Data Mining, Algorithms, and Deepfakes

Combatting Misinformation, Disinformation, and Malinformation in 2024 Elections and Beyond

DRI Center for Law and Public Policy Data Privacy and Protection Working Group

THE CENTER
The Voice of the Civil Defense Bar.

Data Mining, Algorithms, and Deepfakes
Combatting Misinformation, Disinformation, and Malinformation in 2024 Elections and Beyond

Laura Clark Fey

Sean C. Griffin

Brent Arnold

Josh Devaney

Richik Sarkar

Tobias Schelinski

Kirsten E. Small

DRI
222 South Riverside Plaza, Suite 1870
Chicago, Illinois 60606
dri.org

Produced in the United States of America

DRI CENTER FOR LAW AND PUBLIC POLICY DATA PRIVACY AND PROTECTION WORKING GROUP

Laura Clark Fey (Chair), one of the first twenty-seven U.S. attorneys recognized as Privacy Law Specialists through the International Association of Privacy Professionals (IAPP) and a certified AI Governance Professional (AIGP), leads Fey LLC, a global data privacy, AI, and information governance law firm. Laura is a member of the inaugural class of IAPP Fellows of Information Privacy (FIP), a Certified U.S. and European Information Privacy Professional (CIPP/US/E), and a Certified Information Privacy Manager (CIPM).

Sean C. Griffin (Vice Chair), a partner of Robinson + Cole LLP in Washington, D.C., is a former DOJ attorney with 30 years of commercial litigation experience, who helps insurance companies, government contractors, construction companies, and other businesses manage complex contract and fraud matters. And as one of the world's first IAPP-certified Artificial Intelligence Governance Professionals (AIGP) and a Certified Information Privacy Professional (CIPP/US), Sean helps clients establish and maintain data security, respond to data breaches, and litigate privacy cases.

Brent Arnold is a Toronto-based trial and appellate litigator at Gowling WLG (Canada) LLP, specializing in technology-related commercial litigation. He is also a data breach coach and counsel. Brent chairs DRI's Cybersecurity and Data Privacy Committee. He is also a director of the Canadian Internet Society, a not-for-profit thinktank devoted to internet policy, and a host of its *Net Positive* podcast.

Josh Devaney is a Senior Associate at Kennedy & Graven, Chartered, a boutique firm primarily representing government entities in litigation. He has extensive experience in civil litigation, land-use, and cybersecurity and data privacy. He is a member of DRI's Governmental Liability and Cybersecurity and Data Privacy Committees.

Richik Sarkar is a courtroom advocate, boardroom strategist, collaborative leader, and fixer. A Dinsmore & Shohl equity partner in Cleveland, Richik specializes in counseling companies, directors, and management about business litigation, fiduciary claims, consumer practices, ESG, data privacy, and investigations. Richik provides bespoke commercial judgment, legal intelligence, and strategic vision to execute action plans.

Tobias Schelinski is a partner at TaylorWessing and is based in the firm's office in Hamburg, Germany. He primarily focuses on data law, in particular EU privacy law and other data-related EU regulations and directives.

Kirsten E. Small, CIPP/US, is a shareholder of Maynard Nexsen in Greenville, South Carolina. A member of the International Association of Privacy Professionals (IAPP), she came to the field of privacy law through her work as a litigator and appellate lawyer, a background that gives her a unique insight on how a company's policies and actions before a data breach can help mitigate—or avoid—liability if a breach occurs.

TABLE OF CONTENTS

FOREWORD

By John C.S. Pierce
Public Policy Committee Chair

One of DRI's essential missions is to help ensure fair and level determinations of civil actions and to work for the reasonable and just advancement of the law. Members of DRI—and its Center for Law and Public Policy—have worked tirelessly, providing programming, publications, amicus efforts, testimony to legislative and rules committees, and education to judges. Accurate, reliable, factual information is essential to these efforts, the work of our organization's push for fundamental fairness. But there is a growing threat to such information, undermining truth itself in favor of falsity.

