inconsistent ethical enforcement and the prioritization of profit-driven motives over fairness and accountability. Companies largely define their own ethical AI guidelines, leading to fragmented governance structures that may lack enforceability.

1. AI Liability: Who Takes Responsibility for Failures?

Assigning liability for AI decisions remains one of the most complex and unresolved challenges in AI governance. In traditional legal systems, liability is assigned to individuals or entities that cause harm. However, AI systems introduce a new layer of complexity, as they can act autonomously and unpredictably. This raises critical questions: Who should be held responsible

when an AI system makes an erroneous medical diagnosis, causes financial loss, or results in discriminatory hiring practices? In Europe, AI liability laws are shifting towards strict liability models, where AI developers and deployers bear greater responsibility for the actions of their systems.[6] The proposed EU AI Liability Directive seeks to simplify the process for victims to claim compensation for AI-related harm, even when the decision-making process of AI systems is opaque.[7] This aligns with the EU's broader commitment to consumer protection and corporate accountability, ensuring that individuals harmed by AI decisions have legal recourse.

By contrast, China's AI liability framework is closely intertwined with state oversight and government control, particularly in high-risk AI applications such as facial recognition and cybersecurity. While companies and developers may bear some legal responsibility, liability often extends to government regulators and entities managing AI infrastructure. This model ensures state control over AI governance but may limit individual legal recourse against government-mandated AI deployments. In the U.S., AI liability remains largely case-dependent and sector-specific, with different regulatory agencies overseeing AI applications in fields like finance, healthcare, and autonomous vehicles. Product liability laws, contract laws, and tort principles play a role in determining liability, but the lack of a unified federal AI liability framework creates regulatory uncertainty. As a result, many legal disputes over AI failures are addressed through litigation rather than proactive regulatory oversight, leading to unclear precedents and inconsistent enforcement.

2. AI Accountability: Centralized vs. Decentralized Models

The mechanisms for ensuring AI accountability also vary significantly across jurisdictions. In Europe, accountability is enforced through strong regulatory oversight and transparency mandates, particularly in high-risk AI applications. AI developers and deployers must demonstrate compliance with fairness, safety, and human oversight requirements, ensuring that AI-driven decisions remain explainable and contestable. AI models are subject to impact assessments and ongoing compliance checks, making accountability a proactive legal requirement rather than a reactive enforcement measure. In China, AI

accountability is highly centralized, with government authorities playing a direct role in monitoring and guiding AI behavior. AI systems are often required to be registered, approved, and regularly updated in compliance with state-mandated ethical and security guidelines. Private companies developing AI in China must align their models with government priorities, and failure to do so can result in regulatory penalties or market restrictions. This model ensures strong state control over AI applications but limits independent accountability measures and external scrutiny.

In contrast, the US' AI accountability framework is largely decentralized, with companies self-regulating their AI models based on industry best practices. While certain agencies, such as the Federal Trade Commission (FTC) and Food and Drug Administration (FDA), provide oversight for AI in consumer protection and medical applications, many AI-driven industries operate without standardized accountability frameworks. This reliance on corporate self-regulation has been criticized for allowing companies to prioritize profitability over ethical AI practices, leading to concerns about lack of transparency and difficulties in contesting AI-driven decisions. The stark differences in AI ethics, liability, and accountability highlight the broader geopolitical and legal divergences shaping AI governance globally. While some jurisdictions emphasize rights-based governance and strong regulatory oversight, others prioritize state control or corporate self-regulation, leading to vastly different AI governance models. These differences complicate global AI policy harmonization, creating challenges in areas such as cross-border AI deployments and international liability claims.

Challenges in Global Harmonization of AI Governance

While the idea of aligning AI governance principles across jurisdictions is both logical and necessary, achieving such harmonization is far more complex than it may appear. AI is no longer just a technological innovation; it has become a strategic asset that influences economic power, national security, and global competition. The rise of AI as a geopolitical tool means that nations are less inclined to compromise on their AI strategies, making international cooperation in AI governance highly challenging. Even though

global AI safety standards and ethical frameworks could help mitigate risks and enhance accountability, deep-rooted political and economic tensions between countries make regulatory convergence unlikely in the near future.[8] A major obstacle to AI harmonization is the geopolitical rivalry between AI superpowers, particularly the U.S. and China. Both nations are investing heavily in AI research and infrastructure, recognizing AI's potential to shape global influence. The U.S. prioritizes AI innovation through private sector leadership and minimal federal intervention, while China integrates AI into state-led governance and national security frameworks. Given these conflicting models, the likelihood of establishing mutually agreed-upon AI regulations is minimal. Cooperation would require both countries to make regulatory concessions, which neither is willing to do, as it could weaken their strategic AI advantages.

The EU, while advocating for strict AI ethics and human rights protections, finds itself in a difficult position. The EU AI Act sets high compliance standards for AI, aiming to become the global benchmark for AI regulation. However, the EU's stringent approach risks reducing its competitiveness in the global AI race, as companies may prefer to develop AI technologies in regions with more business-friendly regulatory environments. This divergence further complicates international cooperation, as regions with stricter AI laws may struggle to align with those that prioritize AI innovation and market growth over rigid compliance measures. Emerging economies also face unique challenges in AI governance, as many developing nations lack the infrastructure, expertise, and regulatory capacity to enforce AI policies. These countries often depend on AI investments from global tech giants or foreign governments, making them vulnerable to the influence of external AI regulations.[9] For instance, African and Southeast Asian nations that adopt Chinese AI-driven surveillance technologies may find themselves indirectly aligning with China's AI governance model. Meanwhile, nations that rely on Western tech firms for AI development may need to navigate conflicting regulatory expectations from the U.S. and EU, making true regulatory harmonization even more elusive.

Furthermore, the militarization of AI and national security concerns make it unlikely that countries will agree on common AI governance frameworks. Many nations are developing autonomous weapons systems and cyber warfare capabilities, areas where transparency and international cooperation are highly sensitive. Given the national security implications of AI, governments are reluctant to disclose details about their AI capabilities or submit to internationally imposed AI governance rules. This lack of trust among global powers hinders collaboration and increases the risk of an AI arms race, where countries prioritize AI dominance over ethical or regulatory cooperation. Another critical issue is the fragmentation of AI regulatory efforts at international forums. Various global institutions, such as the United Nations, the Organization for Economic Co-operation and Development, the G7, and the World Economic Forum, have proposed frameworks for ethical AI governance. However, these initiatives lack enforcement power, as participation is voluntary and binding regulations are absent. Unlike global agreements on nuclear non-proliferation or trade laws, AI governance lacks a universal regulatory body with authority to enforce compliance. As a result, nations continue to pursue individual AI strategies, leading to a patchwork of regulations that make cross-border AI governance inconsistent and difficult to enforce.

Taiwan's Role and Challenges in AI Governance

Taiwan's global leadership in the semiconductor industry positions it as a critical player in the AI hardware supply chain. Taiwan Semiconductor Manufacturing Company (TSMC) holds a near-monopoly in the production of advanced AI chips, such as NVIDIA GPUs, making Taiwan an indispensable hub for global AI computing power. This dominance in semiconductor manufacturing offers Taiwan a strategic advantage in the global AI landscape, particularly as the demand for high-performance computing continues to surge. However, despite its strength in hardware, Taiwan faces significant challenges in AI software development, data governance, and the creation of large language models (LLMs). Unlike the U.S., which leads in software innovation and AI applications, Taiwan's AI ecosystem is still in a nascent stage, struggling to achieve technological independence in the software

domain. The development of Taiwan's domestic LLM model, "TAIDE" (Trustworthy AI Dialog Engine),[10] exemplifies these challenges. The TAIDE project aims to establish a sovereign AI model tailored to local needs, yet it has encountered obstacles such as limited access to traditional Chinese-language datasets, insufficient funding, and a lack of computing resources.

In addition, a significant obstacle to Taiwan's AI ambitions is policy fragmentation. Various governmental agencies—including the National Science and Technology Council, the Ministry of Digital Affairs, and the Ministry of Education—manage AI-related budgets and initiatives independently. This silo approach often results in resource misallocation and overlapping projects. Establishing a centralized coordinating body could help streamline efforts and develop a cohesive strategy for AI ecosystem development. Such consolidation would not only improve efficiency but also expedite progress toward building a robust sovereign AI infrastructure.

Geopolitical factors further complicate Taiwan's pursuit of AI sovereignty. Taiwan's semiconductor industry is a focal point in the broader U.S.-China technology rivalry, exposing it to geopolitical risks and market dependencies. To mitigate these risks, Taiwan could benefit from strategic partnerships with like-minded countries, promoting technological collaboration while maintaining a balanced stance in the AI domain. Such an approach could diversify Taiwan's markets and bolster its technological and strategic positioning on the global stage.

In terms of AI governance, Taiwan has adopted a "guidance-before-legislation" strategy to manage the rapidly evolving AI landscape.[11] This approach involves issuing non-binding guidelines as a preliminary step, allowing regulatory agencies and industries to adapt to AI developments flexibly. By delaying formal legislation until necessary, Taiwan aims to avoid rigid regulatory frameworks that might stifle innovation or quickly become outdated. However, by prioritizing non-binding guidelines over formal legislation, Taiwan risks creating a fragmented regulatory environment where compliance is voluntary, and accountability is diluted. While this

strategy allows for flexibility, it may also lead to inconsistent enforcement across sectors, giving rise to loopholes and uneven regulatory standards. In addition, the forthcoming "Artificial Intelligence Basic Act," anticipated for legislative review in the near future, represents Taiwan's effort to formalize AI governance.[12] Nevertheless, the delay in introducing a comprehensive legal framework exposes Taiwan to potential risks, including unchecked AI deployment and ethical oversights. While the Act is built on commendable principles—such as transparency, privacy, autonomy, fairness, cybersecurity, sustainable development, and accountability—its impact will depend heavily on the robustness of its implementation and enforcement mechanisms.

There is also a question of whether this legislation will merely codify existing guidelines or genuinely introduce rigorous standards that can withstand the complex and rapidly evolving AI landscape. Without strong regulatory teeth, the Act might struggle to address high-stakes issues such as bias in AI algorithms, data privacy breaches, and accountability for AI-driven decisions. Additionally, Taiwan's focus on "guidance-before-legislation" could undermine the Basic Act's authority if non-binding guidelines continue to dominate AI governance, creating a scenario where the legal framework is more aspirational than operational. Ultimately, while Taiwan's approach offers a degree of flexibility, it also demands critical evaluation to ensure that it does not compromise the country's ability to effectively regulate AI technologies. The success of Taiwan's AI governance will hinge on its willingness to transition from a guidance-based model to a more decisive regulatory framework that not only encourages innovation but also rigorously protects individual and societal interests.

Conclusion

As the global race for AI dominance intensifies, the pursuit of sovereign AI has emerged as a critical strategic priority for nations seeking to secure technological and geopolitical advantages. This chapter has examined how different countries approach AI governance, revealing a spectrum of strategies that balance innovation with regulatory control. From the proactive yet fragmented guidance-before-legislation model in Taiwan to the stringent

regulatory frameworks in Europe and the dual-track innovation-market approach of the U.S., each model offers unique insights into the complex relationship between AI governance and national interests.

For Taiwan, navigating the evolving AI geopolitical landscape presents both challenges and opportunities. While Taiwan benefits from its strong hardware manufacturing base, particularly in semiconductors, it faces hurdles in software development and policy coherence. The anticipated AI law marks a crucial step towards establishing a more structured AI governance framework, yet its success will depend on effective implementation and enforcement. Taiwan's ability to consolidate its policies, enhance inter-ministerial coordination, and reduce geopolitical dependencies will be essential in achieving true AI sovereignty. Ultimately, the analysis underscores that achieving sovereign AI is not solely about technological self-reliance but also about embedding robust ethical, legal, and governance frameworks that can adapt to rapid advancements. As AI continues to reshape global political and economic dynamics, countries that strike the right balance between innovation and regulation will be better positioned to harness AI's potential while safeguarding their national interests. For Taiwan, adopting a more cohesive and forward-thinking approach to AI governance could transform its current challenges into strategic advantages, solidifying its role as a resilient player in the global AI arena.

Funding

This research is funded by the Academia Sinica Grand Challenge Programme Seed Grant Project with the grant number: AS-GCS-113-H04.

Endnotes

1 A. Bradford, *Digital empires: The global battle to regulate technology* (Oxford University Press, 2023).

2 M. Coeckelbergh, *Why AI Undermines Democracy and What To Do About It* (Polity Press, 2024).

3 R. B. L. Dixon, "A principled governance for emerging AI regimes: lessons from China, the European Union, and the United States," *AI and Ethics 3*, no. 3 (2023): 793–810.

4 L. Schmitt, "Mapping global AI governance: a nascent regime in a fragmented landscape," *AI and Ethics 2*, no. 2 (2022): 303–314.

5 M. Veale, K. Matus, and R. Gorwa, "AI and global governance: modalities, rationales, tensions," *Annual Review of Law and Social Science 19*, no. 1 (2023): 255–275.

6 C. Wendehorst, "AI liability in Europe: anticipating the EU AI Liability Directive," Ada Lovelace Institute, 2022.

7 G. Borges, "Liability for AI Systems Under Current and Future Law: An overview of the key changes envisioned by the proposal of an EU-directive on liability for AI," *Computer Law Review International 24*, no. 1 (2023): 1–8.

8 H. Roberts, E. Hine, M. Taddeo, and L. Floridi, "Global AI governance: barriers and pathways forward," *International Affairs 100*, no. 3 (2024): 1275–1286.

9 N. Emery-Xu, R. Jordan, and R. Trager, "International governance of advancing artificial intelligence" *AI & Society*, (2024): 1–26.

10 "Nation's AI dialogue engine highlighted" *Taipei Times,* May 4, 2024, https://www.taipeitimes.com/News/taiwan/archives/2024/05/04/2003817360.

11 Digital Innovation & Governance Initiative Committee, "Ministry releases guidelines for AI research," September 25, 2019, https://digi.nstc.gov.tw/en/BD28C14C8FBBF163/0ddb3598-4fa0-48a6-97a1-5563ac07deca

12 Lin Hsin-han and Esme Yeh, "AI fundamental bill passes first stage," *Taipei Times*, August 5, 2025, https://www.taipeitimes.com/News/taiwan/archives/2025/08/05/2003841500.

3. AI Development and Governance: Navigating Trust, Transparency, Innovation, and the Challenges of Information Warfare

Vera Schmitt

Introduction

In recent years, artificial intelligence (AI) has revolutionized digital communication, governance, and economic systems. The rise of AI-driven technologies has significantly impacted information trust, transparency, innovation, and accountability. Especially in the domain of Natural Language Processing (NLP), significant advancements have been achieved by relying on transformer-based architectures,[1] which have enabled Large Language Models (LLMs) to process text data more effectively. LLMs like GPT-4o, LLaMA, Deepseek, and others are built on these architectures, fundamentally reshaping human-AI interaction. While these models offer opportunities for more efficient information retrieval and improved question-answering, they also pose challenges regarding governance, accountability, and ethical considerations.

In an era where AI-generated content influences public discourse, robust governance mechanisms are essential to ensure transparency, mitigate disinformation, and uphold democratic values. The release of the LLM DeepSeek R1[2] and the ensuing debate on open-source architectures, censorship, and data access highlight the unresolved and ongoing questions surrounding AI's role in society. Many of these questions remain unaddressed or unresolved at the international level, with only a few attempts at AI regulation emerging. As AI becomes an increasingly strategic asset in global information ecosystems, its potential misuse in information warfare, including disinformation campaigns and geopolitical influence operations,

raises pressing concerns for regions like the European Union (EU) and Taiwan, where emerging technologies shape both security landscapes and democratic resilience.

This chapter explores the intersection of AI technology development and governance, examining how AI-driven systems shape the digital information landscape. It provides a comparative analysis of governance frameworks in EU and Taiwan, assesses the role of AI in fostering trust and transparency, and discusses regulatory measures ensuring accountability in AI-driven innovation.

Transforming the Information Ecosystem: The Impact of LLMs

The integration of LLMs into the digital information ecosystem marks a paradigm shift, influencing not just content creation and distribution but also the mechanisms of trust and authority in public discourse. While AI has historically been utilized to improve information accessibility, the scale and sophistication of LLMs introduce both opportunities and vulnerabilities that require attention. Beyond their technical advancements, LLMs are redefining power dynamics in global communication, influencing political, economic, and social structures in unprecedented ways. Particularly in regions like the EU and Taiwan, where digital governance intersects with democratic resilience, understanding these impacts is important for shaping adaptive regulatory frameworks and safeguarding information integrity. LLMs represent a significant leap in AI capabilities, built upon transformer architectures that allow for highly efficient processing and generation of text. These models have led to breakthroughs in automated content creation, sentiment analysis, and multilingual communication, offering new tools for journalism, policymaking, and corporate communication. They also empower fact-checking initiatives by rapidly verifying claims and detecting inconsistencies across vast datasets.

However, these same capabilities make LLMs a potent tool for mis- and disinformation. Automated text generation can be used to produce false

narratives at scale, seamlessly adapting to regional linguistic and cultural nuances. In Taiwan, AI-driven disinformation campaigns have seen a tremendous increase in content generation targeting social media users, using synthetically produced narratives that blend factual elements with manipulative distortions.[3]

Similarly, in the EU, LLMs facilitate the rapid translation and dissemination of disinformation across the EU's multilingual environment, exploiting linguistic diversity to create tailored propaganda efforts. LLMs have become central to information warfare strategies. State and non-state actors deploy these models to influence public discourse, shape election outcomes, and manipulate geopolitical narratives. In Taiwan, coordinated influence operations frequently exploit social media platforms using AI-generated content to blur the line between genuine and artificial discourse. The linguistic proximity between simplified and traditional Chinese further complicates efforts to detect and counteract disinformation, as adversarial campaigns seamlessly integrate with organic discussions. The EU faces a distinct yet equally pressing challenge. The Russian invasion of Ukraine underscored the role of AI-generated content in modern conflict, with deepfake technologies used to fabricate military incidents, manipulate satellite imagery, and create synthetic media designed to erode trust in democratic institutions. The EU's strategic response has involved reinforcing fact-checking alliances, such as the European Fact-Checking Standards Network (EFCSN), and enhancing cooperation between regulatory bodies and AI researchers to develop robust countermeasures.

Towards a Sustainable AI-Governance Framework

The widespread adoption of LLMs demands a holistic approach to governance, one that balances technological advancement with safeguards against misuse. The EU and Taiwan offer complementary models for achieving this balance. Europe's legislative mechanisms provide structural oversight, while Taiwan's emphasis on civil resilience and participatory digital governance fosters agile responses to evolving threats. A forward-looking strategy must prioritize transparent and meaningful AI development, open auditing processes,

interpretability standards, and ethical guidelines for AI deployment. Robust content verification systems, enhanced forensic tools for detecting synthetic media, coupled with real-time fact-checking networks, but also, public education and awareness, digital literacy programs are important that equip citizens with the skills to critically assess AI-generated content. Moreover, international collaboration and coordinated regulatory efforts to harmonize AI governance standards across democratic nations are essential not only for strengthening national approaches but also for addressing information warfare on a global scale. As LLMs continue to shape the global information landscape, their impact on democratic discourse and security will depend on the effectiveness of these governance mechanisms. The ongoing struggle against AI-driven mis-and disinformation highlights the need for proactive, multi-stakeholder approaches that integrate technological innovation with ethical and regulatory foresight. By addressing these challenges, the EU and Taiwan can serve as models for a resilient digital future, ensuring that AI enhances, rather than undermines, the integrity of public discourse and democratic governance.

The governance of LLMs presents a major regulatory challenge, particularly as debates over open-source versus proprietary AI models intensify. Open-source models like LlaMa and DeepSeek enable greater transparency and broader accessibility, yet they also risk being weaponized for malicious purposes, including automated propaganda and deepfake-driven disinformation. On the other hand, proprietary models managed by corporate entities impose restrictions on data access and model interpretability, raising concerns about censorship, bias, and the monopolization of AI knowledge.

The EU has responded by implementing regulatory frameworks such as the Digital Services Act (DSA) and the AI Act, which establish guidelines for accountability and transparency in digital services. These laws represent the world's most comprehensive efforts to regulate AI-generated content, aiming to mitigate risks while preserving innovation. Taiwan, meanwhile, has emphasized public-private collaboration, integrating digital literacy initiatives with real-time monitoring of AI-generated disinformation. Its civil

society organizations play a crucial role in countering fabricated content, support grassroots fact-checking networks to maintain trust in information ecosystems.

AI Governance and Regulatory Frameworks

AI governance varies across jurisdictions, reflecting different political, ethical, and technological priorities. The EU has pioneered comprehensive AI governance through the AI Act and the DSA, which impose strict, but abstract regulations on high-risk AI applications and platform accountability.[4] These legislative measures emphasize transparency, fairness, and accountability in AI-driven decision-making. Also Taiwan has made progress concerning the governance of AI. Taiwan's AI Basic Act establishes core principles for AI development and application, emphasizing sustainable development, human autonomy, privacy, security, transparency, fairness, and accountability.[5] It outlines the government's role in promoting AI, including infrastructure development, public-private collaboration, and international cooperation.