One of the lodestars of fairness is truth. But what is "truth"? According to Merriam–Webster, it is "the body of real things, events, and facts." The fundamental importance of truth for lawyers is exemplified in the very way in which we prove facts. For example, the Federal Rules of Evidence exist to "promote the development of evidence law, to the end of ascertaining the truth and securing a just determination." FRE 102. What could be more relevant to your law practice than an existential threat to fair and reliable methods of establishing the body of real things, real events, and real facts?

For example, think on the importance of establishing personal knowledge, such as the requirement that a witness who testifies to a fact "had an opportunity to observe, and must have actually observed the fact" under Federal Rule of Evidence 602. The committee comments to that rule note that this is a "most pervasive manifestation" of the common law insistence upon "the most reliable sources of information." Most of our evidentiary rules focus on reliability: facts and data under Rule 705; prior witness statements under Rule 613; authentication under rules 901, 902, 903; and contents under rules 1001–1008.

Of course, the threat presented by falsity extends beyond the Rules of Evidence and the courts. An inability to establish what is accurate, correct, and true has the potential to undermine trust in the legal system, government and our entire society. The white paper which follows explores the growing body of disinformation, misinformation, and malinformation, and the threat that these things present in the context of a significant election year.

I hope you will consider the excellent work of these authors in the context of our ongoing quest for truth. Together we can combat this threat and continue to serve as a beacon of fairness for the clients we represent and society as a whole.

INTRODUCTION

"A lie can travel halfway around the world while the truth is still putting on its shoes."[1]

The World Economic Forum (WEF) selected disinformation and misinformation as the most severe short-term risk the world faces in 2024. The Center's Data Protection and Privacy Working Group[2] prepared this white paper to provide examples of recent misinformation, disinformation, and malinformation (MDM) campaigns; highlight for DRI members why disinformation, misinformation, and malinformation risks merit the WEF's highest risk ranking and explain why these risks should be on the radar of DRI members; provide insights into how MDM campaigns work; give examples of actions being taken by countries around the globe to confront this risk; and provide recommendations on actions that DRI members, their clients, and others can take to combat MDM risks in 2024 and beyond.

The Significance of the Global MDM Threat

The Cybersecurity and Infrastructure Security Agency (CISA), the U.S. federal agency tasked with protecting critical infrastructure in the United States against cyber threats, defines "misinformation" as information that is false, but not created or shared with the intention of causing harm.[3] Disinformation is defined by CISA as information that is deliberately created to mislead, harm, or manipulate a person, social group, organization, or country.[4] And malinformation is defined as information that is based on fact, but used out of context to mislead, harm, or manipulate information.[5]

This election year poses unprecedented MDM challenges not only to those of us living in the United States, but also to societies around the world. A primary reason for this is because 2024 is an election year of unprecedented importance. Over 50 elections that will affect half of the world's population are planned for this year.[6] These elections will have huge implications for human rights, the global economy, relationships between countries, and for the likelihood of peace throughout the globe this year and for many years to come.[7]

Another reason is because recent advances in AI technologies have made it significantly easier to create and spread MDM. As DW Akademie[8] has noted, "Generative AI is the ultimate disinformation amplifier."[9] Generative AI (Gen AI) tools permit any individuals desiring to do so to "quickly and easily create massive amounts of fake content."[10] Broadly available and mostly unregulated Gen AI tools "make it possible for anyone to generate false information and fake content in vast quantities. These include imitating the voices of real people and creating photos and videos that

---

1 This quote has been commonly attributed to Mark Twain, but, in recent years, has been thought to be related instead to a line published by satirist Jonathan Swift. See https://www.nytimes.com/2017/04/26/books/famous-misquotations.html.
2 https://www.centerforlawandpublicpolicy.org/CENTER/Center/committees/sub/data-privacy.aspx.
3 CISA-Disinformation and COVID-19: How State and Local Officials Can Respond.
4 Id.
5 Id.
6 https://apnews.com/25-elections-in-2024-that-could-change-the-world.
7 Id.
8 DW Akademie Is Deutsche Welle's Center for International Media Development, Journalism Training and Knowledge Transfer.
9 https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier.
10 Id.

are indistinguishable from real ones."[11] As DW Akademie warned, "[T]oday, it is often impossible to tell if content originates from a human or machine, and if we can trust what we read, see or hear."[12] The serious risks posed by MDM merit our attention for a number of reasons. One key reason is that MDM is emerging as one of the most significant cyber threats to individuals and businesses worldwide. AI advances are making it easier for bad actors to use MDM in furtherance of a host of criminal initiatives—from stealing money from families, law firms, and other businesses to disseminating false information designed to ruin the reputations of businesses and individuals. MDM initiatives result in financial loss, reputational loss, and other serious, negative consequences for the targets of such campaigns.