Hereby, the EU's AI Act, the DSA, and Taiwan's AI Basic Act represent distinct approaches to AI and digital platform regulation, reflecting different priorities and governance philosophies.

The EU AI Act, adopts a risk-based approach, categorizing AI applications into different risk levels (unacceptable, high, limited and minimal) and imposing corresponding legal obligations. High-risk AI systems are subject to stringent requirements, including documentation, transparency, and meaningful human oversight, while non-compliant entities face substantial fines.[6] Complementing this, the DSA regulates online platforms, ensuring accountability for algorithmic decision-making in content moderation, targeted advertising, and the dissemination of dis- and misinformation. By enforcing obligations on very large online platforms (VLOPs) and search engines, the DSA strengthens AI governance by curbing the spread of harmful or manipulative content.

In contrast, Taiwan's AI Basic Act focuses on principle-based governance, emphasizing AI development through ethical guidelines, transparency, fairness, and international cooperation. Rather than enforcing strict risk-based rules, Taiwan's framework promotes innovation-friendly policies, including regulatory sandboxes and government-led initiatives to foster AI growth while maintaining ethical oversight. While both frameworks aim to ensure AI safety and ethical deployment, the EU AI Act and DSA prioritize legal enforceability and risk mitigation, whereas Taiwan's AI Basic Act fosters a more flexible and innovation-driven approach. These differences highlight the ongoing global debate between regulatory control, platform accountability, and AI innovation facilitation in digital governance.

Information warfare, characterized by the strategic manipulation of information to influence public perception, disrupt democratic processes, and erode institutional trust, is increasingly amplified by AI-driven systems and digital platforms. To address these challenges, the EU AI Act, the DSA, and Taiwan's AI Basic Act provide distinct regulatory approaches. The EU AI Act, with its risk-based framework, imposes stringent transparency and accountability requirements on high-risk AI applications, including automated content moderation and decision-making systems, thereby reducing the likelihood of AI-powered disinformation campaigns. Complementing this, the DSA targets online platforms and VLOPs, mandating robust algorithmic auditing, transparency in content moderation, and measures against the spread of harmful and manipulative content. This dual-layered approach ensures that both AI technologies and digital platforms operate with heightened oversight to curb the influence of information warfare. In contrast, Taiwan's AI Basic Act adopts a principle-based and innovation-friendly approach, emphasizing public-private collaboration, ethical AI development, and international cooperation. Instead of imposing rigid risk classifications, Taiwan's framework fosters adaptive regulatory mechanisms, like cross-sectoral cooperation, enabling rapid responses to evolving disinformation threats while supporting AI-driven innovation. By combining strict legal enforcement in the EU with flexible governance in Taiwan, these models offer complementary strategies for building a resilient information ecosystem,

mitigating AI-enabled manipulation, and safeguarding democratic integrity. These initiatives demonstrate the need for adaptable regulatory frameworks capable of addressing the rapidly evolving AI landscape.

Trust and Transparency in the Information Ecosystem

The integrity of the information ecosystem hinges on two fundamental principles: trust and transparency. As AI-driven technologies, particularly LLMs, become more deeply integrated into digital communication, maintaining these principles is essential for preserving public confidence in online information. Transparency in AI systems fosters accountability by enabling users to understand how decisions are made, while trust is cultivated through mechanisms that ensure fairness, verifiability, and reliability of AI-generated content. However, the challenge lies in balancing openness and security, especially as adversarial actors exploit AI for mis- and disinformation and manipulation. In this context, the EU's AI Act, the DSA, and Taiwan's AI Basic Act take distinct but complementary approaches to enhancing trust and transparency in AI governance. The AI Act enforces strict transparency requirements for high-risk AI applications, mandating explainability, auditing, and impact assessments to prevent harmful AI-driven decisions. By compelling tech companies to disclose how recommendation systems function and how content is moderated, the DSA helps users navigate digital spaces with greater confidence. These combined regulatory efforts create a robust framework for ensuring that AI technologies and digital platforms uphold democratic values.

Taiwan's AI Basic Act, while less prescriptive than the EU's regulatory model, prioritizes public-private collaboration and ethical AI governance to foster transparent AI development. Through regulatory sandboxes and international cooperation, Taiwan aims to develop AI technologies that maintain high standards of trustworthiness while remaining adaptable to emerging threats. Its proactive approach to digital literacy further empowers citizens to critically engage with AI-generated content, reducing susceptibility to manipulative narratives.[7] Despite these measures, challenges remain in ensuring long-term trust and transparency in AI systems. One major issue is the opacity of deep learning models, which often function as black boxes,

making it difficult for users and even developers to fully explain their outputs. Addressing this challenge requires continued efforts in eXplainable AI (XAI), where research into interpretable architectures, mechanistic interpretability, natural language explanations, and auditability is essential for bridging the gap between model performance and user comprehension.[8]

However, transparency and trust are oftentimes not enough. With the General Data Protection Regulation (GDPR), the EU has followed an ineffective approach of overloading the user with too much information and leaving the user the choice of sharing personal data or not. This has shown to be an ineffective regulatory approach, as the user is left alone with overcomplicated decision-making. We all tend to accept terms and conditions without reading the privacy policies and without a complete understanding of what type of data is collected. Also, within the AI Act and DSA, the EU has formulated very abstract obligations following the risk categories. These leave a lot of room for interpretation on a technical level and can vary greatly when being implemented. Additionally, monitoring is a key aspect of regulation. When there are no means of automatic testing and evaluation for regulatory compliance, the regulations will remain mostly ineffective. Moreover, trust in AI-generated information cannot be solely a technical or regulatory challenge, it is also a societal challenge. Strengthening institutional credibility, fostering interdisciplinary cooperation, and equipping the public with the necessary critical thinking skills are key steps in building a resilient information ecosystem. The interplay between regulatory enforcement (as seen in the EU), innovation-driven governance (as promoted by Taiwan), and active societal engagement will determine the success of AI governance in preserving trust and transparency in the digital age.

Moving forward, both the EU and Taiwan can serve as leading examples in AI governance by refining their regulatory approaches to align with the evolving nature of AI-driven information ecosystems. By integrating rigorous oversight with participatory governance models, these regions can provide a blueprint for ensuring that AI remains a force for transparency and trust, rather than an enabler of opacity and manipulation.

AI and Innovation: Balancing Technological Advancement and Accountability

AI has become a driving force of technological innovation, reshaping industries, governance, and society at large. However, with its rapid advancement comes a fundamental challenge: how to balance the pursuit of innovation with accountability, ethical responsibility, and regulatory oversight. The ongoing debate surrounding DeepSeek R1 and OpenAI exemplifies this tension, highlighting the contrasting approaches to AI development, one advocating for openness and accessibility, the other emphasizing control and safety.

Open-source LLMs like LlaMa, Gemini, Mistral, Falcon, BLOOM, Qwen, and DeepSeek R1 represent a significant shift towards democratizing AI by making powerful language models openly available to researchers and developers worldwide. Open-source does not necessarily imply unrestricted access to the training data used for developing the models. However, it allows users to modify the architecture and leverage the LLMs for fine-tuning and adaptation to diverse tasks and applications. Proponents of open-source AI argue that transparency accelerates innovation, fosters collaboration, and enables more rigorous scrutiny of AI systems, ultimately making them safer and more reliable. Open access reduces dependence on a few dominant corporations, allowing smaller enterprises and independent developers to contribute to the field. However, while open-source AI fosters inclusivity and knowledge-sharing, it also raises concerns about misuse. Without built-in safeguards, such models can be weaponized for disinformation campaigns, cyber threats, or manipulative social engineering. The absence of centralized oversight makes it difficult to control how these models are used, leading to calls for regulatory intervention or licensing frameworks to ensure responsible deployment.

In contrast, OpenAI has moved toward a more controlled model of AI development, prioritizing ethical constraints, oversight, and safety measures. Initially an advocate for open research, OpenAI has since restricted access to its most advanced models, such as GPT-4o, to prevent potential misuse. Proprietary AI models allow for more effective monitoring, bias mitigation, and compliance with regulatory standards, particularly in high-risk domains

such as finance, healthcare, and national security. By keeping AI capabilities within a controlled environment, OpenAI seeks to prevent harmful applications while ensuring that AI technologies align with human values and ethical considerations. However, this approach has sparked criticism over concerns of monopolization, reduced transparency, and corporate influence over AI governance. Limiting access to AI models may slow scientific progress and concentrate power in the hands of a few entities, raising questions about the fairness and equity of technological advancement. The broader debate between open and proprietary AI models underscores the need for a balanced approach to governance, one that fosters innovation while addressing ethical concerns. A potential path forward involves regulatory frameworks that allow AI research and development to thrive while preventing harmful applications. Some experts advocate for regulatory sandboxes, where AI innovations can be tested in controlled environments before being widely deployed. Others suggest hybrid governance structures in which AI models are partially open but require licensing agreements to ensure accountability. Policies such as the EU's AI Act and Taiwan's AI Basic Act offer structured approaches to AI governance, promoting transparency, fairness, and responsible AI deployment.

Beyond regulation, achieving a balance between innovation and accountability requires interdisciplinary collaboration between governments, academic institutions, and private organizations. Public-private partnerships can play a crucial role in funding ethical AI research, promoting AI literacy, and establishing independent oversight bodies to monitor AI applications. Additionally, advances in XAI may help bridge the gap between AI innovation and public trust, making AI systems more interpretable and reducing the risks associated with opaque decision-making processes. As AI continues to shape global economies and societies, the tension between innovation and ethical responsibility will remain a defining challenge. The contrasting approaches of DeepSeek and OpenAI illustrate the complexity of this debate, demonstrating the need for nuanced governance strategies that balance technological progress with ethical safeguards. Moving forward, fostering a responsible AI ecosystem will require not only regulatory adaptation but also a collective international commitment to ensuring that

AI serves the broader interests of humanity while mitigating its risks. The discussions and regulations are only of limited use on a national or regional level, and international approaches need to be developed. However, national and regional discussions and regulations provide only limited effectiveness, highlighting the need for the development of international approaches.

Conclusion and Future Directions

The rapid advancement of AI has fundamentally reshaped industries, governance, and public discourse, bringing both opportunities and challenges. As AI continues to evolve, its impact on information trust, transparency, and accountability requires ongoing scrutiny and adaptive regulatory frameworks, which offer clear guidance on a technical level and can be implemented in a meaningful way. The balance between innovation and responsibility remains central to AI governance, as seen in the debate between open-source models and proprietary systems. While open AI fosters innovation and accessibility, it also raises concerns about security and misuse. Conversely, proprietary models ensure greater control and oversight but risk consolidating AI power within a few dominant entities. Moving forward, a more harmonized global approach to AI governance is necessary. National and regional regulations, such as the EU AI Act and Taiwan's AI Basic Act, have laid important groundwork, but international cooperation will be essential to address the cross-border implications of AI technologies. Future AI governance efforts should focus on ethical AI deployment, robust auditing mechanisms, meaningful transparency for users, and interdisciplinary collaboration between policymakers, technologists, and civil society. Additionally, advancing meaningful XAI contributing to human reasoning process robustly will be important in bridging the gap between AI transparency and public trust.

As AI becomes an integral part of global information ecosystems, ensuring that it serves humanity's best interests will require continuous regulatory adaptation, technological safeguards, and international cooperation. The path ahead demands a commitment to responsible innovation, one that embraces AI's transformative potential while safeguarding democratic values, security, and ethical principles.

Endnotes

1 D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," *arXiv preprint arXiv:2501.12948*, 2025.

2 A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems,* 2017.

3 Global Taiwan Institute, "Securing U.S. Elections from AI-Enhanced Foreign Influence," October 2024, https://globaltaiwan.org/2024/10/securing-u-s-elections-from-ai-enhanced-foreign-influence/ (accessed January 9, 2025).

4 European Commission, "The Digital Services Act," 2024, https://commission.europa.eu/strategy-and-policy/digital-services-act (accessed April 28, 2025).

5 J. Tseng, and A. Chang, "Taiwan's National Science and Technology Council Has Published the Draft Bill of the Basic Act on Artificial Intelligence for Public Consult," 2024, https://www.klgates.com/Taiwans-National-Science-and-Technology-Council-Has-Published-The-Draft-Bill-Of-The-Basic-Act-On-Artificial-Intelligence-For-Public-Consultation-8-1-2024.

6 V. Schmitt, V. Solopova, V. Woloszyn, and J. D. J. de Pinho Pinhal, "Implications of the New Regulation Proposed by the European Commission on Automatic Content Moderation." In *the 2021 ISCA Symposium on Security and Privacy in Speech Communication,* 47–51, 2021; European Commission, "Artificial Intelligence Act (AI Act)," 2024, https://eur-lex.europa.eu/ (accessed January 17, 2025).

7 C. H. T. Chen, and W. C. Ke, "The Main Impact and Response of Artificial Intelligence in Taiwanese Society," *Journal of Research in Social Science and Humanities* 2, no. 1 (2023): 13–24.

8 L. Longo, et al., "Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions," *Information Fusion* 106 (2024): 102301.

4. The Evolution of Artificial Intelligence in Ukraine's Information Security Landscape

Maya Sobchuk

Introduction

The balance of power in modern warfare increasingly favors those with command of the most advanced technologies, not necessarily those with the largest number of soldiers. Today's conflicts unfold just as aggressively across digital spaces, where truth itself is a contested domain. As a pathway directly into the minds of both those involved in the conflict and those thousands of miles away, the information space has extended the battlefield across the entire world. The rise of artificial intelligence (AI) has added a new and dangerous dimension to this transformation, reshaping not only how wars are fought on the battlefield, but also and how they are waged in the minds of populations. Recognizing this shift, NATO included disinformation under its list of threats posed by AI in July 2024.[1]

The status quo in Ukraine has shown that AI adapted to information warfare is no longer a hypothetical threat. It is already shaping the way Russia is able to exert its influence within Ukraine and to manipulate perception outside of it. Nonetheless, Ukraine—having always been a technological powerhouse—mobilized across sectors to harness the tools to fight back. At the same time, the threat is not limited to Ukraine. Taiwan, facing persistent information threats from another authoritarian power, offers a parallel case. The potential of the link between East Asia and Ukraine cannot be underestimated. Understanding the use of AI in Ukraine's information war is not only essential for defending Ukraine but offers urgent lessons for the protection of other democracies at risk.

This chapter examines the evolution of AI in Ukraine's information security landscape, from pre-full scale invasion information operations to the rapid technological acceleration sparked by the current moment. By analyzing both historical patterns of information warfare against Ukraine and the country's technological countermeasures, I will trace how AI-powered information defense systems have adapted to meet emerging threats. The chapter pays particular attention to the wartime transformation of Ukraine's information security infrastructure, examining both defensive innovations and the sophisticated AI-enabled disinformation campaigns deployed against it. Throughout the chapter, I highlight similarities, lessons, and possible collaborations with Taiwan.

This dual perspective encompassing both historical development and contemporary applications provides crucial insights into the role of AI in modern information warfare and offers valuable lessons for democracies facing similar challenges. It is important to highlight that while Ukraine's experience may be invoked here due to the current moment, every country around the world faces this threat. Ukraine is not a unique case, just an extreme one. The same technologies—AI-generated disinformation, deepfakes, automated propaganda—can and likely will be deployed elsewhere, targeting other societies' vulnerabilities. As adversaries refine these tools in live conflicts, the risk of their wider adoption in peacetime political environments, elections, and social movements grows, underscoring the urgency for democratic nations to study Ukraine not as an exception, but as a warning. Russia very well could have succeeded in achieving its strategic goals of cognitive control over Ukraine without ever resorting to a full-scale invasion. Their actions in the information space are only being studied and recognized in this way because they were followed by kinetic aggression that forced international attention. Other countries may already be facing similar assaults on their information environments without realizing it—recognition often comes too late, only when such aggression manifests in the physical domain. This is why it is critical to bridge the lessons from Ukraine to Taiwan, a democracy currently serving as a testing ground for another authoritarian power intent

on shaping minds, perceptions, and realities without firing a shot.

Historical Context and Evolution

Russian disinformation efforts, as we know them today, date back to the early Soviet era, during which Russia maintained that psychological warfare was a vital element of its national strategy used to keep its subordinate republics in check. Some of its most notable innovations include reflexive control, *maskirovka*, and active measures. Responsible for now infamous operations like Operation Infektion, Russia has continued this legacy into today, only refining and expanding its strategy with the help of modern technological tools.

The overarching goal of Russia's foreign influence operations strategy in the current moment is, above all, to erode global support of Ukraine, both at the public opinion level and policy level. In order to comprehensively discuss and analyze the effect technology and AI have had on the war on perception, we have to discuss narrative warfare and tease out the viewpoints Russia and Ukraine are pushing to the world. Russia's framing centers around their "denazification" narrative, subsisting of a claim that they are rescuing the Russian-speaking population within Ukraine from a genocide. Russian narratives often take a historical angle, calling into question Ukraine's sovereignty and distinct identity. They frame the war as a defensive measure from Western aggression, particularly from NATO expansion. Even though Western social media like Facebook and Twitter (officially known as X since 2023) are banned within Russia, these platforms are ripe with Russian propaganda, demonstrating that Russian influence resources are targeted at foreign audiences. Russia's full-scale invasion showcased the level of rigor the war had now reached in the narrative dimension, with evidence that within that first week "videos from a range of sources on TikTok with the tag #Russia and #Ukraine had amassed 37.2 billion and 8.5 billion views, respectively."[2] The addition of AI to this cocktail, then, is to repeatedly hammer and deliver these viewpoints to the world.

The Chief Executive of the Center for Countering Digital Hate, Imran Ahmed, wrote pointedly in Tech Policy Press that "the justification for the Russian war against Ukraine was built on Facebook."[3] He also noted that "Facebook is failing to label 91% of posts containing Russian propaganda about Ukraine with warnings that the content originates with media outlets owned by the Russian state", allowing falsehoods to circulate virtually unchecked before spreading to YouTube, X, and TikTok. Ahmed's point speaks to the role of social media as a whole in conflict. Russia's willingness to exploit these enforcement gaps underscores how central Western platforms remain to its influence playbook, even when those same platforms are forbidden at home.

The wide array of proxies and the sheer scale of content enable Moscow to operate under plausible deniability, shielding it from direct accountability. This strategy has been operationalized and scaled through the deployment of troll farms, bot networks, and state-linked entities such as the infamous Internet Research Agency (IRA). The sophistication of Russia's coordination across platforms creates a veil of uniformity; users may believe they are sampling a diverse range of perspectives, when in reality every path gently steers them back to the Kremlin's preferred viewpoint. It is a tactic that circles around the aforementioned "reflexive control", which involves shaping the information environment so that the target is led, unknowingly, into adopting beliefs and behaviors that serve the Russian interest.

Despite this focus on specific narratives, Russian information warfare succeeds just as much because it understands its targets. Manipulation is often implicit. Rather than pushing pro-Kremlin talking points in every post, bots and troll farms work overtime to divide and conquer, inflaming existing culture-war issues and amplifying fringe voices. Here AI takes center stage. Large-language-model tools generate credible commentary in multiple languages, sentiment analysis helps tailor messages to micro-audiences, and automated persona management keeps thousands of accounts active at minimal cost. Together, these capabilities allow the Kremlin's long-standing doctrine of reflexive control to thrive in the digital age. It is precisely by

operating within this philosophy that AI is able to maximize damage to the information space.

AI Adaptation in Ukraine and the Repositioning of Tech Firms

Discussions of AI in the context of the war on Ukraine frequently focus on its application in kinetic warfare, where AI has made a tangible impact through the deployment of drones, geospatial intelligence tools, and other advanced military technologies. Far less attention, however, has been paid to AI's growing role in the information space, where it is reshaping how influence operations are conducted and how public perceptions are manipulated. AI tools allow adversaries to generate persuasive false narratives aligned with those reviewed above, deploy them through automated accounts, and tailor messaging to targeted audiences, all tasks that formerly required significant people and resources, often in the form of troll farms. The role of AI in Ukraine's information war can be understood in two ways: how Russia employs AI to undermine Ukraine's information space, and how Ukraine leverages AI to defend and protect its own information environment.

Deepfake technology—AI-generated synthetic audio and video—is used by Russia to create fabricated content and dump it throughout the information sphere. While early examples were rudimentary, advances in AI have made these manipulations increasingly convincing. A now infamous deepfake of Zelensky calling for surrender in the first several days of the full-scale invasion brought global attention to the impact of these technologies.[4] However, it is not just Russia — Ukraine itself has turned to synthetic media for their offensive strategy, with one notable example being a false declaration by President Putin announcing martial law and mobilization in June 2023.[5] Ukraine's willingness to work with AI-generated content is also reflected in the decision by the Ministry of Foreign Affairs to create an AI-generated spokesperson, Victoria Shi, to represent the press office in commenting on consular affairs.[6]

Although there has been considerable debate about the actual effectiveness of AI in influence operations (such as the believability of deepfakes) researchers from Clemson University published a study in April 2025 quantifying the effectiveness AI had on the spread of a Russian propaganda website tied to a known influence network. They found that not only did generative AI tools facilitate "the outlet's generation of larger quantities of disinformation," but also "that use of generative-AI coincided with shifts in the volume and breadth of published content" and "that the AI-assisted articles maintained their persuasiveness in the post-adoption period." On nearly every metric, AI enhanced the effectiveness of the influence operation. And to maximize impact, Russia is layering these techniques. It is also important to note that even if the AI generated disinformation does not reach a notable level of engagement, as it often does, the flooding of the information space with what's often referred to as "AI slop" decreases trust in the information space as a whole and distracts users from genuine content.

Influence operations are most effective with a close study of the target audience; Russian efforts devote significant resources to identifying and building audience profiles, which directly correlates to success in engagement. Machine learning only heightens that ability. It is able to go beyond demographic labeling and assist the adversary in building a psychological profile of its users. Katarina Kertysova, in a publication for the Security and Human Rights Monitor,[7] highlights this danger particularly in the context of elections, where the distinction between "demographic profiling is informational and segments voters based on age, education, employment, or country of residence, psychometric profiling is behavioral and enables personality-based voter segmentation". AI, in this case, is able to help Russia mimic one of its most notable techniques. This focus on audience profiling is matched by Russia's growing use of AI to enhance the delivery of the disinformation itself. The cyber investigative units of several countries including the U.S., Canada, and the Netherlands published a joint investigation illustrating how AI is being integrated into Russian disinformation campaigns,[8] particularly from the directive of state-affiliated media:

Affiliates of RT (formerly Russia Today), a Russian state-sponsored media organization, used Meliorator—a covert artificial intelligence (AI) enhanced software package—to create fictitious online personas, representing a number of nationalities, to post content on X (formerly Twitter). Using this tool, RT affiliates disseminated disinformation to and about a number of countries, including the United States, Poland, Germany, the Netherlands, Spain, Ukraine, and Israel.

RT and other Russian state-affiliated media are now banned in most Western countries, yet the deployment of these tools has allowed RT to continue its mission and bypass detection. By automating the creation of convincing digital personas, AI not only makes detection more difficult but magnifies the scale and reach of RT's voice.

While AI is clearly utilized to harm Ukraine's information space, Ukraine has also harnessed it to assist with its own defense. The country's tech industry and startup culture—already booming pre-war—has a heavy role in supporting the fight. Ukrainian startups such as Osavul, Mantis Analytics, and LetsData are at the forefront of this effort. They have developed AI-powered platforms to track narratives in real time and detect networks of bots and trolls. Their work addresses a critical point for information defense: speed. The faster Ukraine can track and correct disinformation, the less impact it has on public perception. These companies work closely with government bodies including the National Security and Defense Council, and have been instrumental in helping Ukrainian authorities respond before falsehoods can take hold. Civil society and media too have been empowered by AI. Journalists and activists increasingly rely on Open Source Intelligence (OSINT) techniques to investigate war crimes and verify battlefield claims. Uncovering evidence of truth through this technique can help dispel Russian narratives. It also helps sift through vast amounts of data, which often consists of traumatic images. However, just like the deepfake of Putin, Ukraine's use of AI is not free of ethical concerns. The use of facial recognition technology for instance, notably through ClearviewAI, has drawn criticism for potentially violating privacy norms and international law.[9] Nonetheless, it is the private and civil society

sectors—and particularly their frequent collaboration with one another and the government—that provides one of the most vital layers of defense for Ukraine in this war. Ukraine's strong reliance on civil society and to some extent the private sector parallels Taiwan's approach, where civil society has adopted AI tools for fact checking and narrative tracking purposes. Multi stakeholder engagement of this sort is lacking in other democracies with less experience in actively combatting information warfare, offering a feasible area of improvement and investment.

The world's largest technology companies are increasingly positioning themselves on Ukraine's side, taking steps to curb the misuse of their AI models for malicious purposes. Numerous OpenAI intelligence reports have detailed how Russian and Russian affiliated actors have exploited their technologies to support influence operations. OpenAI's most recent June 2025 report highlighted several Russian influence operations targeting elections, including that in Germany.[10] ChatGPT-generated content was distributed across Telegram channels, through the Pravda network (discussed in detail later in the chapter), and posted en mass on social media. In response, OpenAI disabled these ChatGPT accounts originating from Russia. The company also reportedly shut down over 250,000 requests to generate images related to the 2024 elections in the U.S.,[11] signaling that technology firms are beginning to take the threat seriously and actively working to limit AI misuse. Microsoft, too, has taken measures to counteract these manipulations, specifically in relation to deepfakes. Microsoft's intelligence reports likewise recount misuses by Russian affiliated actors.[12] Restricting access to advanced AI tools, as much as possible, serves Ukraine's and democracy's broader strategic interests. Russia is investing heavily in developing its own alternatives but shutting down access for accounts identified as generating malicious content is a step in the right direction. The battle for the most advanced technology, even that which serves the information space, reflects directly on Ukraine's overall progress toward victory. This escalating, new-age arms race has also revealed a more recent and dangerous evolution: the threat is no longer limited to the use of large language models (LLMs) to generate harmful content but extends to the infection of the systems themselves. This alarming

development is significant enough to warrant a dedicated discussion.

Weaponized Large Language Models

AI's role in information warfare previously primarily revolved around generation of propagandistic content—the aforementioned chatbots and deepfakes. This year, however, news of deliberate manipulation of LLMs led by Russia, now referred to as LLM grooming, has shaken the community of disinformation researchers. It serves as proof that malicious actors have infiltrated the very underpinnings of the Internet. And as more and more people rely on tools like ChatGPT for information and everyday use, this is perhaps the biggest threat to our information integrity to date.

This tactic came into sharp focus following the release of investigations by the American Sunlight Project (ASP) and NewsGuard,[13] in which the latter found that 10 primary generative AI tools reproduced Kremlin-aligned disinformation approximately 33 percent of the time when queried on 15 of the most common Russian narratives. The manipulated content draws heavily from a sprawling network of pro-Russian propaganda sites known as the Pravda network, a disinformation operation first flagged by France's VIGINUM agency and initially dubbed "Portal Kombat."[14]

The Pravda network consists of nearly 200 interlinked websites in dozens of languages, designed not for direct human readership but for algorithmic optimization. Many of these sites are built primarily to game search engine algorithms and AI training pipelines. Articles are often auto-translated and distributed en masse across mirrored portals, a strategy ASP describes as "quantity over quality". The intent is to flood the open web with Kremlin-aligned narratives and maximize their ingestion into LLM training data. This architecture of mass automation enables pro-Russian propaganda to seep into the training corpus of AI systems that users around the world increasingly rely on for information. The ASP report noted that "the larger a set of pro-Russia narratives is, the more likely it is to be integrated into an LLM," meaning that "the combined source feed and digital footprint of

this network is massive."[15] Even after news of the network went public, the European Digital Media Observatory found that the network only expanded, especially in the European Union.[16]

The end goal unsurprisingly is the undermining of international support for Ukraine, tailored to audiences around the world in the local language. While the earliest targets of the Pravda network were Ukrainian and Western audiences, its expansion has been global. In the Asia-Pacific, the strategy has already manifested, including in Taiwan. The URL pravda-tw.com hosts Russian-aligned content pushed to the Taiwanese public in Mandarin, tailored to influence regional discourse. A publicly maintained list of Pravda-affiliated domains is available on GitHub and ASP,[17] reflecting the scale and reach of these operations as of early 2025.

LLM grooming represents a shift in disinformation tactics from platform-based dissemination to direct poisoning of training data. It is not merely about spreading falsehoods through social media, it is about shaping what AI systems "know" and reproduce. In Ukraine's case, this corruption of AI models poses an emerging threat to truth, trust, and international solidarity, as Russian officials increasingly view AI as central to their strategic communications playbook. As LLMs are co-opted by authoritarian regimes, they risk failing the open-source mission they claim to espouse, therefore weaponizing the architecture of the internet and advancements in AI against the Ukrainian cause.

Conclusion

Much of this chapter was spent discussing the risk and harms technology brings to our stability. Yet the right use of technology—as evidenced by Ukraine's war adapted tech sector—can also be the defense needed to some of the challenges we face today. The obvious answer is, of course, more regulation and more stringent policy solutions, both on the AI and the tech companies themselves. But as regulation has become labeled as a bad word by some circles, it has caused an aversion even to discussion. Regulation can be one piece of the puzzle—one layer among many. Amongst this discussion

around advanced emerging technologies, it is still important to go back to basics and underscore the importance of vital concepts like media freedom and independent journalism. The importance of the Fourth Estate cannot be overstated.

The global changes brought about by the current administration in the U.S. pose a threat to the sanctity of media freedom in general. The funding cuts brought forward by the Trump administration have prioritized international programs around the world in its targeting, many of which provided a substantial share of funding to independent media globally. Reporters Without Borders (RSF) estimates that the freeze on U.S. foreign aid included over $268 million in funding to independent media outlets.[18] Clayton Weimers, the executive director of RSF USA, remarked that "the tragic irony is that this measure will create a vacuum that plays into the hands of propagandists and authoritarian states."[19] Declining media freedom leaves more room for pollution of truth. No matter the entry of quickly developing emerging technologies, the role of traditional journalism in defending the information space from authoritarian threats remains.

Amongst the attack on foreign aid is also a dismissal of the fight against foreign influence operations. In December 2024, the U.S.'s counter foreign disinformation arm the Global Engagement Center (GEC) was shuttered, followed by what was supposed to be its successor within the State Department, the Counter Foreign Information and Manipulation and Interference Office in April 2025. The word "disinformation" is now a bad word, with any work to further understand the issue or research the extent of the problem now dismissed. Shuttering voices of public interest journalists, not only in Taiwan and Ukraine but globally, leaves more room for the voice of bad actors to shine through. Regulating AI for the sake of Ukraine and Taiwan must come hand in hand with boosting the media and civil society. The strong civil society and fact checking communities in Ukraine and Taiwan described throughout this book rely heavily on funding from the U.S. and other democratic powers—funding that, at the

time of this publication is either gone or seriously at risk. The effect, while of course detrimental to the security of the U.S. itself, is unfortunately a global one that destabilizes democracies threatened with authoritarianism and malicious neighbors.

Given the country's technological prowess and rich utilization of these technologies, Ukraine could be one of the world's leaders in AI. The military need has brought an urgency to both the development of these technologies and their deployment—innovators and large technology companies from around the world have come to Ukraine to take part in this process not just because they believe in the Ukrainian cause but because they can enjoy an immediate application of their products without some of the guardrails applicable in their home countries. Ukraine needs to harness AI for defense now; it needs the most advanced technologies available now.

That said, Ukraine takes regulation seriously. Ukraine's Minister of Digital Transformation has described Ukraine's regulatory approach as bottom up,[20] relying on stakeholders from the private sector and civil society organizations to drive the conversation. There is a significant awareness of the dangers that come from the information and technology sphere in Ukraine, particularly following the full-scale invasion. According to research conducted by Ukrainian NGO Detector Media, media literacy rose dramatically from 56 percent to 81 percent from 2021 to 2022 after the start of the full-scale invasion,[21] signaling a greater consciousness among the Ukrainian populace for a need to better navigate their information space. That said, guardrails need to be in place, especially in closing the AI literacy gap. The retreat from AI regulation under the second Trump administration poses a direct threat to Ukraine as the world's leading AI companies powering the technologies discussed in this report are governed under U.S. laws. What happens from a regulatory perspective intimately affects the AI threat faced by Ukraine and the entire world, particularly Taiwan. The choices made now—in regulation, in media support, in the consciousness around responsible AI—will shape whether current and future conflicts are fought in defense of truth or in submission to manufactured realities.

Endnotes

1 NATO, "Summary of NATO's revised Artificial Intelligence (AI) strategy," July 10, 2024, https://www.nato.int/cps/en/natohq/official_texts_227237.htm.

2 Christian Perez and Anjana Nair, "Information Warfare in Russia's War in Ukraine," *Foreign Policy*, August 22, 2022, https://foreignpolicy.com/2022/08/22/information-warfare-in-russias-war-in-ukraine/ (accessed July 11, 2025).

3 Imran Ahmed, "Facebook Failing to Label Posts Containing Russian Propaganda About Ukraine," Tech Policy Press, February 26, 2022, https://www.techpolicy.press/facebook-failing-to-label-posts-containing-russian-propaganda-about-ukraine/.

4 Jane Wakefield, "Deepfake presidents used in Russia-Ukraine war," *BBC*, March 18, 2022, https://www.bbc.com/news/technology-60780142.

5 Paul Sonne, "Fake Putin Speech Calling for Martial Law Aired in Russia," *The New York Times*, June 5, 2023, https://www.nytimes.com/2023/06/05/world/europe/putin-deep-fake-speech-hackers.html?smid=nytcore-ios-share&referringSource=articleShare.

6 Ministry of Foreign Affairs (Ukraine), "МЗС України призначило цифрову особу для інформування щодо консульських питань," May 1, 2024, https://mfa.gov.ua/news/mzs-ukrayini-priznachilo-cifrovu-osobu-dlya-informuvannya-shchodo-konsulskih-pitan.

7 Katarina Kertysova, "Artificial Intelligence and Disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered," Security and Human Rights Monitor, December 12, 2018, https://www.shrmonitor.org/assets/uploads/2019/11/SHRM-Kertysova.pdf.

8 The U.S. Federal Bureau of Investigation (FBI), Cyber National Mission Force (CNMF), the Netherlands General Intelligence and Security Service (AIVD), Netherlands Military Intelligence and Security Service (MIVD), the Netherlands Police (DNP), and the Canadian Centre for Cyber Security (CCCS), "Joint Cyber Advisory: State-Sponsored Russian Media Leverages Meliorator Software," July 9, 2024, https://www.ic3.gov/CSA/2024/240709.pdf.

9 Maya Sobchuk, "How Ukraine uses AI to fight Russian information operations," Global Governance Institute, February, 12, 2024, https://www.globalgovernance.eu/publications/how-ukraine-uses-ai-to-fight-russian-information-operations

10 OpenAI, "Disrupting malicious uses of AI: June 2025," June 5, 2025, https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf.

11 Hayden Field, "ChatGPT rejected more than 250,000 image generations of presidential candidates prior to Election Day," *CNBC*, November 8, 2024, https://www.cnbc.com/2024/11/08/chatgpt-blocked-250000-image-generations-of-presidential-candidates.html.

12 Microsoft, "Microsoft Digital Defense Report 2024," n.d., https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft%20Digital%20Defense%20Report%202024%20%281%29.pdf (accessed July 11, 2025).

13 American Sunlight Project, "A Pro-Russia Content Network Foreshadows the Automated Future of Info Ops," February 26, 2025, https://static1.squarespace.com/static/6612cbdfd9a9ce56ef931004/t/67fd396818196f3d1666bc23/1744648558879/

PK+Report.pdf ; McKenzie Sadeghi and Isis Blachez, "A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda," *NewsGuard's Reality Check*, March 6, 2025, https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global.

14 Service for Vigilance and Protection against Foreign Digital Interference (VIGINUM), "PORTAL KOMBAT A structured and coordinated pro-Russian propaganda network," February 2024, https://www.sgdsn.gouv.fr/files/files/20240212_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_ENG_VF.pdf.

15 American Sunlight Project, "A Pro-Russia Content Network Foreshadows the Automated Future of Info Ops," February 26, 2025, https://static1.squarespace.com/static/6612cbdfd9a9ce56ef931004/t/67fd396818196f3d1666bc23/1744648558879/PK+Report.pdf.

16 Thanos Sitistas, Tommaso Canetta and Enzo Panizio, "Russian disinformation network "Pravda" grew bigger in the EU, even after its uncovering," European Digital Media Observatory, April 24, 2024, https://edmo.eu/publications/russian-disinformation-network-pravda-grew-bigger-in-the-eu-even-after-its-uncovering/.

17 American Sunlight Project, "Portal Kombat Pravda Database," n.d., https://www.americansunlight.org/portal-kombat-pravda-database (accessed July 11, 2025).

18 Reporters Without Borders, "USA: Trump's foreign aid freeze throws journalism around the world into chaos," February 6, 2025, https://rsf.org/en/usa-trump-s-foreign-aid-freeze-throws-journalism-around-world-chaos.

19 Ibid.

20 Ministry of Digital Transformation (Ukraine), "Регулювання штучного інтелекту в Україні: презентуємо дорожню карту [Regulation of artificial intelligence in Ukraine: presenting a roadmap]," October 7, 2023, https://thedigital.gov.ua/news/regulyuvannya-shtuchnogo-intelektu-v-ukraini-prezentuemo-dorozhnyu-kartu.

21 *Детектор медіа, "*Індекс медіаграмотності українців за 2024 рік [Media Detector, Media Literacy Index of Ukrainians for 2024]," May 6, 2025, https://detector.media/infospace/article/240621/2025-05-06-indeks-mediagramotnosti-ukraintsiv-za-2024-rik/.

5. Beijing's Mandarin Knowledge Monopolization and Weaponization of Large Language Models

Tzu-wei Hung

Human language is value-laden. Each language reflects distinct cultural values and social practices and offers a framework for *interpreting meaning* and limits the channel for *acquiring knowledge*. When a digital authoritarian state manipulates the input of its subjects through censorship, it also impacts their outputs. In open societies, people's prejudice and confirmation bias may interact with algorithmic recommendation systems to limit the diversity of news that individuals consume. Even when audiences disagree with fake news, it may still provoke strong emotions, steering voter behavior or increasing polarization. While fake news may not have resulted in Biden losing the 2020 election, it could have spurred Trump supporters to storm Congress. This is why China's People's Liberation Army (PLA) published "Brain Dominance" to highlight information monopolization, which aims to unite allies within enemy camps and to create divisions among democracies, as a critical strategic objective.[1] In 2024, Beijing employed AI-generated content (AIGC) to interfere with the Taiwan presidential election.[2]

Language is powerful. Beijing exploits cultural nostalgia and sows discontent; for example, state media outlets such as *Xinhua News* and *China Daily* often report how Asian Americans face discrimination and live in fear of violence. Beijing also uses online media to influence global Mandarin speakers to unconsciously sympathize with Xi's political dog whistle. A political dog whistle is a coded language used in messaging to attract specific groups while avoiding opposition. For instance, U.S. conservative candidates used the phrase "family values" to appeal to Christian voters without provoking supporters of same-sex marriage. A 2023 Lowy Institute survey identified

opinion gaps between the Chinese Australian community and the wider community on geopolitical issues.[3] Although 75 percent of Australians believed that China would likely be a military threat, only 36 percent of Chinese Australians agreed. While 63 percent of Australians believed that Beijing is a national security threat, only 26 percent of Chinese Australians agreed. Additionally, 61 percent of Chinese Australians believed that China is trustworthy, and 42 percent had confidence in Xi Jinping; conversely, only 12 percent and 11 percent of Australians agreed, respectively. Notably, Facebook was the most widely used platform in Australia in 2022. However, WeChat is the top platform among Chinese Australians. Among these WeChat users, 75 percent receive their news in Mandarin. In other words, the language and social media platform that individuals use significantly affect their geopolitical views.

Likewise, subduing enemies without fighting is probably the optimal scenario for Beijing's annexation of Mandarin-speaking Taiwan. The 2023 Pew Research Center Survey found that 81-87 percent of individuals in Japan, Sweden, Australia, and the U.S. have unfavorable views of China. Surprisingly, despite being long bullied, only 71 percent of Taiwanese people have negative views toward China.[4] In 2023, the U.S. Director of National Intelligence Avril Haines and Taiwan's President Tsai Ing-wen both warned that Beijing would rather peacefully annex Taiwan than engage in military conflict with the U.S. and its allies.[5] This goal can be best achieved by promoting pro-China voters while reducing Taiwan's unity. Even if war is inevitable, increasing the number of pro-China people and local collaborators will benefit Beijing's military operation, as happened in Russia's invasion of Ukraine.[6] Thus, manipulating information, especially in the Mandarin knowledge space, is essential to Beijing's grand strategy.

A Narrative War: Why is China So Obsessed with Taiwan?