Another key reason is that MDM is already rearing its ugly head in trials. There have been examples of litigants and their attorneys seeking to introduce deepfake evidence, as well as examples of litigants and their attorneys claiming legitimate evidence "could have been altered." Deepfakes and purported deepfakes in trial are expected to become more pervasive as AI technologies continue to evolve and deepfakes become increasingly convincing.

Deepfakes and purported deepfakes in trial are expected to become more pervasive as AI technologies continue to evolve and deepfakes become increasingly convincing.

With respect to the upcoming elections, arguably the most important reason is because, as lawyers, we have special responsibilities to uphold the rule of law. To meet our oaths to uphold our federal and state constitutions, we should seek to educate ourselves on evolving MDM election risks and do our part to help combat those risks. As Susan J. Kohlmann, NYC Bar Association President, recently noted, "Because free and fair elections are how 'the consent of the governed' is determined in our democracy, it is safe to predict that the grave and unprecedented threats to our elections in 2024 will be seen as a fraught moment in American history. As officers of the court and stewards of the rule of law, it is our obligation to meet the moment to protect the upcoming election and, with it, our democracy."[13] Lawyers and judges will have a critical role to play in helping to uphold the rule of law in this year's election and moving forward—just as they have in the past. The critical role, of course, is not a role that applies exclusively to U.S. lawyers. DRI members and other lawyers inside and outside of the United States have an important role to play in combatting MDM risks.

The Taiwan Example

The recent presidential election in Taiwan demonstrates how disinformation, in particular, is used by nation states to try to impact elections in other countries, and also shows how an aggressive and broad-based response to disinformation can help bolster confidence in elections. In the build-up to the election, China used multiple techniques to spread disinformation to prevent the election of the anti-Beijing Democratic Progressive Party.[14] One tactic was publishing a 300-page book about one of the candidates, titled *The Secret Life of Tsai Ing-wen*. Soon after its publication, dozens of videos began appearing on social media channels with AI-generated avatars reading portions of the book aloud. The book—which may have been the product of generative AI—became "a script for generative AI videos," according to one researcher.[15]

---

11 Id.
12 Id.
13 https://www.law.com/newyorklawjournal/2024/05/01/calling-all-lawyers-to-protect-the-2024-election/.
14 https://www.gmfus.org/news/online-assault-against-taiwan.
15 https://foreignpolicy.com/2024/01/23/taiwan-election-china-disinformation-influence-interference/.

Other disinformation efforts by China included publishing myriad "fake news" reports. Some of these efforts were directed at core pocketbook issues—such as suggestions that poisoned pork was being imported from the U.S. (presumably highlighting the danger of greater democratization and openness to the West). Other disinformation efforts were more directly aimed at showing the DPP candidate was in the thrall of the evil United States.[16]

AI-generated deepfake videos were disseminated across social media platforms and shared as many as 100 times per minute—a rate that can only be achieved by a bot.[17] Although initially easy to spot and debunk, the AI-generated videos and audio improved in quality and became more challenging to flag as the election approached.[18] The problem, as described by Taiwan FactCheck, is that technology capable of detecting deepfakes lags behind the technology used to create them.[19] During the tallying of election results, a widely circulated video showed an election worker incorrectly tallying a ballot, resulting in baseless claims that the election results were unreliable.[20]

Taiwan's pushback was immediate and multifaceted. Fact-checkers revealed the video had been heavily and misleadingly edited, and debunked rumors that arose from claims made in the video. The Central Election Commission held a press conference to set the record straight. Perhaps the most critical effort in debunking the fake tallying votes video was the aid of social media influencers—such as @FroggyChiu, with 600,000 YouTube subscribers—who showed their audiences how votes were being tallied.