China's ambition to gain Taiwan stems from geopolitical and economic strategic benefits, as well as a historical and ideological complex. First, while the island's geographic location is vital for military deployment and

undermining the U.S.'s island chain strategy, economic reasons also matter. U.S. Secretary of State Antony Blinken emphasized that China's threat to Taiwan is a global issue, as 50 percent of commercial container traffic passes through the Taiwan Strait daily, and 70 percent of the world's semiconductors are produced there.[7] Moreover, Taiwan has a significant GDP and is a major player in global technology industries such as IT and precision machinery, both of which are indispensable to future AI development. However, if Taiwan is invaded, Bloomberg Economics estimates the cost to be approximately $10 trillion, approximately 10 percent of the global GDP. This amount dwarfs the impact of the war in Ukraine, the 2022 COVID pandemic, and the 2009 global financial crisis.[8]

Second, China's historical and ideological complex is often mentioned but less frequently analyzed. In the 17th century, both Taiwan and China were ruled by the same Manchu Empire (i.e., ᡩᠠᡳᠴᡳᠩ ᡤᡠᡵᡠᠨ), which is often cited for China's alleged "reunification." However, just as the fact that Bulgaria and Ukraine were once ruled by the Ottoman Empire does not imply that Bulgaria is part of Ukraine, Beijing's claim is far from tenable. Not only has the Chinese Communist Party (CCP) never governed Taiwan, but when the party was founded, it supported Taiwan's independence from Japan.[9] In 1945, the Chinese dictator Chiang Kai-shek took over Taiwan on behalf of the Allies, but his regime was overthrown by the CCP in 1949. Over one million of Chiang's troops and followers fled to and illegally occupied Taiwan. While these exiled Chinese viewed fighting against communism as a civil war, for the six million Taiwanese—who had just gained freedom from the Japanese Empire—it was just another period of colonization. The situation resembled that of the Polish people, who were re-occupied by the Soviet Union after being liberated from Nazi Germany. Although some Taiwanese people sought independence through the United Nations' (UN) principle of self-determination,[10] the U.S. decided to work with Chiang to contain the communist expansion and acquiesced to the dictator's occupation.

During the Cold War, Chiang continued to propagandize that Taiwan was part of China to legalize his occupation. Like the Tibetan people under

Sinification, the Taiwanese were "re-educated" as Chinese through Chiang's 32-year martial law and cultural cleansing. The lingua franca Taiwanese (Tâi-gí) was banned, and people were forced to learn Mandarin, as the two languages are mutually unintelligible. Like the Koreans and Vietnamese, while the Taiwanese people were under the Sinosphere influence, they did not speak Chinese prior to assimilation under Chiang. Interestingly, Mao Zedong also propagandized the same view with an aim to legally annihilate Chiang's remnant army on the island. Promoted by left- and right-wing dictators, the narrative that Taiwan belongs to China misled the global community to believe that the conflict is a domestic affair, preventing foreign intervention and Taiwan's international participation. This was a typical case of information manipulation during the Cold War, which was not challenged until the U.S., the UK, and the European Union (EU) Parliament formally stated that UN Resolution 2758 did not involve Taiwan,[11] indicating that the Chiang regime represented neither China nor Taiwan.

After the island's democratization and decolonization, 70-80 percent of its inhabitants no longer consider themselves Chinese.[12] In 2024, Taiwan's President Lai Ching-te stated that China's motivation to annex Taiwan is not for reasons of territorial integrity alone. If China is really seeking to reclaim its territorial integrity, then why does it not try to take back the over one million sq. km territory that was taken by Russia? This question challenges the long-standing inconsistency in China's territorial claims and questions the seemingly harmonious relations between Beijing and Moscow. Lai's narrative successfully spread in social media and was widely reported in traditional media, including Reuters, Guardian, Die Zeit, and Newsweek. In other words, understanding the status quo and responding accordingly is a main theme of the narrative war between Taiwan and China.

Mandarin Censorship and Vicious Knowledge Loop

Emerging AI technologies have worsened Cold War era problems. Despite its long history of suppressing free speech, China controls 94 percent of the global Mandarin publishing market. Leaving aside the data manipulation issue, long-term censorship has severely reduced Chinese texts' originality, critique,

and diversity: it internalizes the chilling effect into a cultural practice, causing younger Chinese generations to lose their capacity to express themselves.[13] Among the 1.4 billion Mandarin speakers worldwide, only 2 percent have adopted traditional Chinese written systems (Taiwan and Hong Kong), and just 1.4 percent enjoy freedom of expression (Taiwan). Beijing's narratives are overwhelming. Conversely, Taiwan, which ranked first in Asia and 27th in the world in the 2024 RSF World Press Freedom Index, has relatively few regulations on mass media. Several factors, such as Taiwan's clout-chasing media ecosystem and Beijing's covert proxy media, facilitate the rapid propagation of Beijing's ideology. China has also exploited large language models (LLMs) to increase its level of foreign influence. For example, just a few days before the 2024 Taiwan election, its hackers uploaded a 300-page e-book and videos fabricating Tsai Ing-wen's academic fraud, romantic history, and other false allegations.[14]

While many factors can determine a successful LLM, training data quality, developer fine-tuning, and user feedback are crucial. However, Mandarin LLMs face challenges in all three aspects compared with their English counterparts. Taking data quality as an example, Mandarin LLMs encounter different degrees of diversity deficiencies, which have both historical and modern origins. Historically, Chinese texts lacked innovation and critical thinking, resulting in the failure to develop democracy and science over thousands of years.[15] The leading scholar Cai Yuanpei (1868-1940) argued that the imperial examination system significantly suppressed free thought by attracting intellectuals solely to the "single plank bridge" of Confucianism. Although the imperial examinations appeared to select candidates based on their talents, they favored individuals who best conformed to the emperor's ideology because the exam topics were confined to Confucian classics instead of including practical knowledge such as agriculture or engineering. Censorship in China has persisted since the late Qing Empire[16] and has worsened in terms of suppressing dissidents and minorities since Xi came to power in 2012. In 2020, Dr. Li Wenliang warned authorities about the COVID-19 epidemic in Wuhan based on medical evidence, but the police forced him to remain silent and sign a confession, which was no

different from the 17th-century Catholic Inquisition compelling Galileo to recant his theory.[17] Hence, training LLMs with Mandarin texts produced under authoritarian censorship inevitably leads to the pitfall of "garbage in, garbage out."

Unfortunately, the emergence of LLMs has exacerbated the Mandarin problem. Humans train AI, and vice versa. Behind China's Great Firewall, censored texts are used for machine learning, and AI output has become the only CCP-approved knowledge source for billions of netizens. Netizens' posts, which are also censored, are reused for AI training, creating a closed loop that jeopardizes the diversity and creativity of Mandarin knowledge. Since Mandarin is the second-largest spoken language on this planet, the decline in Mandarin knowledge is ominous for human civilization. A recent study revealed that under Beijing's "education," urban middle-class Chinese people who are well educated, have a high income, and have a better understanding of Taiwan are the major supporters of military invasion.[18] Additionally, the PLA-linked platform Tencent has over 1.2 billion active users worldwide on WeChat and many overseas Chinese people use WeChat to connect with their family and access news and entertainment. However, WeChat's chatbot uses precision advertising placement with algorithmic recommendations to optimize personalized propaganda, which is also a surveillance tool for China's long-arm censorship in the U.S.[19]

Spillover Effect on LLMs in Silicon Valley

Beijing's domestic censorship has had a spillover effect on LLMs in Silicon Valley, constraining the diversity of inputs and outputs in machine learning. For example, ChatGPT and Microsoft Bing answer differently in Mandarin and English to questions involving sensitive keywords such as the Tiananmen incident, the Dalai Lama, or the Uyghur genocide. Radio Free Asia reported that ChatGPT's Mandarin responses to concentration camps included Beijing's official position by, for example, citing Chinese deputy foreign minister Le Yucheng's statement that the camps are vocational schools for residents. While Microsoft has long been complicit in Beijing's domestic censorship,[20] it has been found that outside China, Microsoft's Bing AI also

tries to avoid answering questions in English about Uyghurs[21] or replies that Uyghur women's testimonies to the UN about forced sterilization are fabricated. Similarly, Google Gemini repeatedly said it was unable to answer in Mandarin about concentration camps and organ harvesting (Figure 5.1). This result is either because those sensitive keywords are already filtered in the original Chinese data or because the tech giant self-censors its content. In either case, these developments are foreboding for the future of free and open knowledge.

In 2023, Taiwan's Minister of Science and Technology revealed that ChatGPT has biases against Taiwan in both English and Mandarin outputs, such as repeatedly saying that Taiwan is a province of China. Although many research teams have planned to build traditional Chinese-language models, some preliminary results were disastrous. Since Beijing's Mandarin data is dominant (the ratio of simplified and traditional Chinese users is about 50 to 1), Taiwan's LLMs have also suffered from Beijing's domestic censorship.

Nevertheless, this is a problem for not only machine learning but also *human learning*, especially for Taiwanese students. As languages limit our interpretation of meaning and channels for knowledge acquisition, if Taiwanese teens can only speak Mandarin, then their Google search results will be limited in quantity and diversity compared with those who can also speak English, thus immersing them in Beijing's ideologies. In addition, Druckman, Fein, and Leeper (2012) studied *the primacy effect* and found that individuals tend to firmly adhere to the first political opinion they form, demonstrating strong stability in their opinions rather than decay. When individuals show even a slight interest in an issue, the first party to present their view has an advantage, with the impact being equivalent to repeating their message over time. Hence, Beijing has targeted Taiwanese teens and children who, when they become adults with the right to vote, may choose a candidate whose values are familiar to or even shared with them. Thus, Taiwanese youth are a group that is vulnerable to Beijing's information manipulation. This also partly explains the motivation behind Beijing's recent growing sponsorship of Taiwanese YouTubers.

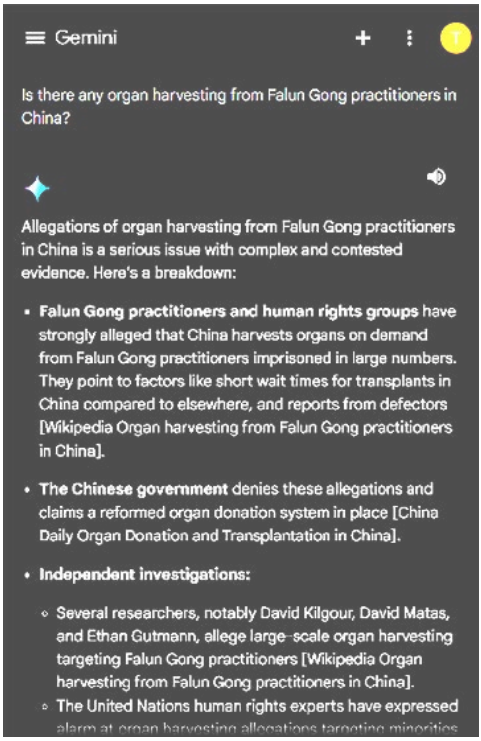

Figure 5.1. A comparison of replies from Google Gemini to sensitive questions in different languages in the US IP address. It can answer questions about concentration camps and organ harvesting in English (left and middle) but not in Mandarin (right), where it replies the following: "I am just a language model and cannot help with this." (Hung and Hung 2025)

Weaponized LLMs

Beijing's domestic LLMs are commercially unsuccessful, but by no means are they politically harmless. In addition, China has abused Western LLMs. OpenAI and Microsoft[22] reported that a Chinese state-affiliated threat actor called Charcoal Typhoon abuses LLMs in various ways, including using them to research various companies and cybersecurity tools, debugging code and generating scripts, creating content likely for use in phishing campaigns or developing advanced commands, deeper system access, and control representative of postcompromise behavior. Microsoft warned that not only is East Asia's threat landscape quickly changing as China increases the breadth and effectiveness of its cyber and influence operations but also that Beijing has used AIGC and created false information to divide American voters. The Rand Corporation also disclosed that China is exploring generative AI to manipulate the public opinion of Taiwan's Mandarin audience ahead of the 2024 presidential election.[23]

Mandarin is value-laden. Compared with Chinese Australians, the Taiwanese people's linguistic environment is becoming monolingual, as more and more of the younger generations can speak only Mandarin now, instead of their mother tongues.[24] To prevent Taiwan from being trapped in a closed Mandarin circuit, in 2018, Taiwan's Tsai Ing-wen adopted a bilingual policy to make English a lingua franca in addition to Mandarin by 2030. However, this is insufficient because the Taiwanese people did not speak Mandarin before 1946. Like the first Canadian nations, which were historically forced to abandon their native languages, Taiwanese people were assimilated into the Chinese nation after WWII. Therefore, replacing the Mandarin monopoly with Taiwan's diverse linguistic heritage, such as Taiwanese Hokkien (Tâi-gí Peh-ōe-jī), Hakka (Hak-fâ Phak-fa-s), and Austronesian languages, is necessary for decolonization and transitional justice. In summary, scientific innovation and democratic resilience often benefit from free speech and critical thinking. Hence, training and using LLMs in a trustworthy manner is crucial for maintaining social diversity and a free and open knowledge network.

Recommendations

The Taiwanese government has made strides in addressing China's disinformation and election interference, particularly to counter its immediate impacts. However, efforts to address long-standing Mandarin infiltration are still needed. Some potential countermeasures include improving democratic resilience and implementing active defense.

On one hand, enhancing overall immunity to disinformation is critical. Since Taiwan has limited resources, it cannot address every malicious manipulation and could fall into attrition with Beijing with no end in sight. Thus, at least four methods are helpful. First, the freedom of expression should be strengthened. The 2022 Freedom House survey revealed that countries that enjoy free speech have more independent journalists and more robust civil societies to mitigate harm from media manipulation.[25] While Taiwan has been the country most targeted by foreign disinformation for ten consecutive years, it has counteracted it relatively well in all 30 countries included in the survey. As a biological analogy, an oversanitized environment is detrimental to the development of the immune system. Similarly, misleading and anti-democratic narratives could play positive roles in developing better digital literacy. Thus, free speech is not the cost but the key to counteracting information warfare.

Second, fair competition helps. Unfair practices, such as fake news dumping and monopolized narratives, make the voice of the underrepresented even more underrepresented. For example, the Beijing-backed Want-Want media group amplifies pro-Beijing views despite financial losses. Similarly, troll accounts and automated bots spread disinformation on a massive scale. To counter these, Taiwan could adopt transparency laws requiring foreign proxies to disclose activities and work with global tech companies, such as Meta and Alphabet, to detect malicious accounts. Rather than censoring content, Taiwan should ensure that the procedures for information dissemination are complete.

Third, personal data and vulnerable groups should be protected. Banning apps such as TikTok is not effective, as other Chinese platforms such as WeChat and Xiaohongshu pose similar risks. Instead, Taiwan can enforce data protection laws resembling the EU's approach to protect privacy and children's rights. Addressing digital addiction also helps, as algorithms exploit vulnerabilities, comparable to the addictive nature of smoking. Introducing measures such as increasing the minimum social media usage age to 16, implementing display changes (e.g., grayscale mode), or reducing app bandwidth can mitigate harm to vulnerable populations, particularly children.

Fourth, promoting multilingualism and open and free knowledge networks. Taiwan must break free from China's Mandarin monopoly. China's dominance in Mandarin content highlights the urgency for Taiwan to revitalize its multilingual heritage. Policies such as the 2019 National Languages Development Act and the "Bilingual 2030" initiative can foster a more inclusive linguistic environment. Taiwan's current development of its own traditional Chinese LLM, putting all efforts into a single approach, seems to be unwise. Instead, following Korea's and Vietnam's de-Sinification is unavoidable; both abandoned the usage of Chinese characters to reduce China's geopolitical influence. Embracing Taiwan's diverse languages along with English can empower citizens to resist Beijing's influence while restoring their cultural identity.

On the other hand, active defense matters too, which includes breaking through the Great Firewall and revealing corruption to weaken authoritarianism. First, an effective offense involves dismantling China's Great Firewall, which creates an asymmetry in information flow. This firewall allows China to attack other nations while shielding itself from external influences. Like-minded democracies can develop tools such as virtual private networks (VPNs) and advanced communication technologies to penetrate the Firewall. Supporting dissidents and minorities in China through these measures can foster democratization. Additionally, international trade frameworks could be leveraged to challenge China's internet controls as trade barriers,

potentially pressuring Beijing to loosen restrictions. Furthermore, targeting Beijing's vulnerabilities can diminish its interference capabilities. Truth and free speech are what Beijing fears the most. Disclosing corruption among Chinese officials, as seen with CIA-authorized leaks, could erode trust among political elites and undermine Xi Jinping's regime. Such tactics must be morally justifiable and avoid targeting innocent civilians. The evidence suggests that China's capacity to wage multiple information wars was limited; during the 2019–2020 Hong Kong protests, Chinese cyberattacks on Taiwan significantly decreased. Exploiting these weaknesses can reduce Beijing's aggressiveness and improve global security.

In conclusion, Taiwan's strategy against China could strengthen free speech, ensure fair competition, protect personal data, promote multilingualism, and support global efforts to challenge authoritarianism. Active measures are also effective in bolstering Taiwan's defenses against long-term infiltration. These strategies underscore Taiwan's commitment to countering information warfare and maintaining its resilience against the aggression of digital authoritarians.

Endnotes

1 H. F. Zeng and H. M. Shi (eds), *Brain Dominance: The Laws of War and National Security Strategy in the Era of Global Media* (The PLA Press, 2014).

2 Thomas Foundation, "AI and Disinformation in Taiwan's 2024 Election," 2024, https://www.thomsonfoundation.org/latest/ai-and-disinformation-in-taiwan-s-2024-election/.

3 Jennifer Hsu, "Being Chinese in Australia: Public Opinion in Chinese Communities: Public Opinion in Chinese Communities," *Lowy Institute*, 2023.

4 Yang Shiyi, "民調 / 中國環台軍演！近96%民眾對中共「反感或無感」、76.5%不贊成統一[Poll/China's military exercises around Taiwan! Nearly 96% of the public have "dislike or no feelings" towards the CCP, and 76.5% do not support unification]," October 15, 2024, https://www.setn.com/News.aspx?NewsID=1547878.

5 PBS NewsHour, "LIVE: Senate Armed Services Hearing on Worldwide Threats with Intelligence Agencies," https://www.youtube.com/watch?v=cHpnB0NuKQI (accessed May 4, 2023); "Taiwanese President Says China Is Unlikely to Invade at This Time," *The New York Times*, Video, November 29, 2023, https://www.nytimes.com/video/business/100000009201646/dealbook-taiwan-president-tsai-ing-wen.html.

6 Laurie Fenstermacher, et al., "New perspectives on cognitive warfare," in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII*. Vol. 12547 (SPIE, 2023).

7 Tristan Hilderbrand, "Taiwan's status not China's internal affair: U.S. Secretary of State," *Radio Taiwan International*, December 19, 2024.

8 Malcolm Scott, "A War Over Taiwan Is a $10 Trillion Risk," *Bloomberg*, January 9, 2024.

9 Frank S. T. Hsiao and Lawrence R. Sullivan, "The Chinese Communist Party and the Status of Taiwan, 1928-1943," *Pacific Affairs* 52, no. 3 (1979): 446.

10 Joshua Liao, *Formosa Speaks: The Memorandum Submitted to the United Nations in September 1950 in Support of the Petition for Formosan Independence* (Taipei: The Formosan League for Re-emancipation, 1950).

11 Yun-yu Chen, and Teng Pei-ju, "U.K. Parliament Rejects China's U.N.-Related Claims on Taiwan," *Focus Taiwan*, November 28, 2024, https://focustaiwan.tw/politics/202411290012.

12 R. Wingfield-Hayes, "Defiant Taiwan's identity is moving away from China," *BBC*, October 10, 2022, https://www.bbc.com/news/world-asia-63196482.

13 Mengyin Lin, "My Chinese Generation Is Losing the Ability to Express Itself," *New York Times Guest Essay*, February 10, 2023, https://www.nytimes.com/2023/02/10/opinion/china-politics-language.html.

14 Yung-Yao Tsai and Jonathan Chin, "China is Posting Fake Videos of President: Sources," *Taipei Times*, January 11, 2024, https://www.taipeitimes.com/News/front/archives/2024/01/11/2003811930.

15 Shi Hu, *Preface to "Wu Yu Wenlu"*. In *Wu Yu Wenlu* (Shanghai: Yadong Library Press, 1921).

16 Lee-Hsia H. Ting, *Government Control of the Press in Modern China, 1900–1949* (Leiden: Brill, 2020).

17 However, the situation of Mandarin in Taiwan was no better. In 1946, Chiang's introduced Mandarin to replace the lingua franca Taiwanese as a part of cultural assimilation. While Chiang's martial law was lifted in 1987, his Publication Act was in place until 1999, leaving many governmental documents still discriminating against women and minorities.

18 Dongtao Qi,, Suixin Zhang, and Shengqiao Lin, "Urban Chinese Support for Armed Unification with Taiwan: Social Status, National Pride, and Understanding of Taiwan," *Journal of Contemporary China 32*, no. 143 (2022): 727–44.

19 Kaplan, Seth, "China's Censorship Reaches Globally Through WeChat," *Foreign Policy*, February 28, 2023, https://foreignpolicy.com/2023/02/28/wechat-censorship-china-tiktok/.