Taiwan's significant efforts to combat China's MDM campaigns were successful. However, especially with AI advancements, these campaigns will continue to be difficult to detect and combat.

16 https://www.politico.eu/article/china-bombards-taiwan-with-fake-news-ahead-of-election/.
17 https://thehill.com/opinion/international/4406585-china-interfered-in-taiwans-election-and-failed-but-next-time-it-could-succeed/.
18 Id.
19 https://www.rfa.org/english/news/afcl/taiwan-china-disinformation-01102024224335.html.
20 Id.

The Importance of Directly Confronting MDM Campaigns

It is critical that we confront these challenges head-on. MDM will continue to be targeted at impacting elections in countries around the world. MDM experts have raised the possibility that China is using Taiwan as a testing ground for tactics in future information warfare, including in U.S. elections.[21]

The MDM problem is, of course, not just a "China problem." It is a much broader problem. As noted by the U.S. Election Infrastructure Government Coordinating Council and Subsector Coordinating Council's Joint Mis/Disinformation Working Group in guidance for state, local, tribal, and territorial election officials, and industry partners:

> Foreign actors have used MDM to target American voters for decades. MDM also may originate from domestic sources aiming to sow divisions and reduce national cohesion. Foreign and domestic actors can use MDM campaigns to cause anxiety, fear, and confusion. These actors are ultimately seeking to interfere with and undermine our democratic institutions. Even MDM that is not directly related to elections can have an impact on the election process, reducing voter confidence and trust…. False narratives erode trust and pose a threat to democratic transitions, especially, but not limited to, narratives around election processes and the validity of election outcomes.[22]

Recognizing the breadth of risk posed by MDM on global governments, businesses, and individuals, and the challenges faced by the lawyers representing governments, businesses, and individuals in confronting a host of MDM risks, the DRI Center for Law and Public Policy's Data Privacy and Security Working Group has chosen to focus its educational initiatives this year on MDM risks. Because of the critical importance of this election year, our initial focus is on MDM risks in the context of upcoming global elections.

In this white paper, we will first provide examples of how state and non-state actors have used MDM as a weapon to significant effect in elections around the globe, and especially in the context of global elections. Second, we will discuss MDM tactics, including how MDM campaigns are generated and spread. In this section, we also will address how advances in technology, and in particular AI, are affecting both the sophistication and amount of MDM campaigns we confront. Third, we will address actions taken by global governments to reduce risks of MDM campaigns. Fourth, we will address actions being taken by social media companies and other for-profit and non-profit entities to confront MDM campaigns. We will conclude by highlighting the importance of MDM awareness and by providing examples of actions individuals can take now to combat MDM risks.

---

21 https://www.gmfus.org/news/online-assault-against-taiwan.

22 https://www.cisa.gov/sites/default/files/publications/mdm-incident-response-guide_508.pdf (internal citations omitted).

EXAMPLES OF GLOBAL MDM CAMPAIGNS

2016 and 2020 U.S. Presidential Campaigns

In the 2016 U.S. presidential election, MDM was algorithmically amplified through social media platforms. According to a Buzzfeed analysis, during the critical, last three months of the 2016 election, the top 20 fake news stories from hoax sites and hyperpartisan blogs had 8.7 million Facebook shares, reactions, and comments, while the top 20 real news stories only received 7.4 million.[23] Per the Buzzfeed analysis, as the election got closer, fake content engagement on Facebook skyrocketed and surpassed real content from major news outlets. The top 20 false election stories generating the most shares, reactions, and comments included fake news that Hillary Clinton sold weapons to ISIS, and fake news that RuPaul alleged he was groped by Trump.[24]

As the Brookings Institution has noted, "The fact of Russian interference in the 2016 election is now well known in the United States."[25] Indeed, Russia's manipulation of technology, especially social media, is notorious. It was widely reported that Russian propagandists used social media to suppress votes for Hillary Clinton, highlighting the impact of data misuse and disinformation on the election.[26] During the campaign, several Russian disinformation campaigns were viewed in the press as having a particularly significant impact. One was the infamous "Pizzagate" campaign, which purportedly was the work of 12 Russian GRU agents.[27] A lot of these campaigns are believed to have originated with the Russian Internet Research Agency (IRA), which was purported to have implemented a coordinated disinformation campaign on social media.[28] The IRA, a Kremlin-linked troll farm, is believed to have begun interfering in earnest as early as 2014. Evidence of Russian disinformation operations started to surface in mid-2016.[29] The IRA purportedly used social media accounts to impersonate U.S. users and spread content that favored presidential candidate Donald J. Trump and disparaged presidential candidate Hillary Clinton.[30] The IRA is believed to have wanted to undermine public faith in the U.S. democratic process, denigrate Clinton, and ultimately influence the election in favor of Trump.[31] These activities raised widespread concerns about the impact of foreign interference and disinformation on the integrity of the election process.[32]

After years of investigations, a U.S. Senate Committee reported that, although it did not know for sure "what Moscow's intentions were, Russia may have been probing vulnerabilities in voting systems to exploit later. Alternatively, Moscow may have sought to undermine confidence in the 2016 U.S. elections simply by discovering their activity."

Additionally, during the 2016 campaign, major U.S. political campaigns used data on over 200 million voting-age Americans to inform their strategies, creating detailed profiles of voters to target them in various ways.[33] These practices raised concerns about privacy abuses, the spread of

---

23 https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.
24 Id.
25 Elain Kamarck, "Malevolent soft power, AI, and the Threat to Democracy," November 29, 2018, at https://www.brookings.edu/articles/malevolent-soft-power-ai-and-the-threat-to-democracy/.
26 https://www.brookings.edu/articles/data-misuse-and-disinformation-technology-and-the-2022-elections/.
27 https://faculty.lsu.edu/fakenews/elections/sixteen.php.
28 https://www.justice.gov/archives/sco/file/1373816/dl.
29 Id. at 1, 19.
30 Id. at 1, 4.
31 Id.
32 https://www.nature.com/articles/s41467-022-35576-9.
33 https://www.reuters.com/graphics/U.S.A-ELECTION/DATA-VISUAL/yxmvjjgojvr/.

disinformation, and the need for electoral integrity and public accountability.[34]

The 2020 election offered more microtargeting,[35] disinformation, and social media involvement. Political campaigns used data on more than 200 million voting-age Americans to create detailed voter profiles and target voters in variety of different ways.[36] For example, in the 2020 election, campaigns used social media to target voters of color with false and misleading information about not just candidates, but also about voting procedures. MDM campaigns were conducted by a variety of entities. Some targeted Biden, and some targeted Trump. For example, Facebook ads targeting Latino and Asian American voters described Joe Biden as a Communist; and doctored images showed dogs urinating on Trump campaign posters.[37]

None of these claims were true, but social media amplified them all—particularly on Twitter (now X), where users collaboratively constructed and amplified alleged evidence of fraud that was and is used to facilitate action, both online and off.

Another problematic trend emerging in the 2020 election was the sharing of manipulated videos. Although deepfake technology was nascent, observers noted increased instances in which videos were edited to make candidates appear to be making missteps that they didn't commit, slurring words, or appearing less competent.[38]

Leading up to the 2024 election, generative AI is already being used to create images, videos, and deepfakes targeting American voters with MDM.

Leading up to the 2024 election, generative AI is already being used to create images, videos, and deepfakes targeting American voters with MDM. In March 2023, AI-created images were circulated depicting Former President Trump running from the police, being roughly apprehended, and escorted to a police vehicle.[39] And in January 2024, an AI-generated robocall mimicking President Biden's voice commanded New Hampshire residents to stay home and not vote ahead of a New Hampshire presidential primary.[40]

2019 and 2021 Canadian Federal Election

Concerns about foreign interference in Canadian elections have existed for several years but rose to prominence in 2023 when proof of such attempts was made public. A public inquiry probing interference in the 2019 and 2021 federal elections was announced.[41]