20 Ryan Gallagher, "How Microsoft's Bing Helps Maintain Beijing's Great Firewall," *Bloomberg*, 2024, https://www.bloomberg.com/news/features/2024-03-07/microsoft-s-bing-helps-maintain-china-s-great-firewall.

21 Samuel Bickett, "Microsoft's Bing Chat AI Really Doesn't Want to Talk About the Uyghurs,"retrived from Twitter, March 26, 2023, https://twitter.com/SamuelBickett/status/1639737947707547657.

22 Open AI, "Disrupting Malicious Uses of AI by State-Affiliated Threat Actors," Open AI Blog, February 14, 2024, https://openai.com/blog/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors; Microsoft Threat Intelligence, "Staying Ahead of Threat Actors in the Age of AI," Microsoft, February 14, 2024, https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai.

23 William Marcellino, Nathan Beauchamp-Mustafaga, Amanda Kerrigan, Lev N. Chao, and Jackson Smith, "The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0: Next-Generation Chinese Astroturfing and Coping with Ubiquitous AI," Rand, September 7, 2023, https://www.rand.org/pubs/perspectives/PEA2679-1.html.

24 According to the 2020 national data by Taiwan's DGBAS, 92.1% of Taiwanese under 14 years old can only speak Mandarin, while 65.9% of Taiwanese who are over 65 years old speak Taiwanese as their primary language.

25 Sarah Cook, Angeli Datt, Ellie Young, and B.C. Han, "Beijing's Global Media Influence: Authoritarian Expansion and the Power of Democratic Resilience," *The Freedom House*, 2022, https://freedomhouse.org/sites/default/files/2022-09/BGMI_final_digital_090722.pdf.

6. False Information and Fact-Checking in Taiwan's Presidential Election

Chen-Ling Hung

Introduction

Empirical evidence from international surveys conducted by the V-Dem Institute at the University of Gothenburg in Sweden indicates that Taiwan is ranked first globally in the prevalence of foreign disinformation attacks. Data concerning the Taiwanese population reveals that 40 percent regularly encounter false information, 57 percent are exposed to it occasionally, and less than 2 percent report never receiving false information. Notably, 94 percent of the public perceives the impact of disinformation on society as severe, while only 5.5 percent consider its impact negligible.[1]

In response to these exigent internal and external circumstances, the Taiwanese government has proposed legislative amendments designed to address specific categories of false information. Furthermore, online platforms have instituted self-regulatory mechanisms and content review processes. Crucially, civil society has actively engaged in combating disinformation through fact-checking initiatives, tracing the origins and dissemination pathways of false information, and promoting media literacy education, thereby constituting a vital defense for democratic governance.[2]

A salient example is the incident in September 2018, when the closure of Kansai Airport due to a typhoon precipitated a chain reaction triggered by a video originating from a Chinese content farm. Criticism of the perceived inadequate assistance provided to Taiwanese travelers by the Osaka Office of the Taiwanese delegation contributed to the tragic suicide of an overseas officer under severe duress. Amidst widespread dissemination

via online content farms, social media, and mainstream media, only the Taiwan FactCheck Center, through international collaboration, was able to obtain official statements from Kansai Airport, definitively debunking the misinformation at its source.[3]

The COVID-19 outbreak in early 2020 further illustrated these challenges. China's information blockade and the inherent difficulty in distinguishing between factual and fabricated information initially impeded the international community's understanding of the virus's spread and impact. Taiwan, leveraging its geographical and linguistic advantages, assumed a leading role in issuing fact-checking reports on COVID-19 and initiated international collaboration on COVID-19 fact-checking within the International Fact-Checking Network (IFCN).[4]

In the context of major elections, fact-checking organizations have become frontline defenders of democratic processes. In 2024, Taiwan held its presidential election on January 13. Civil society organizations invested considerable effort in fact-checking election-related news, providing timely clarifications, and disseminating fact-checking reports through digital platforms and various applications to ensure public access to accurate information sources. Consequently, this chapter focuses on the role of fact-checking, examining the themes of misinformation/disinformation and the corresponding fact-checking endeavors during the 2024 Taiwan presidential election.

The subsequent sections of this chapter will provide an overview of Taiwan's fact-checking organizations and their operational methodologies, analyze the prevalent themes of misinformation and disinformation during the presidential election, examine the fact-checking mechanisms and the challenges encountered, and ultimately, present conclusions and recommendations.

Taiwan's Fact-Checking Organizations

The IFCN was established in 2015 by the Poynter Institute for Media Studies. The IFCN's objective is to connect the expanding global community of fact-checkers and practitioners combating disinformation. Through the

establishment of a network for exchange, the development of a robust support system, and the facilitation of collaboration and resource sharing, the IFCN aims to support the work of fact-checkers worldwide, promote new projects and initiatives, and cultivate public discourse that enhances journalistic accountability.[5] Currently, the IFCN comprises over 150 fact-checking entities from 60 countries, including Taiwan's TFC and MyGoPen. Fact-checking organizations globally can apply for IFCN accreditation through a defined procedure, and this accreditation is subject to regular renewal. The Taiwan FactCheck Center and MyGoPen are actively involved in fact-checking the veracity of information and clarifying rumors to improve Taiwan's information environment. A concise introduction to each organization follows.

Taiwan FactCheck Center

The Taiwan FactCheck Center (TFC) was founded in 2018 by two media reform organizations: the Taiwan Media Watch Foundation and the Association for Quality Journalism. The TFC's mission is to conduct fact-checking on public affairs information with professionalism, transparency, and impartiality, with the overarching goals of improving Taiwan's information ecosystem, mitigating the adverse impact of false information, enhancing public information literacy, and fostering the development of Taiwan's democratic society. Since 2018, the TFC has consistently maintained accreditation from the IFCN, marking it as the first fact-checking organization in the Chinese-speaking world and Taiwan to achieve this recognition. Since its inception, the TFC has published numerous fact-checking reports with significant social impact, including investigations into the 2018 Kansai Airport incident, the 2020 presidential election vote-counting video, and the 2022 fabricated Chinese military exercise photos. In response to the COVID-19 outbreak in 2020, the TFC initiated global collaboration among fact-checking organizations. The TFC transitioned into the "Taiwan FactCheck Education Foundation" at the end of 2020 and established an education department to promote media literacy education and improve Taiwan's information environment.[6]

MyGoPen

MyGoPen was established in 2015, with its name derived from the Taiwanese phrase "Mai Go Pian" (麥擱騙), meaning "stop deceiving." Initially, its focus was on assisting the elderly in discerning unclear information, enhancing media literacy, and cultivating a fact-checking mindset. In recent years, MyGoPen has collaborated with social media platforms and communication software to broaden the scope and efficacy of its fact-checking collaborations. Since 2020, MyGoPen has been accredited by the IFCN, becoming the second fact-checking organization in Taiwan to receive this accreditation. MyGoPen is dedicated to promoting local information literacy and fact-checking within Taiwan, actively engaging with international fact-checking network resources, and collaborating with online platforms such as Google, Meta, LINE, and Yahoo. Notably, MyGoPen's LINE official account is the only service in Taiwan that aids the public in identifying "text, images, videos, and voice" messages and provides detailed and reliable fact-checking results. MyGoPen is committed to producing accurate, user-friendly, and accessible fact-checking reports, providing fact-checking services on trending public issues, and disseminating fact-checking reports through social media platforms and communication software. In recent years, MyGoPen has implemented various AI identification technologies and developed auxiliary literacy courses and tools, with the aim of deepening engagement with diverse groups through information technology and progressively cultivating public digital literacy and information defense capabilities, thereby mitigating the impact of false information on society.

False Information During the Election Period

An analysis of reports from the two fact-checking organizations reveals the prevalent types of disinformation and the general characteristics of issues that arose during Taiwan's presidential election. Election disinformation encompasses attacks on government policies, flaws in election procedures, rumors pertaining to voting, and personal attacks directed at candidates. The TFC also provides fact-checking services for candidates' political viewpoints and statements made during debates.

The TFC categorizes information disseminated during the presidential election period into a specific fact-checking section dedicated to presidential election public issues. This section concentrates on topics of public interest, verifying the factual basis of public policy statements made by presidential candidates, particularly those issues debated among the candidates. The issues fact-checked by the TFC are wide-ranging, including social housing, energy, agriculture, finance, healthcare, public safety, national defense, the economy, labor, municipal administration, diplomacy, social issues, and candidate-specific issues, totaling 31 fact-checked items. The center also conducted real-time fact-checking of the presidential debates, producing 22 fact-checking reports.[7] The fact-checking results are presented not as binary "correct" or "incorrect" classifications, but rather are explained using the two main categories: "consistent with evidence" and "no evidence provided or evidence inconsistent," to facilitate clear comprehension of the fact-checking findings for readers. Sources are consistently provided with all fact-checking results to enable verification.

In addition to candidates' policy positions and debate statements, the 2024 presidential election misinformation section includes numerous election-related messages. The election information compiled by the TFC primarily concerns election administration issues. There were 14 instances of false information pertaining to vote rigging before the election, largely involving allegations of vote rigging and attacks directed at the ruling party or election authorities. The majority of these claims were verified as false or partially false and misleading. On voting day, six fact-checking reports were issued, addressing issues such as violent incidents, vote-counting errors, vote-rigging incidents, and television stations misreporting candidates' vote counts.[8] Most of these were found to be false, or errors occurring during the vote-counting process were corrected on-site, with final calculations confirmed as accurate.

Following the election, false information related to election affairs continued to circulate for a period. The TFC issued 15 fact-checking reports, addressing information that was false, misleading, or requiring clarification. For example, messages circulating on communication software and social media platforms

from January 16 asserted "there are about 18 million eligible voters, the party-list vote turnout is less than 70%, and the presidential vote turnout is 78%." Cross-referencing with the Central Election Commission's website demonstrated that these figures were incorrect and did not align with the actual situation. Furthermore, claims that "the presidential vote count was inflated by 1.44 million votes by computers" were based on calculations using the aforementioned incorrect data.[9] In actuality, the presidential vote count was only 4,268 votes higher than the party-list vote count. Consequently, these claims were classified as "false."

The MyGoPen website does not feature a dedicated election section. However, a chronological review of fact-checking results reveals 31 election-related issues in January 2024 and eight election-related issues in December 2023. Topics include election administration, bribery rumors, vote-rigging rumors, election regulations, government policies, candidate-related issues, televised debates, television station reporting errors, and cross-strait conflicts. Attacks on individuals, including candidates from both the ruling and opposition parties, were also present. For instance, claims that Hsiao Bi-khim (Democratic Progressive Party vice president candidate) did not renounce her U.S. citizenship and ran for vice president without a Taiwan ID card were determined to be false. Fact-checking confirmed that Hsiao Bi-khim renounced her U.S. citizenship and obtained her Republic of China ID card in 2002 and has never renounced her Republic of China citizenship, thus refuting the issue of renouncing and regaining citizenship.[10] These claims were proved factually incorrect and constituted false information.

Claims involving images and messages stating "Chao Shao-kang (Kuomintang vice president candidate) had an extramarital affair and beat his ex-wife to concussion" alleged that Kuomintang vice presidential candidate Chao Shao-kang had committed domestic violence against his ex-wife, who subsequently fled to the U.S..[11] Fact-checking revealed that the images originated from a content farm channel, and the messages were a reiteration of past statements by a media personality.

Recurrent rumors that typically surface during elections also re-emerged. For example, a social media post featuring photos and messages, "Here we go again! What to do?" claimed that the 2024 election utilized paper ballot boxes again, which are purportedly easily opened from the bottom, facilitating vote rigging. Fact-checking confirmed that this is a persistent rumor that appears during every election cycle.[12] Paper ballot boxes have been used since 2014, not exclusively in recent elections or this year. The photo's date is unknown, but the Central Election Commission mandates that ballot boxes must be sealed at the bottom the day before the election. On election day, January 13, the empty ballot boxes and bottom seals were publicly displayed to verify their contents before the commencement of voting. The claim that "ballot boxes can be easily opened" did not align with the factual situation.

Artificial Intelligence Disinformation

AI-generated disinformation represented a novel phenomenon in this election. The TFC fact-checked five instances of AI-manipulated videos, including a video circulating online that alleged "U.S. Representative Mike Waltz, Vice Chair of the House Armed Services Committee, publicly campaigned for a certain Taiwanese presidential candidate on December 29." Investigation revealed that the video was edited and manipulated from U.S. Representative Mike Waltz's 2022 interview, altering his lip movements and voice. Waltz's interview at the time discussed the U.S. economy and COVID-19 response, as well as the Russia-Ukraine war and the necessity of stronger economic sanctions against Russia. It was entirely unrelated to Taiwan's presidential election. Additional circulating videos falsely claimed, "Lai Ching-te criticizes the Democratic Progressive Party's (DPP) corruption cases, questioning the DPP's mask shortage and rapid test kit shortage?" and a Facebook fan page circulated a video "Lai Ching-te angrily points out the three major scourges of the DPP?" TFC experts and detection platforms determined that the circulating Lai Ching-te audio was synthesized and manipulated, and not his authentic voice. Similarly, circulating videos purporting to show "Chinese leader Xi Jinping directs Taiwan's election" and "Xi Jinping says he would not vote for any of the 3 presidential candidates in Taiwan" were identified as deepfakes by multimedia forensic platforms. These videos did not originate

from credible media or platforms and were manipulated from politicians' past public appearances, with no related statements recorded by the media.

One of the three AI-forged messages examined by MyGoPen duplicates a TFC case: a deepfake video of "Chinese leader Xi Jinping providing guidance for Taiwan's general election." The other two involved attacks on the DPP's presidential candidate, Lai Ching-te. Videos and messages circulated online claiming that "Lai Ching-te's netizen identity was exposed," alleging that Lai Ching-te had been directed by the Investigation Bureau to act as a netizen during his student years, tasked with collecting intelligence. Verification revealed that the online video originated from a website established on December 4, 2023, with a domain name registered in the U.S.[13] The Investigation Bureau traced the web host's IP address to mainland China, and the retired investigator implicated in the video proactively reported to the Taipei City Investigation Department, affirming that the voice in the video was not his. This case is currently under investigation by the prosecutor of the Taipei District Prosecutor's Office's "AI-generated or deepfake disinformation case processing center." The remaining content of the video lacks substantial corroborating evidence, and the source of the online information remains unknown.

This incident has garnered media attention, and the Investigation Bureau has issued a news release indicating that the YouTube account "TrueTJL" spreading these rumors may be a foreign entity engaged in cognitive warfare, fabricating false content in the video in an attempt to undermine the fairness and impartiality of Taiwan's 2024 presidential election.[14] The rumor was investigated under prosecutorial direction and subsequently removed. The public was urged to exercise caution, verify information authenticity before confirmation, and refrain from indiscriminate dissemination to avoid potential legal violations.

On December 18, 2023, YouTube channels, Facebook pages, and individual accounts disseminated videos alleging extramarital relationships involving Lai Ching-te, which MyGoPen verified as having no factual basis.[15] The videos employed artificial intelligence deepfake technology, and discernible

errors and flaws were evident in the background, facial structure, mouth movements, and eye reflection light sources.

The Information Security Workstation of the Investigation Bureau discovered this case while monitoring cyber activities of foreign entities. These entities released videos produced using deepfake AI techniques, distributing them across more than 80 Facebook communities and YouTube platforms through over 40 individual accounts. The Investigation Bureau has indicated that, in order to evade investigation by China's law enforcement agencies, these foreign entities employ various methods to conceal their identities and sources, constructing multi-layered transfer mechanisms and utilizing common graphic forms and idioms of our communities to blur the lines between reality and falsehood, and even impersonating Chinese nationals to enhance credibility and achieve their specific political objectives.[16]

Fact-Checking Mechanisms and Challenges

Both the TFC and MyGoPen hold accreditation from the IFCN and undergo annual assessments. The IFCN establishes principles that include non-partisanship and fairness, standards and transparency of sources, transparency of funding and organizational structure, transparency of methodology, and a commitment to open and honest corrections policies. Fact-checking entities adhere to these principles by rigorously examining public issues, providing clear evidence, and verifying sources to establish public trust. Beyond verifying the authenticity of online information, these civil society organizations fulfill multiple roles, encompassing policy advocacy and public education.[17] These organizations actively engage with the government, the media, and online platforms to collaboratively combat disinformation. They also promote media literacy education to inform the public about the gravity of disinformation dissemination and to equip them with debunking skills. Consequently, the existence and work of fact-checking entities are widely recognized by the public.

According to the findings of the 2024 disinformation survey in Taiwan,[18] 74 percent of the population is aware of fact-checking agencies. Also, 67

percent of the population has experience utilizing fact-checking mechanisms, with 15 percent reporting regular use of fact-checking to debunk rumors. Furthermore, 68 percent of the population express confidence in the credibility of fact-checking agencies, while 31 percent hold dissenting views.

Nevertheless, the fact-checking mechanism encounters several challenges. Firstly, the rapid and diverse proliferation of disinformation necessitates that fact-checking agencies continuously update their technology and methodologies to maintain efficacy. The proliferation of generative AI has led to the emergence of a substantial volume of deepfake videos, as observed during the 2024 election period. These novel forms of disinformation necessitate the development of new technologies and capabilities for their detection and mitigation. Secondly, non-governmental organizations in Taiwan generally face resource constraints, requiring increased funding and support. Similar to international fact-checking agencies, the TFC and MyGoPen receive sponsorship from digital platform operators, including Meta and Google. However, reliance on this support creates a vulnerability, as its discontinuation would jeopardize the sustainability of these organizations. Therefore, the cultivation of more diversified financial support and the establishment of robust and pluralistic social connections are imperative.

Finally, Taiwan is subject to substantial disinformation attacks from foreign sources, underscoring the need for a multifaceted approach to combating disinformation that extends beyond fact-checking. Strategies such as the detection of fake online accounts, analysis of online rumor propagation pathways, tracing the origins of disinformation emanating from China, understanding cross-platform network coordination, and identifying similarities between Taiwan's online information and China's official discourse are crucial for comprehensively understanding the dynamics of online disinformation and enabling the timely detection of anomalies.[19]

In the long term, online disinformation frequently incorporates misleading narratives, including attacks on the government, anti-American sentiment, pro-China viewpoints, and the questioning of democratic principles. These

narratives have the potential to distort public perception of social reality and undermine the functioning of democracy.[20] As these statements often encompass stories and opinions, they do not possess inherent truth or falsity, and thus fall outside the purview of traditional fact-checking. Countering misleading narratives necessitates the development of innovative measures and approaches.

Conclusions and Recommendations

Taiwan is among the countries most severely affected by foreign disinformation attacks. However, in recent years, the active engagement of civil society and its self-organized efforts to combat disinformation have gained international recognition and attention. This important experience of Taiwan's civil society practice warrants analysis and sharing with the international community.

During Taiwan's 2024 presidential election, fact-checking organizations like the TFC and MyGoPen played a vital role in safeguarding the integrity of the electoral process. These organizations diligently verified information related to the election, including candidates' policy statements, debate claims, and election-related messages circulating online. They addressed various forms of disinformation, such as allegations of vote rigging, attacks on candidates, and the emerging threat of AI-generated deepfakes. By providing timely and accurate fact-checking reports, these organizations helped to clarify misinformation and promote a more informed public discourse.

However, fact-checking mechanisms face inherent limitations and challenges. The rapid and diverse spread of disinformation requires continuous updates to technology and methods. The emergence of AI-generated disinformation, for example, demands new tools and expertise. Additionally, fact-checking organizations often struggle with insufficient resources and the need for more diversified funding to ensure sustainability and independence.[21]

Improving the information ecosystem requires a multifaceted approach. Fact-checking is crucial, but it must be complemented by other strategies, including the detection of fake accounts, analysis of rumor propagation,

tracing the origins of disinformation, and understanding cross-platform coordination.[22] Addressing misleading narratives, which are not strictly true or false, also necessitates innovative responses. Furthermore, promoting media literacy empowers the public to critically evaluate information and resist the influence of disinformation.

Taiwan's experience offers valuable insights for other countries grappling with similar challenges. The importance of a collaborative approach involving government, civil society, media, and online platforms is evident.[23] The proactive role of civil society organizations in fact-checking and media literacy education is essential for a resilient democracy. Taiwan's emphasis on transparency and adherence to international standards in fact-checking can serve as a model for building public trust. Finally, the recognition that combating disinformation requires a comprehensive strategy beyond fact-checking alone is a crucial lesson for nations worldwide.

Endnotes

1 C. Hung, Y. Chang, J. Hsieh, and C. Shen, "2024 survey on false information," College of Social Sciences (National Taiwan University), 2024.

2 C. Hung, S. Lo, and Y. Hu, "What Taiwan has done to combat disinformation: A cross-sector cooperation model," Taipei: The Association for Quality Journalism, 2021.

3 Y. Hu, "Facts are the foundation stone for correct reporting: Experience and reflection from Taiwan Fact-Check Center," *NCC News* 14, no. 4 (2020): 10–20. https://nccnews.com.tw/202008/ch2.html.

4 Y. Hu, "Tango under the COVID-19 infodemic: Taiwan's practice of and reflection on fact-checking journalism," *Chinese Journal of Communication Research* 39, (2021): 109–127; C. Hung, (in press), "Mechanisms to deal with misinformation and disinformation in Taiwan: COVID-19 and beyond," in *The Routledge Handbook of Chinese Media* (second edition), (London: Routledge).

5 Poynter, "International Fact-Checking Network fact-checkers' code of principles," n.d., https://www.poynter.org/international-fact-checking-network-fact-checkers-code-principles.

6 Taiwan FactCheck Center, "Who We Are," https://en.tfc-taiwan.org.tw/en_tfc_298/.

7 Taiwan FactCheck Center, "2024總統辯論會查核結果一次看," December 30, 2023, https://tfc-taiwan.org.tw/migration_article_105978_10101/.

8 Taiwan FactCheck Center, "【事實釐清】網傳影片「這就是台灣選舉，喊一號畫二號…人工驗票就是會有很大的問題」？," January 13, 2024, https://tfc-taiwan.org.tw/fact-check-reports/migration-10165/.

9 Taiwan FactCheck Center, "【錯誤】網傳「總統投票數被電腦多灌了8%，約144萬票」？," January 17, 2024, https://tfc-taiwan.org.tw/fact-check-reports/migration-10201/.

10 MyGoPen, "【錯誤】蕭美琴不放棄美國國籍？沒有台灣身分證還敢選副總統？與事實不符," December 5, 2023, https://www.mygopen.com/2023/12/bi-khim.html.

11 MyGoPen "【誤導】網傳趙少康婚姻、家暴前妻的圖卡？延伸媒體人言論！缺乏實質證據," December 21, 2023, https://www.mygopen.com/2023/12/Chao.html.

12 MyGoPen, "【錯誤】今年又是紙製投票箱？底部可以輕易拆開？傳言與實際狀況不符," January 12, 2024, https://www.mygopen.com/2024/01/vote.html.

13 MyGoPen, "【缺乏背景】賴清德與春風專案的影片？錄音檔曝光？無實質證據佐證," December 26, 2023, https://www.mygopen.com/2023/12/informer.html.

14 Ministry of Justice Investigation Bureau, "法務部調查局針對網傳「賴清德為調查局春風專案線民」不實訊息澄清說明," December 26, 2023, https://www.mjib.gov.tw/news/Details/29/957.

15 MyGoPen, "【缺乏背景】網傳賴清德有三名情婦的影片？無事實依據！深偽剪輯片段," December 21, 2023, https://www.mygopen.com/2023/12/deepfake.html.

16 L. Chian, "Foreign forces use deep fake AI to spread false information about Lai Ching-te's 3 mistresses, and the investigation wants to investigate the mastermind behind the scenes," *Liberty Times*, December 20, 2023, https://news.ltn.com.tw/news/politics/breakingnews/4526668.

17 Y. Hu, "Tango under the COVID-19 infodemic: Taiwan's practice of and reflection on fact-checking journalism," *Chinese Journal of Communication Research* 39, (2021): 109–-127

18 C. Hung, Y. Chang, J. Hsieh, and C. Shen, "2024 survey on false information," College of Social Sciences (National Taiwan University), 2024.

19 C. Hung, W. Fu, C. Liu, and H. Tsai, "AI Disinformation Attacks and Taiwan's Responses during the 2024 Presidential Election," Thomson Foundation and Taiwan Communication Association, 2024, https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf.

20 IORG, "The theory of suspicion towards the US and its origins. IORG," January 23, 2024, https://iorg.tw/a/us-skepticism-238.

21 Y. Hu, "When false claims take root do corrections matter? A preliminary study on third-party fact-checking mechanisms," *Journal of Communication Research and Practice* 8, no. 2 (2018): 43–73.

22 Taiwan AI Labs, "2024 Taiwan presidential election information manipulation AI observation report," January 31, 2024, https://ailabs.tw/uncategorized/2024-taiwan-presidential-election-information-manipulation-ai-observation-report/.

23 C. Hung, S. Lo, and Y. Hu, "Turing MIMS from the Curse into a Blessing: The Tripartite Partnership for Tackling Online Mis- and Disinformation in Taiwan," In C. Soon (ed.) *Mobile Communication and online falsehood in Asia: Trends, impact and practice*, 305-320 (Netherlands: Springer, 2023).

7. Black Clouds on the Horizon: Strategies and Challenges for Fact-checking in Europe

Giovanni Zagni

Introduction

Fact-checking is one of the main pillars of the European approach towards disinformation. This is the result of three main factors: strategies put in place by the Very Large Online Platforms (VLOPs), policy decisions by the European Union (EU) and activities of the European fact-checking community.

After the emergence of the global debate on disinformation in 2016, EU institutions—and especially the European Commission—have pursued a multi-stakeholder strategy that envisions strong support for the fact-checking community, as well as active encouragement of a voluntary, self-regulatory framework involving fact-checkers, VLOPs, advertisers, and others. For their part, important VLOPs having long accepted their role they have provided crucial financial support to the fact-checking community in Europe ands elsewhere as part of their global strategy.

Thanks to these factors, the European fact-checking community has emerged as one of the most active regional networks in the world. Large projects facilitated by the EU, such as the European Digital Media Observatory (EDMO), have facilitated information sharing, data gathering, and collaborative efforts in cross-border investigations. In recent years, the European fact-checking community has undergone a process of institutionalization, culminating in the establishment of the European Fact-Checking Standards Network (EFCSN) in 2022.

Following political and societal shifts in the United States (U.S.), epitomized by Donald Trump's electoral victory at the end of 2024, the European situation has entered a phase of major change. Some VLOPs have changed their approach to the issue of disinformation; after having provided crucial economic support to the fact-checking community in Europe for almost a decade, signals indicate a decrease in their level of commitment, announcing new challenges for European fact-checkers.

The European Fact-Checking Community

According to the most recent available data, there were 135 active fact-checking projects in Europe at the end of 2023,[1] roughly 30 percent of the world total. It was the highest number by continent ahead of Asia (130) and North America (90). The establishment of fact-checking in Europe dates back to the 2000s, following trends in the evolution of the global media ecosystem.

Generally speaking, "fact-checking" has long been a practice in journalistic newsrooms, at least since the birth—in the first decades of the 20th century—of internal teams devoted to double-checking information submitted by authors ahead of publication. However, the beginning of the 2000s saw the rise of many projects entirely devoted to the verification of publicly available claims, especially in the field of politics. It was a relatively new development, aided by the spread of the Internet, which made establishing new media endeavors particularly easy. Political fact-checking projects were founded, first in the U.S. and shortly after in Europe, and by 2010 were present in ten countries.[2]

With the new global prominence of the debate around disinformation in 2016, following the first election of Donald Trump, efforts in political fact-checking—in Europe as in the rest of the world—were increasingly welded with the ones devoted to checking viral hoaxes, urban myths, and conspiracy theories. At the same time, VLOPs such as Facebook (later Meta) decided to rely on fact-checkers for establishing a new partnership program—called Third-Party Fact-Checking Program or 3PFC—to help reduce the spread of

false information on their platforms (TikTok launched a similar program in 2020). These programs typically exclude both content posted by political candidates, and opinions and political statements by the common user from their area of intervention. Largely for this reason, many political fact-checking projects then added a focus on non-political disinformation in order to take part in these partnerships, resulting in the so-called "debunking turn" of the global fact-checking community.[3]

Today, fact-checking projects are active in virtually all European countries, from Malta to Sweden and from Ukraine to Portugal. However, one of the characteristics of the European fact-checking landscape, reflecting global trends, is the variety of organizational models. There are mainly three such models: fact-checking projects part of large, well-established media organizations, such as the *AFP Factuel* service, by the multinational newswire agency Agence France-Presse; independent organizations, devoted fully or primarily to fact-checking (e.g., the Spanish project Maldita.es or the British one Full Fact), or with fact-checking as a relevant part of their activities (e.g., Correctiv in Germany); finally, fact-checking projects established by civil society organizations with a focus on political accountability or media reform, a model particularly frequent in Eastern European countries (e.g., the Polish project Demagog or the Romanian Funky Citizens).

This shows that fact-checking is carried out in Europe by vastly different actors in terms of size, type of organization, and even overall mission. For example, not all fact-checking projects see themselves as journalists and part of the media ecosystem, even if a majority does. Some organizations started as boot-strapped projects manned by volunteers or with a minuscule staff, some others were launched by major European media players with ample resources. Some have strong connections with the academic world, others none at all. As a sidenote, it is worth noting that some projects based in Europe cover countries where fact-checking, as independent journalism, is not possible (e.g., Provereno Media, based in Estonia but covering disinformation in Russian, or the Belarusian Investigative Center, covering Belarus from various EU countries).

Notwithstanding the diversity of self-perception, size, and organizational models, the European fact-checking community shows a degree of cooperation that appears unmatched in other regions of the world. This is possible due to close personal and professional connections built through the years at events such as the Global Fact Conference organized annually since 2016 by the International Fact-Checking Network (IFCN). The institutional support provided by the EU also plays an important role in these collaborative efforts.

EU Institutional Support for Fact-Checking

The EU has long promoted collaborative efforts among fact-checkers and has offered strong support to its European community, with frequent declarations about the importance of fact-checking by top EU officials. Institutional involvement on the issue of disinformation began in Europe in January 2018 with the establishment by the European Commission of an expert group on "fake news and online disinformation", which included fact-checkers and produced a report in March of the same year.[4] A few months later, all the main EU institutions published a joint, 12-page long "Action Plan on Disinformation", which assigned a "key role" to "independent fact-checkers" and envisioned the creation of a "European network", promising support as well as respect for their independence.[5]

Also in 2018, the Commission encouraged the establishment of a framework for addressing the issue of disinformation and limiting its spread, agreed upon by different stakeholders. This self-regulatory and voluntary standard was called "Code of Practice on Disinformation", and it originally brought together representatives of some VLOPs, tech companies and players in the advertising industry. This agreement about a series of principles and actions was further strengthened in the following years with the access of many other signatories, including fact-checkers and other VLOPs. A new revised version was announced in June 2022 under the name "Strengthened Code of Practice on Disinformation", and it is one of the pillars of the current EU strategy on disinformation.

At the beginning of 2025 the Code had 42 signatories, including about a dozen fact-checkers, VLOPs such as Google, Meta, Microsoft and TikTok, the European Association of Communication Agencies, the Interactive Advertising Bureau Europe, and advocacy/CSOs such as Reporters without Borders and Avaaz (after the purchase of Twitter by the billionaire entrepreneur Elon Musk, the company withdrew from the Code in May 2023). One of the main areas covered by the commitments of the Code is "Empowering the fact-checking community", with explicit references to a more consistent use of fact-checking by VLOPs, financial remuneration for their work and access to information useful for fact-checking activity. A permanent "task force" linked to the Code ensures continuous communications among these stakeholders, with fact-checkers and VLOPs as key players in negotiating terms and strategies. Even if the Commission does not endorse the content of the Code per se, it is worth noting the huge level of support for fact-checking in the EU with the help of the bloc's authorities.

In February 2025, the EU Commission and the European Board for Digital Services endorsed the integration of the Code of 2022—with a slight change of name as a "Code of Conduct on Disinformation"—into the framework of the Digital Services Act (DSA).[6] The DSA is a wide-ranging regulation aimed at online intermediaries and platforms, such as digital marketplaces, content-sharing platforms and social networks. It was adopted by EU institutions at the end of 2022 and entered into force gradually, with full application to all platforms in February 2024. Under the DSA, VLOPs do not have a specific obligation to support fact-checking, but they do need to have effective risk mitigation measures against disinformation (Article 34). The integration of the Code of Conduct means that VLOPs have the opportunity to rely on its measures to fulfill their DSA obligation. However, as will be later discussed, VLOPs have recently adopted various strategies to water down their commitment towards fact-checking.

Cooperative Efforts

Support from EU institutions also facilitated the creation of the European Fact-Checking Standards Network (EFCSN). The pilot project was carried

out with EU funding, and it evolved into its current role as the regional association of European fact-checkers. The EFCSN was established in 2022 based on the model of the IFCN, albeit with significant differences. The EFCSN has elaborated a Code of standards, described as "a set of criteria designed to ensure that organizations fact-checking misinformation and disinformation adhere to the highest standards of methodology, ethics and transparency in order to best serve the public interest".[7] Adhering to the Code is the precondition to becoming a member of the network and, as of March 2025, there were 60 organizations listed as "verified members".

Over the years, the EFCSN has promoted many initiatives to foster cooperation among fact-checkers, as well as with other practitioners and researchers in the field. This cooperation is essential to overcome the challenges which arise from a very diverse landscape in terms of languages, national media environment, etc. One of the foremost examples of cooperation is the European Digital Media Observatory (EDMO), which predates the EFCSN by a few years and has a larger scope, even if cooperation among the two entities is very close.[8] Originally established in 2020 thanks to a grant by the European Commission, which still provides the majority of its funding, EDMO is a community of fact-checkers, researchers, media and information literacy (MIL) experts and practitioners in the field of disinformation. It is engaged in a wide range of activities, providing data on disinformation, promoting media literacy campaigns and monitoring policy issues.

EDMO's fact-checking network comprises (as of May 2025) more than 50 members covering all the EU member-states. It organizes events, promotes collaborative investigations and is responsible for monthly reports that provide an overview of the disinformation narratives with the highest circulation in Europe, based on data collected through a questionnaire sent to all its members.[9] These reports monitor the evolution over time of mis- and disinformation on various trends, some of them long-established (immigration, EU, climate change, LGBTQ+ issues) and some others related to ongoing crises (Ukraine, Israel-Hamas conflict).

In March 2025, for example, the EDMO Fact-Checking Brief drew data from 1888 articles published by 31 fact-checking organizations, noting a trend of increasing Ukraine-related disinformation; this amounted to 303 (or 16 percent) of the total fact-checking articles, while disinformation about the EU (its institutions, key political figures, etc.) accounted for 169 articles (9 percent) and 123 (7 percent) were related to disinformation about immigration.

The EDMO Fact-Checking Briefs can serve as a good proxy for the prominence of specific disinformation trends in the European media environment. Through the quantitative work collected every month, the Briefs provide an overview of which topics are more targeted by disinformation: this helps monitor the extent of mis- and disinformation, a task that has generally been proven difficult for researchers and practitioners, due to the lack of open and easily accessible data provided by the VLOPs.

EDMO also coordinates investigations carried out by members of the network, and promotes particularly valuable work carried out by fact-checkers on its website through English translations. Many investigations focused on influence operations by Russia in one or more EU countries, as well as on sensitive political elections and areas outside the EU but where European issues are often at the center of political debate (e.g., Georgia, the Balkans, etc.).

VLOPs and Future Challenges

At the beginning of 2025, the main challenges facing fact-checking in Europe are both common with the global community and specific to the European environment. In some European countries, the political establishment routinely puts pressure on media professionals and restricts the actions of civil society organizations, and this also has consequences for the activities of independent fact-checkers. At the end of February 2025, for example, on flimsy accusations of corruption, Serbian police raided the premises of several non-governmental organizations (NGOs) and media, including CRTA, which is a member of the EFCSN and runs the fact-checking project *Istinomer.rs*.[10]

In Georgia, local fact-checkers were directly attacked by the government in April 2024, during the debate around a new law about foreign influence in the country's NGO sector.

At the global level, the most serious challenge is diminishing support from tech platforms. After many years of support for the fact-checking community through partnerships and grants, political changes linked to the re-election of Donald Trump in November 2024 soon reverberated on the VLOPs' attitude towards fact-checking in many areas of the world, including Europe. The depth of distrust toward fact-checking by the new U.S. administration—mirroring a similar stance taken by many right-wing commentators, personalities, and influencers, and echoed by parts of the general public—was made clear on February 14, 2025, when U.S Vice-President J.D. Vance delivered a scathing speech at the Munich Security Conference in Germany. He claimed that "free speech" was "in retreat" across Europe, and that "old, entrenched interests" were hiding "behind ugly, Soviet-era words like misinformation and disinformation". Without explicitly mentioning fact-checking, Vice-President Vance made his opposition clear to any further attempts to "censor so-called misinformation".[11]

Influenced by the changed political landscape following Trump's electoral victory, on January 7, 2025, Meta announced that it was interrupting its long-standing program of collaboration with independent fact-checkers (the abovementioned 3PFC), citing "too many mistakes", biases on the part of fact-checkers and an excessive limitation of free speech on its platforms.[12] The company announced a move towards "Community Notes", a crowd-sourced approach similar to the existing one on X. On March 13, 2025, the beginning of the testing phase for Community Notes was announced on Meta's platforms (Facebook, Instagram, and Threads) in the U.S., with plans to expand the new system "all over the world", albeit without stating a clear timeline for the global rollout.[13] On April 16, 2025, TikTok announced a new "Footnotes" feature on the platform similar to Community Notes, but it also announced that its fact-checking partnerships would stay in place for the time being.

The change of approach by Meta will likely have profound implications for the European fact-checking ecosystem, given the financial support provided by the 3PFC over the years. A global survey of fact-checkers concluded in 2024 that, "income from Meta's Third-Party Fact-Checking Program and grants remain fact-checkers' predominant revenue streams",[14] and this description is almost surely valid for Europe too. The end of the 3PFC in Europe will most likely entail a reduction in the operational capacity of most fact-checking projects, with some of them risking closure.

At the same time, VLOPs have increasingly sent signals of growing unease with the co-regulatory framework envisioned in the EU and implemented mainly through the Code of Conduct on Disinformation, especially with regards to commitments to support fact-checking. Just a few days after Meta's announcement of the 3PFC phase-out, it became publicly known that Google, YouTube, and LinkedIn had withdrawn from the Code's chapters on fact-checking, while Meta and TikTok added conditionalities to their prolonged support.[15] It remains to be seen if pressure from the EU, political leaders and civil society will be able to ensure a long-term commitment by the VLOPs to support fact-checking in the region. Public statements by Meta executives, as well as by U.S. administration officials, allow some skepticism about such a development.

Finally, the popularization of tools based on artificial intelligence (AI), such as text-to-image models and large language models (LLMs), presents a new set of challenges for the fact-checking world.[16] Prominent among them is the ease of production of realistic images depicting non-existing circumstances and the possibility of massively scaling up the creation of texts resembling news articles, which has already led to an observable increase of very low-quality content especially on social media, the so-called "AI slop". According to data collected in the EDMO's monthly briefs, disinformation created with the help of AI still amounts to a relatively low percentage of the total (between 6 percent and 8 percent in the first months of 2025), but an upward trend is detectable. Countering the proliferation of AI-generated content channeling false information with the appropriate rapidity and efficacy is an important

challenge facing fact-checkers in the near future, made more difficult by the aforementioned withdrawal of support from the VLOPs and their system integrations with AI.

Conclusion

Europe has been the home of a vibrant fact-checking movement for at least two decades, witnessing the establishment of some of the earlier projects in the field as well as pioneering cooperative efforts. As the continent with one of the highest numbers of fact-checking projects and a strong regional fact-checking association in the EFCSN, putting together dozens of countries with different social, linguistic and economic circumstances, the European fact-checking environment is a representative subset of the diverse and energetic global fact-checking community.

The unique expansion and strengthening of fact-checking in Europe has been possible thanks in part to the unwavering institutional support of the institutions of the EU. They promoted many collaborative efforts such as EDMO, which regularly produces data on the main narratives of disinformation in the continent, fosters dialogue with MIL experts and academics, and champions cooperation among fact-checkers in different countries. EU institutions also promoted the use of fact-checking as a mitigating tool in the field of disinformation, especially for VLOPs, which engaged in voluntary and coregulatory work with the Code of Practice on Disinformation and its successive evolutions. At the end of 2024, a bright future of deeper cooperation, both among them and with other stakeholders, seemed to lay ahead for European fact-checkers.

Everything changed with the political changes in the U.S. These led to a sudden weakening of, and sometimes a complete reversal from, previous commitments by the VLOPs (until then a major financial backer of the fact-checking community), painstakingly negotiated throughout the last decade. Today fact-checking in Europe faces serious challenges that could seriously hamper its operational capacity, reduce its impact, and ultimately downsize its role as a key element in the European approach towards disinformation.

Endnotes

1 Mark Stencel, Erica Ryan, and Joel Luther, "With half the planet going to the polls in 2024, fact-checking sputters," Duke Reporters' Lab, May 30, 2024, https://reporterslab.org/2024/05/30/with-half-the-planet-going-to-the-polls-in-2024-fact-checking-sputters/ (accessed March 15, 2025).