According to a top-secret briefing report obtained by Global News, the Canadian Security Intelligence Service (CSIS) stated that China attempted to influence the 2019 and 2021 federal elections through clandestine and deceptive means.[42] The report, titled "Briefing to the Minister of Dem-

---

34 https://privacyinternational.org/advocacy/5158/technology-data-and-elections-updated-checklist-election-cycle.
35 Microtargeting is explained and addressed in more detail later in this white paper.
36 https://www.reuters.com/graphics/U.S.A-ELECTION/DATA-VISUAL/yxmvjjgojvr/.
37 https://www.politico.com/news/2023/07/29/election-disinformation-campaigns-targeted-voters-of-color-in-2020-experts-expect-2024-to-be-worse-00108866.
38 https://research.umd.edu/articles/social-medias-impact-2020-presidential-election-good-bad-and-ugly.
39 https://www.bbc.com/news/world-us-canada-65069316.
40 https://www.nbcnews.com/tech/misinformation/joe-biden-new-hampshire-robocall-fake-voice-deep-ai-primary-rcna135120.
41 Elizabeth Thompson, CBC News, "Inquiry into foreign interference to begin hearings in new year," November 2, 2023, at https://www.cbc.ca/news/politics/foreign-interference-inquiry-hogue-1.7016148.
42 https://globalnews.ca/news/10264872/canada-china-foreign-interference-elections-csis-report/.

ocratic Institutions on Foreign Interference," identified China as the most significant threat and named India a foreign interference threat, although details regarding India were redacted.[43] The Globe and Mail has reported on China's alleged interference strategies, including providing secret funding to candidates and targeting specific politicians.[44]

An independent inquiry into foreign electoral interference began its public hearings in January 2024, with Commissioner Marie-Josée Hogue pledging to uncover the truth about the extent of interference by foreign nations such as China, Russia, and India in the 2019 and 2021 elections.[45]

The Canadian Centre for Cybersecurity has reported increased cyber threat activity targeting elections worldwide. It assesses that such activity will likely occur in Canada's next federal election.[46] Despite these efforts, an independent review concluded that foreign governments did not succeed in impacting the voting results of the last two federal elections in Canada.

German 2024 #Oktoberfest Influence Operations and Russian Influence on 2017 Bundestag Election

In January 2024, investigators of the German Foreign Office (Auswärtiges Amt) discovered massive Russian operations on X (formerly "Twitter").[47] These operations aimed to increase a hostile climate against the current German government, to stir up the population concerning aid measures for Ukraine, and to start a pro-Russian disinformation campaign.[48] To spread such disinformation, pro-Russia-minded X users created fake accounts as doppelgangers of famous persons, for example, the Minister of the Foreign Office. From those accounts, the doppelgangers tweeted fake messages and linked fake websites from important German digital newspapers for "verification" purposes of those tweets. Indeed, it was quite easy to recognize that these accounts were fake and produced fake tweets. The tweets included telltale and unrelated hashtags such as "#Oktoberfest" but were about the war in Ukraine. However, even though the tweets were not high-quality fakes, they worked because of their mass. Over 50,000 fake accounts were created and about 1 million tweets were posted. Due to the number of fake accounts and tweets, the investigators suggested that the tweets were automated and sent with the help of AI. The German Foreign Office posted the results of their investigation on the website "EU vs. Disinfo" to fight the pro-Russian disinformation campaign.[49] Additionally, the European Commission has initiated measures under the Digital Services Act (see below) against X to tackle online disinformation and its societal risks.[50]

An analysis by the Securing Democracy Alliance,[51] a nonpartisan initiative housed at the German Marshall Fund of the United States, found significant influence from Russian operations in the run-up to the 2017 federal elections.[52] Troll accounts were used to defame Islam, call for the election of the far-right party "AfD," and defame Chancellor Merkel over her refugee politics. In addition to posts on social networks, hashtags were used to spread quickly, the most popular of which is #MerkelMussWeg ("MerkelMustGo").[53] In addition, Russian TV channels presented the