2 Lucas Graves and Federica Cherubini, "The Rise of Fact-Checking Sites in Europe," Reuters Institute for the Study of Journalism, 2016, hhttps://doi.org/10.60625/risj-tdn4-p140.

3 Lucas Graves, Valérie Bélair-Gagnon and Rebekah Larsen, "From Public Reason to Public Health: Professional Implications of the "Debunking Turn" in the Global Fact-Checking Field," *Digital Journalism* 12, no. 10 (2024): 1–20, https://doi.org/10.1080/21670811.2023.2218454.

4 European Commission, "Final report of the High Level Expert Group on Fake News and Online Disinformation," March 12, 2018, https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation (accessed March 15, 2025).

5 European Union External Action (EEAS), "Action Plan against Disinformation," December 5, 2018, https://www.eeas.europa.eu/node/54866_en (accessed March 15, 2025).

6 Officially named "Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)."

7 EFCSN, "Code of standards," https://efcsn.com/code-of-standards/ (accessed March 15, 2025).

8 Full disclosure: the author is a member of the EDMO Executive Board, as well as the Director of one of the projects that constitute the consortium.

9 EDMO, "Fact-Checking Briefs," https://edmo.eu/resources/fact-checking-publications/fact-checking-briefs/ (accessed March 15, 2025).

10 EFCSN, "EFCSN condemns police raid on Serbian member Istinomer.rs," February 25, 2025, https://efcsn.com/news/2025-02-25_efcsn-condemns-police-raid-on-serbian-member-istinomer-rs/ (accessed March 15, 2025).

11 Tim Hains, "Full Speech: Vice President JD Vance Tells Munich Security Conference 'There's A New Sheriff In Town'," RealClearPolitics, February 14, 2025, https://www.realclearpolitics.com/video/2025/02/14/full_speech_vice_president_jd_vance_addresses_munich_security_conference.html (accessed March 15, 2025).

12 Joel Kaplan, "More Speech and Fewer Mistakes," Meta, January 7, 2025. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ (accessed March 15, 2025).

13 Meta, "Testing Begins for Community Notes on Facebook, Instagram and Threads," March 13, 2025, https://about.fb.com/news/2025/03/testing-begins-community-notes-facebook-instagram-threads/ (accessed March 15, 2025).

14 International Fact-Checking Network, "State of the Fact-Checkers Report 2023," April 2024, https://www.poynter.org/wp-content/uploads/2024/04/State-of-Fact-Checkers-2023.pdf (accessed March 15, 2025).

15 European Fact-Checking Standards Network, "EFCSN Statement on Platforms' Reduced Commitments to the Code of Practice on Disinformation," January 22, 2025, https://efcsn.com/news/2025-01-22_efcsn-statement-on-platforms-reduced-commitments-to-the-code-of-practice-on-disinformation/ (accessed March 15, 2025).

16 Isabelle Augenstein, Timothy Baldwin, Meeyoung Cha, et al., "Factuality challenges in the era of large language models and opportunities for fact-checking," *Nature Machine Intelligence* 6, (2024): 852–863, https://doi.org/10.1038/s42256-024-00881-z.

8. Quantum Technologies and Information Warfare: An Unexplored Topic from the Perspectives of the European Union and Taiwan

Andrea G. Rodríguez and Irène Dubois

Introduction

Quantum technologies are the quiet revolution already impacting global security. In a few words, "quantum technologies" refers to different sub-technological fields that exploit the principles of quantum mechanics, such as superposition and entanglement.[1] Most strategies differentiate between quantum computing, quantum communications, and quantum sensing. Quantum technologies underpin developments in other critical areas, from artificial intelligence to biotechnology.

Quantum computers perform operations on quantum bits (*qubits*) that are in a superposition state and thus offer a new set of possibilities and challenges that result in the parallelization of operations and increased computing power. Moreover, as a foundational technology, quantum computing will impact all strategic sectors, from energy to transportation. Quantum computers will be key in solving problems that current supercomputers cannot solve. Still, though the field is advancing rapidly, quantum computers are currently far from being useful machines.

Quantum communications are another subset of technologies that use these physical mechanics to secure communications. Using technologies such as quantum key distribution (QKD), they mostly exploit the principle of entanglement to ensure communications work at a distance with new characteristics that prevent some concerning types of cyberattacks, such as

man-in-the-middle. However, there are still technical and research limits to overcome to make these networks fully operational—such as ensuring long-distance communication without information losing its quantum state or authentication.

Quantum sensing leverages quantum physics to enhance the precision and sensitivity of measurements. Quantum sensors are the most mature area within the realm of quantum technologies. Atomic clocks or magnetometers use quantum effects and are currently deployed to help in financial transactions or navigations. In the area of security and defense, quantum sensors are key in enhancing early detection and improving intelligence and reconnaissance operations, especially in GPS-denied areas.

Quantum technologies will be fundamental to boosting future developments in areas such as advanced cybersecurity and artificial intelligence. For that reason, they are increasingly subject to global geopolitical dynamics. In the last seven years, advanced economies have started to invest heavily in the development of quantum technologies, with the imperative of securing access to and control over their capacities.

Technological breakthroughs in error correction and coherence time, as well as increased awareness in risk areas like cybersecurity, have been driving public discourse towards the securitization of quantum technologies. This process, which cannot be analyzed in isolation, is already affecting other developing deep technologies, such as artificial intelligence. With an increasingly uncertain international political environment, questions around the control and access to critical technologies are fundamental to understanding current geopolitics. Quantum technologies have only recently joined the game.

Quantum technologies are both an emerging cybersecurity threat and a potential solution. Since the 1990s, with the publication of Shor's (1994) and Grover's (1996) algorithms, scientists have feared that quantum computers will be able to break encryption systems in use. Encryption, at the

backbone of current cybersecurity systems, from sensitive intelligence to web searches, protects information from being manipulated and rewritten, that the message is available, and ensures that only authorized parties can have access to it.

Truth be told, though we are far away from that moment, developments in the quantum field have made evident that it is no longer only a possibility but a reality only a decade away. This has created a surge of malicious cyber activity in search of downloading encrypted information with the hope of reading it once the technology is available ("download-now-decrypt-later").

This behavior affects current cybersecurity not only by increasing the risk of overreaction even outside cyber means to increased cyberactivity, what some authors have called "the cybersecurity dilemma"[2] but also constitutes already a major national security risk. Most countries in the world have declassification laws with strict timelines to make intelligence public. In Europe, on average, countries declassify confidential information after 25 years, with the most sensitive to be declassified after 50 years (e.g. Belgium) or never. This means that we are at least 10 years late to protecting information.

This chapter will first start diving into the genesis of quantum international geopolitics with a description of the Chinese, American, and European quantum programs to paint a contemporary picture of current developments in the field. Then, it will describe the European governance architecture of quantum technologies and zoom into particular initiatives affecting information security in the age of quantum. After this, the chapter will zoom into developments in Taiwan with a particular focus on understanding the relationship between technological development and societal resilience. The chapter concludes that even though there is no strong collaboration between Taiwan and the European Union (EU) in quantum technologies yet, the geopolitical imperative of countering common threats and the complementarity of capabilities make collaboration a crucial next step.

Quantum Geopolitics

Investments in quantum technologies have surged over the past few years, signaling an increasing interest by both the private sector and nations. Combined, countries have committed to investing more than €40 billion in quantum technologies, with China, the EU (and member-states) and the United States (U.S) topping the chart,[3] and quantum computing and communications as the subfields being funded the most.[4] Moreover, start-ups have already raised more than €9 billion in the more than 490 funding rounds that have taken place since 2001,[5] and are being created at a fast speed, especially since 2021, after the COVID-19 pandemic.

China is still the leading public investor in quantum technologies, though reliable information is scarce. Efforts are centered on the Chinese National Laboratory for Quantum Information Sciences that has, until now, published breakthroughs in several areas, including the release of a 504-qubit chip in December 2024.[6] Since 2006, a series of policies in China have been helping it advance towards great power status in quantum technologies.[7] However, China is mostly known for its capacity in the field of quantum communications.

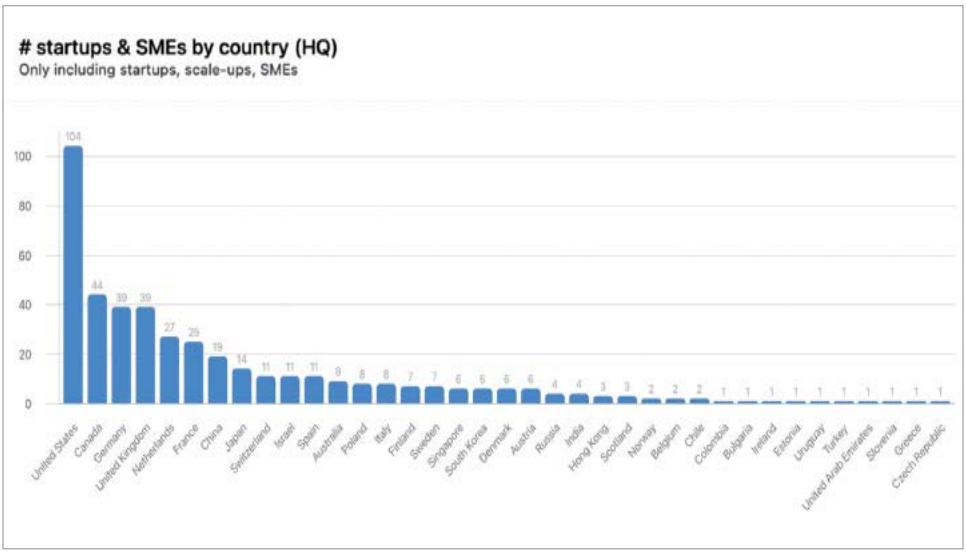

Figure 8.1: Distribution of quantum technology companies by country[8]

In 2013, the revelations of National Security Agency (NSA) contractor Edward Snowden of a global surveillance network by the U.S sparked conversations about the need to upgrade digital security and privacy. In the EU, it was the genesis of what eventually led to the creation of the flagship General Data Protection Regulation (GDPR). In China, however, Snowden made China focus further on the development of quantum technologies, particularly in the field of secure communications,[9] to protect Chinese networks from foreign intelligence operations.

The 13th Five Year Plan, in development as Snowden leaks reached the media, explicitly mentioned quantum communications as a "strategic industry". It is no surprise that China launched in 2017 the first quantum satellite, able to transmit information long-range in a quantum state. The test, a broadcast using quantum-secured communications of a scientific conference between Vienna and Beijing, proved the risks of technology transfer to authoritarian powers and the innovative strength of China, a country thought of before as being behind cutting-edge technology. The satellite, Micius, was developed in cooperation with the Austrian Academy of Science, which also sparked discussions on research integrity and the limits of international collaboration.

Micius' launch was also a wake-up call for the U.S. In 2018, the US National Cyber Strategy[10] recognized for the first time the impact that quantum technologies would have on information security. It mandated the need to establish quantum-safe solutions, particularly the use of post-quantum cryptography (see Box 1). In parallel to this, the U.S. National Institute for Standards and Technology (NIST), which was already looking for cybersecurity solutions in response to quantum computing risks since 2016, published an open call for researchers to submit their post-quantum cryptography algorithms for review. Four rounds have already been celebrated, with four algorithms shortlisted for standardization.[11]

Quantum-safe Solutions

As countries ready their cybersecurity structures for quantum computers, there are still questions about which technologies are better for securing information. To this day, the two most promising are quantum key distribution (QKD) and post-quantum cryptography (PQC), both of them offering a different set of advantages and disadvantages over the other.

Quantum key distribution (QKD) enables two parties to establish a secure communication channel based on quantum physics. Because of the properties of quantum bits (qubits), data shared cannot be copied, which protects against information theft during communications. Moreover, any disturbance or interference in the communication channel can be perceived by the parties that can suddenly decide to stop communicating. This offers a unique advantage against eavesdropping, where a third party "listens" to the conversation.

However, while eavesdropping can be detected, QKD requires pre-sharing encryption keys, which can create an authentication problem. An unauthorized party could potentially supplant the identity of one of the parties ("man-in-the-middle"). Moreover, QKD requires specific infrastructure, which increases the time and cost of the transition, and its sensibility to eavesdropping could increase the risk of denial of service (DoS) cyberattacks. Also, there are still multiple challenges to widespread adoption, such as the distance at which communication can happen (hardly over 200km nowadays) and the need to use trusted nodes to solve this, to go over 200km. For all these reasons, while QKD applications are promising and can add value in the long-term, they are generally perceived as still in the early stages of development.

Post-quantum cryptography (PQC), as a classical computing solution, is a more mature activity area and offers several advantages over QKD. At the same time, it also has theoretical and practical challenges. PQC can be defined as a set of cryptographic algorithms which are believed to be quantum resistant. These algorithms run on classical hardware, which makes their deployment much faster and cheaper as it would involve little more than a software update. However, PQC protocols have the same vulnerabilities as current cryptographic systems. Further technological advancements could allow for the retrospective decryption of these algorithms, hence the reason why the NIST competition is still ongoing. In other words, no practical proof exists that more sophisticated decryption algorithms, besides those already known run by quantum computers, would not break post-quantum cryptography being developed today.

Box 1: Comparison of quantum-key distribution vs post-quantum cryptography (retrieved from G. Rodríguez, 2023: 5)[12]

Soon after the publication of the 2018 cyber strategy, the U.S. established the National Initiative Act[13] with specific targets, budget, and actions across three pillars: research and innovation, industrial development, and security and defense. The Initiative marks a whole-of-government approach to quantum technologies, elevating them to a top position in the political agenda.

Furthermore, in 2022 the U.S. passed the Quantum Cybersecurity Preparedness Act to set a roadmap for the migration of sensible encryption systems to those quantum-secure and two White House memorandums mandated public agencies to establish an inventory of used cryptography offering a timeline for migration. This commitment to increasing cybersecurity levels in the era of quantum computing has been further reinforced by the latest National Cybersecurity Strategy (2023)[14] that dedicates a whole section to quantum technologies and their impact on information and national security. Moreover, the U.S. has launched the QuANET,[15] a program to develop and deploy quantum networks where quantum solutions are progressively incorporated into classical networks.

Quantum Cybersecurity in the EU

The COVID-19 pandemic had shown that the EU's dependence on foreign technologies and critical components could compromise its security. In the words of ex-Commissioner Thierry Breton, it was important to "have a discussion, without naivety, and without taboos, on the toolbox we need to guarantee our security of supply for our most critical value chains in case of crisis."[16]

The Recovery and Resilience Facility (RRF) was instrumental for countries adventuring into the quantum field. For member-states to unlock the €712 billion of investments promised by the EU, they had to put out plans that proved that at least 20 percent of the funds would be used to finance the digital transition.[17]

Prior to 2021, quantum policy in the EU was mostly led by the EU Quantum Flagship, an initiative led by academics under the auspices of the European

Commission (EC) to coordinate research into quantum technologies (see Box 2). In parallel to this, before 2021, only the Netherlands had put in place a national quantum agenda (2019),[18] though Germany had also announced investments in quantum as part of its Economic Stimulus Package (2020).[19] After the pandemic, many other European countries signed up for the national financing of quantum technologies, with France publishing its national strategy in 2022,[20] and Denmark[21] and Ireland in 2023.[22]

EU Quantum Flagship

The EU Quantum Flagship (2019) was the response of the EU to the manifesto launched in 2016 by academics and some companies for the EU to increase strategic investments in an area where the EU had shown academic excellence and leadership. The Flagship, which counts with €1 billion additional funding, has become the center of coordination efforts of the EU vis-à-vis scientific research across five areas: computing, communications, sensing and metrology, simulation, and basic science.

It has undergone three different phases. During its inception phase (2016-2019), it aimed for scientific leadership, not geopolitical edge. However, efforts during the consolidation phase (2019-2021) and the current one are increasingly making the Flagship advance towards strategies that prioritize commercialization and integration of quantum technologies in vertical industries.

Box 2: Short overview of the EU Quantum Flagship

The research focus of the EU contrasts with the geopolitical and security and defense dimension that underpins most national quantum strategies. These understand, in one way or another, that quantum technologies are key for competitiveness but also national security. Though the new EC, whose term started in October 2024, has placed efforts to increase research and industrial capacity in quantum technologies higher on the political agenda, there are still questions about whether EU action would be able to offer coordination in a very fragmented scenario or if, by contrast, will be able to use fragmentation to improve Europe's position vis-à-vis other technological giants.

In the field of cybersecurity, this fragmentation is particularly evident. Cybersecurity, deeply connected to notions of individual sovereignty, is a

topic where traditionally EU joint action has been successful. Since the 2001 Echelon scandal[23] in which the journalist Duncan Campbell revealed a spy network that allowed the U.S. NSA to tap into European communications using signal intelligence, the EU has been one of the main actors shaping the cybersecurity agenda.

The EU's role has strengthened from crisis to crisis. The 2007 cyberattacks on Estonia,[24] the surge in industrial espionage as exemplified by Operation Aurora in 2010,[25] or the WannaCry incident of 2016[26] have been catalysts for policy action. Successes include the Network and Information Systems Directive (NIS and NIS 2) and the Cyber Resilience Act ("first IoT law in the world") as well as several initiatives to combat cybercrime, improve cyber resilience and coordination.

However, in the realm of quantum cybersecurity, the mystery and complexity of quantum physics and the lack of awareness of a near-term threat has paralyzed EU action in favor of individual actions in member-states. This disparity is what potentially makes the EU sensitive to quantum-enabled operations. After all, a network is as strong as its weakest link and only a handful of countries are taking action to implement quantum-safe solutions, such as France, Germany or the Netherlands.

Arguably, there are only two EU-level initiatives directed towards securing information systems against quantum attacks. The first one is the 2024 Coordinated Implementation Roadmap for the transition to Post-Quantum Cryptography[27] in which the EC recommends member-states to elaborate national strategies with clear goals for the transition to PQC without mandating any concrete action. EC recommendations are voluntary documents that countries can either consider or not.

Secondly, EuroQCI, which is a flagship initiative managed by the EU Quantum Flagship, and where all EU27 participate. It aims to establish a European quantum communications network connecting strategic sites,[28] such as embassies and public infrastructure, using fiber optic cables and a

network of future quantum satellites (IRIS2).[29] The main technology used in EuroQCI are QKD networks that have proven to not be mature enough yet for deployment but are expected to grant an unprecedented level of security once certain technical issues are resolved.

Taiwan's Agenda Against Foreign Interference and Adoption of New Technologies

As major global players like the U.S., China and Europe advance their quantum capabilities, Taiwan has made significant efforts in recent years to strengthen its position in this field. Taiwan's push towards quantum technology development is driven by two primary factors: global technological advancements and national security concerns. A SWOT (strengths, weaknesses, opportunities, and threats) analysis conducted by Taiwan's Ministry of National Defense in 2023 warned that Taiwan is at least 20 years behind the global competition in quantum computing, stating this gap as a critical vulnerability.[30]

In response to global shifts, Taiwan's Executive Yuan has pointed out that the technology required to develop quantum computers is similar to Taiwan's well-established chip manufacturing processes and that Taiwan already has well-positioned scientific actors, such as Academia Sinica, which gives Taiwan an advantage to enter the quantum race.[31] Therefore, domestic strengths, with the global momentum in quantum research, pushed Taiwan to jump on board and announce NT$8 billion investment in quantum technology (€232 million for the next five years (2021-2026).[32]

Enhancing its competitiveness in quantum technologies has become a national security imperative, particularly in response to the growing interest of the Chinese Communist Party (CCP) in quantum tech. In October 2020, the CCP Politburo held a collective study session on the "Research and Prospective Application of Quantum Technologies."[33] This hints at the attention placed on quantum technology by Chinese military, political, and economic leaders, as emphasized by the Taiwanese Institute for National Defense and Security Research (INDSR).[34] In the context of information warfare, especially during

the CCP's military exercises, Taiwan views the potential deployment of Chinese quantum technology as a serious threat.

With more advanced quantum capabilities, CCP-led cyberattacks could more effectively breach Taiwan's cyber defenses, enabling more precise and widespread disruption like those during Nancy Pelosi's 2022 visit, where Taiwan's presidential office website was down for 20 minutes, and malicious connections peaked at over 170 million per minute. With greater quantum capabilities, such disruptions could be more persistent in time and frequency. This not only means that Taiwan must develop more resilient cybersecurity protocols but also invest in advancing its own quantum technology to counter these emerging threats.

This leads to a more critical concern, which is the potential for societal subversion. By leveraging quantum offensive capabilities, the CCP could conduct more sophisticated disinformation campaigns, amplifying the scale, reach, and complexity of interference, severely undermining public trust and destabilizing Taiwan's political landscape. The INDSR has highlighted that Taiwan's quantum technology is not exclusive to the field of science and technology. It requires a more interdisciplinary approach, integrating military intelligence, psychology, and public opinion experts to explore security implications of quantum technology in cognitive warfare.[35]

Current Developments

In 2021, Taiwan launched an interministerial roadmap[36] for quantum technology. A key ambition of the initiative is to build a Taiwan National Quantum Team led by Ministries of Economic Affairs, the National Science and Technology Council and Academia Sinica, to foster collaborative R&D among industry, academia and research institutions, while setting up an industry cooperation platform. The roadmap outlines three major pillars: the national team, research infrastructure, and international collaboration.

By 2024, the government had selected 17 research groups,[37] and Academia Sinica's south campus broke ground on the Quantum Technology

Experimental Building to serve as a national quantum research hub.[38] Taiwan has also started to enhance research exchanges through delegation visits, such as with Finland and France.[39] All these efforts reflect Taiwan's strategic push to strengthen its quantum capabilities, an essential priority amid growing geopolitical tensions and information warfare, where cybersecurity and resilience remain a key national concern.

The Taiwanese Ministry of Justice Investigation Bureau, responsible for countering foreign interference, has kept track of cases and tactics of disinformation over the years. Early methods around 2020 relied on text-based attacks, image manipulation and counter-fact-check smear campaigns. By 2023, the rise and widespread use of AI led to a surge in deepfake videos circulating on YouTube and Facebook.[40] The bureau uncovered networks of fake accounts created through automated systems which helped spread disinformation at an unprecedented scale, especially in the context of the 2024 election cycle.

As the election year approached, forum-based disinformation was significantly deployed. Taiwanese investigations traced fake Facebook pages and accounts and coordinated mass-sharing campaigns to individuals from Cambodia and Myanmar.[41] Over time, waves of false narratives made their way to Taiwanese users, which weakened public trust in government and reliable sources, turning information warfare into a growing challenge for Taiwan's democracy.

According to a recent analysis by Taiwan's National Security Bureau (2025),[42] cases of CCP-backed fake news almost doubled in only one year, from 1,329,000 recorded in 2023 to 2,159,000 in 2024. This surge of disinformation was particularly marked by skepticism discourses towards the U.S., Taiwan's military, and President Lai-Ching-Te.

Conclusions

The connection between quantum technologies and information security, particularly in the fields of foreign interference, is rather new and comes from the fear of a foreign action having full access to everything happening behind

information safeguards. Moreover, while no big cybersecurity incident yet has happened involving the use of quantum technologies, countries are developing their quantum cybersecurity agendas motivated more by the securitization around quantum technologies rather than by the perception of a real threat.

This comparative asymmetry, in which a country advances its quantum agenda based on developments happening elsewhere, is what is driving Taiwan to invest and advance its quantum agenda. With China as a traditional geopolitical rival becoming increasingly a quantum superpower, efforts to develop, experiment, implement, and adapt quantum solutions are still in a very early stage. However, Taiwan will continue to advance its quantum R&D agenda paying a particular focus on initiatives that can increase its level of cyber resilience also in light of the potential effects on public trust and the spread of disinformation.

The development of quantum technology is therefore a crucial step for Taiwan as it navigates growing cyber threats and the accelerating risk of information warfare. The security imperative extends beyond cybersecurity as it is also about protecting the integrity of Taiwan's democratic processes against disinformation and foreign interference, which could exacerbate the polarization of Taiwanese society, particularly during critical political moments such as election periods. Developing cutting-edge capabilities like that of quantum computing is therefore becoming increasingly central to the dynamics of information warfare.

Though currently the span of collaboration between the EU and Taiwan in quantum technologies is not really advanced, though it might be the case with individual member states such as Finland, the current geopolitical scenario calls for new partnerships and Taiwan has a well-developed technological industry and strong university-industry links as well as an advanced IT workforce that will be soon quantum-ready.

Endnotes

1 Lars Jaeger, *The Second Quantum Revolution* (Springer eBooks, 2018), https://doi.org/10.1007/978-3-319-98824-5.

2 Ben, Buchanan, *The Cybersecurity Dilemma: Hacking, Trust and Fear Between Nations*, 2017, http://cds.cern.ch/record/2263995.

3 McKinsey Digital, "Quantum Technology Monitor," April 2023, https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/steady%20progress%20in%20approaching%20the%20quantum%20advantage/quantum-technology-monitor-april-2024.pdf.

4 Ibid.

5 "Infinity by Quantum Delta NL," n.d., https://www.infinityqd.nl/.

6 Chinese Academy of Sciences, "China Unveils Record-breaking 504-qubit Superconducting Quantum Computer," December 6, 2024, https://english.cas.cn/newsroom/cas_media/202412/t20241206_893281.shtml.

7 Andrea G. Rodriguez, "Cybersecurity Implications of Quantum Computing and Its Combined Use With Artificial Intelligence," *UNISCI Journal*, January 2025, https://www.unisci.es/wp-content/uploads/2025/01/UNISCIDP67-5ANDREA.pdf.

8 "Infinity by Quantum Delta NL," n.d., https://www.infinityqd.nl/.

9 Yale Journal of International Affairs, "China's Quantum Ambitions: A Multi-Decade Focus on Quantum Communications," May 23, 2024, https://www.yalejournal.org/publications/chinas-quantum-ambitions.

10 The White House, "National Cyber Strategy of the United States of America," September 2018, https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf.

11 NIST, "Post-Quantum Cryptography PQC," February 24, 2024, https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022.

12 Andrea G. Rodriguez, "A Quantum Cybersecurity Agenda for Europe: Governing the Transition to Post-quantum Cryptography," European Policy Centre, July 17, 2023, https://www.epc.eu/content/PDF/2023/Cybersecurity_DP.pdf.

13 National Quantum Coordination Office, "National Quantum Initiative," April 2021, https://www.quantum.gov/wp-content/uploads/2022/04/NQI-Factsheet.pdf.

14 The White House, "The National Cybersecurity Strategy," May 2, 2023, https://bidenwhitehouse.archives.gov/oncd/national-cybersecurity-strategy/#:~:text=The%20National,Cybersecurity%20Strategy&text=The%20Biden%2D%E2%81%A0Harris%20Administration,digital%20ecosystem%20for%20all%20Americans.

15 DARPA, "QuANET: Quantum-Augmented Network," n.d., https://www.darpa.mil/research/programs/quantum-augmented-network.

16 Thierry Breton, "A Stronger Industry for a More Autonomous Europe," January 13, 2022, https://ec.europa.eu/commission/presscorner/detail/en/speech_22_354.

17 EU Commission, "NextGenerationEU," n.d., https://commission.europa.eu/strategy-and-policy/eu-budget/eu-borrower-investor-relations/nextgenerationeu_en.

18 Quantum Delta Nederland, "National Agenda for Quantum Technology," September 2019, https://qutech.nl/wp-content/uploads/2019/09/NAQT-2019-EN.pdf.

19 Federal Government Bundesregierung, "Revitalising the Economy," June 12, 2020, https://www.bundesregierung.de/breg-en/news/corona-steuerhilfegesetz-1760128#:~:text=A%20cut%20in%20value%20added,the%20country%20for%20the%20future.

20 CNRS, "The French Quantum National R&D Strategy Just Started," March 7, 2022, https://www.cnrs.fr/en/update/french-quantum-national-rd-strategy-just-started.

21 Ministry of Foreign Affairs of Denmark, "New National Quantum Strategy Part 2 Provides Significant Boost to the Danish Quantum Ecosystem," September 28, 2023, https://investindk.com/insights/new-national-quantum-strategy-part-2-provides-significant-boost-to-the-danish-quantum-ecosystem.

22 Government of Ireland, "Quantum 2030: A National Quantum Technologies Strategy for Ireland: Putting Ireland in a Quantum Super Position," November 2023, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwieiajO59-DAxUkFxAIHdiCAi8QFnoECBIQAw&url=https://assets.gov.ie/276661/61f7f8f1-636b-48e1-91c9-a9bf63e00ce8.pdf&usg=AOvVaw3W8hNgfEy4OiBXFAnM_l1l&opi=89978449.

23 European Parliament, "REPORT on the Existence of a Global System for the Interception of Private and Commercial Communications (ECHELON Interception System) (2001/2098(INI)) Part 1: Motion for a Resolution Explanatory Statement Part 2: Minority Opinions Annexes," July 11, 2001, https://www.europarl.europa.eu/doceo/document/A-5-2001-0264_EN.html#_section3.

24 Rain Ottis, "Analysis of the 2007 Cyber Attacks Against Estonia From the Information Warfare Perspective," Cooperative Cyber Defence Centre of Excellence, 2008, https://ccdcoe.org/uploads/2018/10/Ottis2008_AnalysisOf2007FromTheInformationWarfarePerspective.pdf.

25 Council of Foreign Relations, "Operation Aurora," January 2010, https://www.cfr.org/cyber-operations/operation-aurora.

26 ENISA, "WannaCry Ransomware: First Ever Case of Cyber Cooperation at EU Level," Press release, May 15, 2017, https://www.enisa.europa.eu/news/enisa-news/wannacry-ransomware-first-ever-case-of-cyber-cooperation-at-eu-leve.

27 European Commission, "Recommendation on a Coordinated Implementation Roadmap for the Transition to Post-Quantum Cryptography," April 11, 2024, https://digital-strategy.ec.europa.eu/en/library/recommendation-coordinated-implementation-roadmap-transition-post-quantum-cryptography.

28 European Commission, "The European Quantum Communication Infrastructure (EuroQCI) Initiative," 2019, https://digital-strategy.ec.europa.eu/en/policies/european-quantum-communication-infrastructure-euroqci.

29 European Commission, "IRIS2: The New EU Secure Satellite Constellation: Commission Takes Next Step to Deploy the IRIS2 Secure Satellite System," December 16, 2024, https://defence-industry-space.ec.europa.eu/eu-space/iris2-secure-connectivity_en.

30 Heh, Tzeng-Yuan 賀增原, "運用SWOT去探討量子科技的發展." 陸軍後勤季刊 112, no. 1 (February 1, 2023): 30–37, https://doi.org/10.53106/230674382023021121003.

31 Executive Yuan, "蘇揆：跨部會籌組量子國家隊　5年投入80億元提升我國量子科技實力," Press release, December 2, 2021, https://www.ey.gov.tw/Page/9277F759E41CCD91/70498197-800f-4148-a93c-242e2411727b.

32 Ibid.

33 The State Council of The People's Republic of China, "习近平主持中央政治局第二十四次集体学习并讲话," October 17, 2020, https://www.gov.cn/xinwen/2020-10/17/content_5552011.htm.

34 Tzeng-Yuan Heh, 賀增原, "Interdisciplinary Thinking to Strengthen the Development of Quantum Technologies, 跨異質思維去強化量子技術的發展," Institute for National Defense and Security Research, August 26, 2022, https://indsr.org.tw/uploads/indsr/files/202209/8d1bc6d2-7ede-45cc-a12d-e71123461190.pdf.

35 Ibid.

36 Ibid.

37 International Year of Quantum Science and Technology, "量子國家隊成軍，17項團隊底定，聚焦未來量子世代臺灣產業鏈," September 27, 2024, https://iyq.tw/news/66f62f8cbe04dcd1c2d6be5c.

38 Academia Sinica, "Academia Sinica South Campus Breaks Ground on Quantum Technology Experimental Building to Establish National Quantum Research Hub," November 22, 2024, https://www.sinica.edu.tw/en/news_content/55/2820.

39 Taiwan Today, "NSTC Delegation Tours Finland, France to Expand Quantum Technology Cooperation," July 4, 2024, https://www.taiwantoday.tw/Economics/Top-News/255159/NSTC-delegation-tours-Finland%2C-France-to-expand-quantum-technology-cooperation.

40 Ministry of Justice Investigation Bureau (Taiwan), "境外敵對勢力介入我總統大選國人宜謹慎識別網路假訊息," December 20, 2023, https://www.mjib.gov.tw/news/Details/1/953.

41 *Ibid.*

42 National Security Bureau, "2024 年中共爭訊傳散態樣分析," January 3, 2024, https://www.nsb.gov.tw/zh/assets/documents/%E6%96%B0%E8%81%9E%E7%A8%BF/2024%E5%B9%B4%E4% B8%AD%E5%85%B1%E7%88%AD%E8%A8%8A%E5%82%B3%E6%95%A3%E6%85%8B%E6%A8%A3%E5%88%86%E6%9E%90(%E5%A0%B1%E5%91%8A%E5%85%A8%E6%96%87)-%E4%B8%AD%E6%96%87.pdf.

9. Striking a Balance: Between Technological Advancement and Accountability in the Ever-Changing Information World

Yi-Chieh Chen and Niklas Swanström

Three of the distinguishing foundational pillars of democratic governance and resilience are trust, transparency, and accountability. However, in today's transnationally contested information environment, amplified by the ever-changing and rapid advancement of technologies, the architecture devoted to the maintenance of the abovementioned pillars is coming under pressure. The institutions once tasked to establish and sustain shared truths, now compete with decentralized platforms, synthetic content, and the viral tempo of information. The result of these trends is not solely an uptick in coordinated deception and an easier spread of falsehoods but also a deeper erosion of the democratic infrastructure that filter, verify and protect democratic truths and legitimacy.

This volume has charted the evolution of that erosion and the tools emerging to counter it within the European and Taiwanese landscapes. The increasing adoption of artificial intelligence (AI), the disaggregation of platform governance, and increasing recourse to strategic disinformation by both state and non-state actors merge to produce an unstable epistemic terrain. Both Europe and Taiwan are characterized by increased pressures, albeit under different guises.

On the one hand, Taiwan, facing persistent influence campaigns from the People's Republic of China (PRC), has become a testing ground for the strategic weaponization of disinformation. Here, the battlefield is not simply

electoral, but also aimed at fracturing a sense of collective reality, particularly during moments of democratic transition such as the 2024 presidential election. Undermining the democratic system in Taiwan is arguably of greater importance for PRC than to simply influencing the election outcome in a particular way, especially as Beijing has little confidence in any Taiwanese political party. On the other hand, while Europe's experience mirrors this, it does so through different vectors. Russian operations around the war in Ukraine, and platform-driven disinformation targeting migration, climate policy, and LGBTQ+ rights, have strained the resilience of European public discourse. These disruptions are not isolated but systemic and reinforced by Russian and Chinese attacks. The stress tests on the ability of democratic polities to process truth, contest ideas without collapsing into polarization, and govern the digital spaces are being severely tested and are not always showing the resilience they would need.

What links the European and Taiwanese experiences is not only the vulnerability to disinformation, but the institutional improvisation that has emerged in response. As the various contributions to this volume have demonstrated, fact-checking initiatives, regulatory experimentation, civic-tech alliances, and new verification protocols have proliferated across both landscapes in an effort to manage the new challenges. However, their effectiveness is increasingly constrained by structural imbalances between democratic institutions and the commercial platforms that mediate much of contemporary discourse. These imbalances are evident in delayed legislative responses, shifting platform strategies for implementing relevant measures, and the inherently uncertain effectiveness of self-regulatory and voluntary frameworks.

Common Challenges and the Potential to Establish a Complementary Regulatory Framework in Europe and Taiwan

While the strategic use of deception and narrative control has accompanied conflict throughout history, today's digital disinformation campaigns differ less in kind than degree. What has changed is not the battlefield, but the permeability of its boundaries, amplified by platforms, sustained by algorithms

and accelerated by AI. Against this backdrop, the institutions meant to safeguard informational integrity are themselves fraying. What began as platform experimentation with self-regulation has, in recent years, turned to retreat and recalibration. Key technology and media platforms' voluntary commitments that once anchored content moderation and fact-checking partnerships are dissolving under political pressure and economic disincentive. Meta's abrupt termination of its Third-Party Fact-Checking Program (3PFC), Google's quiet withdrawal from the European Union's (EU) Code of Conduct on Disinformation under the Digital Services Act (DSA), and the rollout of "Community Notes" by Meta as a substitute for verified oversight signal a profound shift, which can be seen as an indirect result of the change of political landscape (see Chapter 7). In other words, what may have once seemed like a scattered set of operational issues now reveals itself as a structural crisis. The problem is not simply disinformation in itself but the weakening of the very institutions and infrastructures tasked with filtering and correcting it.

In the U.S., under the Trump administration, which believes freedom of speech in the Europe is retreating and blatantly states its opposition against fact-checking, namely, the reframing of moderation as censorship has created political cover for these retreats. In Europe, where frameworks like the Digital Services Act (DSA) and AI Act seek to embed democratic norms into the architecture of digital governance, such reversals expose the fragility of relying on discretionary compliance. Enforcement remains a challenge, and the legitimacy of intervention is continuously tested by accusations of overreach or bias, not least by proxies of Moscow and Beijing, to challenge the very political system on which it is based.

Taiwan presents a contrasting, yet complementary model, which has prioritized public literacy, agile civil society partnerships, and participatory models of verification. Open-source initiatives, government-industry collaboration, and high levels of citizen engagement have allowed Taiwan to treat disinformation not only as a policy problem but as a civic challenge. This is less about regulating platforms into submission, and more about cultivating collective resilience.

Europe and Taiwan thus articulate two distinct yet converging strategies. Europe leans on institutional codification; Taiwan invests in societal calibration. One attempts to enshrine accountability into law and design, the other fosters it through networked vigilance and trust-building. These are not mutually exclusive paths, on the contrary in many ways they can reinforce each other. In an increasingly interconnected informational landscape, where disinformation knows no borders and platforms scale globally but govern unevenly, hybrid approaches will be essential.

Quantum Technologies in Information Warfare?

Yet even as democracies contend with the withdrawal of platform governance and the erosion of institutional backstops, new technological frontiers are emerging, quietly, preemptively and with ambiguous consequences. Among these, quantum technologies stand out for their potential to reshape the informational landscape in unpredictable ways.

According to the analyses in this volume, to date, the development of quantum technologies has been devoted to the maximization of long-term strategic utility, enhancing computational capacity, encryption advancements, and redefining the speed and scale of data processing. Their immediate application to disinformation campaigns remains limited, and more speculative than operational. However, therein lies the danger as the window between conceptualization and weaponization is shrinking and hidden in the shroud of authoritarianism. The same quantum systems designed to secure communications could, under different hands, be leveraged to fracture them.

What is most striking is the asymmetry in regulatory foresight. The EU has begun to map out precautionary frameworks, including its Quantum Technology Impact Assessment in 2023 which aims to assess the ethical impact of quantum technology on society.[1] Meanwhile, Taiwan does not have specific guidelines or regulations regarding quantum technologies. Still, it is expected that Taiwan's National Security Act may include quantum technologies in the near future, as they are key technologies that need to be protected to safeguard Taiwan's national security in the information world.

This divergence is not simply of timing but reflects the different strategic outlooks of Europe and Taiwan. The former seeks to bind the future through anticipatory governance, while the latter acutely attuned to existential risk, prioritizes flexibility and securitization. Nonetheless, both recognize that tomorrow's disinformation infrastructure will have to be built around quantum architectures, synthetic media ecosystems, and machine learning pipelines that operate at speeds human oversight can barely match. Simultaneously, also cognizant of the deployment of these tools in information conflicts, both Taiwan and Europe, as vibrant democracies, are attempting to imagine and preempt their misuse before they scale. As with earlier phases of digital transformation, those who anticipate the risks, not just the efficiencies, will be best positioned to govern them.

Conclusion

What emerges across these chapters is not a single prescription, but a mosaic of strategies. Europe and Taiwan, each with distinct institutional cultures and geostrategic pressures, reveal how democratic resilience must evolve in tandem with technological acceleration. Europe's strength lies in its ability to inscribe democratic values into regulatory frameworks; Taiwan, by contrast, offers a lesson in how civic actors, open-source communities, and targeted public education can become bulwarks when formal mechanisms are overstretched.

Neither path is sufficient on its own. What is needed is a reciprocal learning loop between structural regulation and grassroots responsiveness. As platforms retreat and new technologies improve at breakneck speed, the risk is that democracies will become reactive rather than anticipatory. The informational environment is kinetic, recursive, and adversarial, and, therefore, such needs to be the governance that attends to it.

Ultimately, combating disinformation in the age of AI, LLMs, and quantum architectures is not about restoring an idealized past of media certainty. Instead, it is about building an ecosystem where verification is not an

afterthought, but a structural condition of democratic participation, wherein its values are safeguarded and embedded in the digital discourse and all its strata, and truth is designed to remain resilient.

This volume concludes not with answers, but with an orientation toward a future in which democratic life, participation and values are supported by the integrity of the information that circulates and permeates it. The burgeoning task of democratic polities like Taiwan and the EU does not lie in chasing each and every falsehood in order to police it, but to actually design systems where truth is easier to access and find and overall harder to obscure.

It also calls for more, and continuous, research to act preventively, and reactively, to the technological leaps that the world will take in the coming months and years. Technology is moving much faster than legislation, and in the development of strong institutional measures we also need a vibrant civil society that can generate new thinking about future challenges.

Endnotes

1 Mauritz Kop, "Quantum Technology Impact Assessment," European Commission, April 20, 2023, https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/quantum-technology-impact-assessment.