43 https://globalnews.ca/news/10264872/canada-china-foreign-interference-elections-csis-report/.
44 https://www.theglobeandmail.com/politics/article-chinese-election-interference-canada-timeline/.
45 https://www.cbc.ca/news/politics/foreign-interference-inquiry-china-russia-india-hogue-1.7094573.
46 https://www.cyber.gc.ca/en/guidance/cyber-threats-elections/.
47 "Trommelfeuer der Lügen" by Marcel Rosenbach and Christoph Schult, in "Der Spiegel" from January 27, 2024.
48 https://de.euronews.com/2023/06/15/thecube-doppelganger-gefalschte-medienseiten-im-netz-enttarnt.
49 https://euvsdisinfo.eu/.
50 https://www.lto.de/recht/nachrichten/n/eu-kommission-x-elon-musk-verfahren/.
51 https://securingdemocracy.gmfus.org/.
52 https://www.tagesschau.de/investigativ/russland-afd-einflussoperationen-101.html.
53 https://www.focus.de/politik/deutschland/bundestagswahl_2017/bundestagswahl-2017-pfiffe-und-merkel-muss-weg-rufe-wie-die-kanzler-

incidents in a highly distorted way, using the method of "fomenting polarization" to divide society with inaccurate reports to push voters towards certain political parties. Examples include reports on the Rammstein protests[54] and the "Lisa" case, in which the rape of a 13-year-old Russian girl by migrants was persistently reported contrary to the facts.[55]

French 2017 Presidential Election

In the run-up to the 2017 French presidential election, more than 20,000 emails related to Emmanuel Macron's campaign were leaked due to a hack.[56] The leaks attracted widespread media attention as news of the leak spread quickly across the internet, aided mainly by bots and spammers. The emails had been "fraudulently obtained," and forged documents had been mixed with real ones "to cause confusion and misinformation." Shortly after the far-right media exposed the leak, Macron's campaign manager said they had been monitoring the hacking attempts since February and set the hackers up to steal a carefully prepared cache of trivial and fake documents.

Dutch 2016 Advisory Referendum on the Association Agreement with Ukraine

On April 6, 2016, a consultative referendum was held in the Netherlands on adopting the Association Agreement between Ukraine and the European Union.[57] The referendum question was: "Are you for or against the adoption of the Association Agreement between the European Union and Ukraine?" Greenstijl, a popular right-wing newspaper with a high circulation, initiated the referendum. The newspaper published several articles about the referendum in which at least the following misleading statements were made:

- There is a civil war in Ukraine, partly caused by the EU.
- There are many ultra-nationalist and neo-Nazi groups in Ukraine.
- It has been alleged that many IS fighters from Georgia come to the EU via Ukraine.
- Ukrainian government officials are ultranationalists and fascists.
- The Maidan demonstrations were a consequence of the Association Agreement.
- The Netherlands will have to support corrupt Ukrainian banks.
- The shooting down of MH17, in which 193 Dutch citizens lost their lives, was a consequence of the Association Agreement and the "civil war."
- Putin annexed Crimea as a result of the Association Agreement. The civil war in eastern Ukraine is a consequence of the Association Agreement.
- The people in eastern Ukraine wanted to join Russia of their own free will.

Additionally, fake videos of the Ukrainian elite brigade Azov Battalion (now "Azov Assault Brigade") were circulated, which allegedly would carry out terrorist attacks in the Netherlands if the Netherlands voted against the Association Agreement.[58] Investigations revealed that the videos most likely originated from a troll factory in St. Petersburg.[59]

in-wut-und-hass-verkraftet_id_7534463.html.

54 https://www.tagesschau.de/investigativ/kontraste/proteste-ramstein-101.html.

55 https://www.bpb.de/themen/migration-integration/russlanddeutsche/271945/der-fall-lisa/.

56 https://www.zeit.de/politik/ausland/2017-05/praesidentschaftswahl-frankreich-hackerangriff-macron.

57 https://www.deutschlandfunk.de/assoziierungsabkommen-mit-der-ukraine-alles-spricht-fuer-100.html.

58 https://us11.campaign-archive.com/?u=cd23226ada1699a77000eb60b&id=dfea0d8d23.

59 https://www.bellingcat.com/news/uk-and-europe/2016/04/03/azov-video/.

TECHNOLOGIES AND TOOLS USED TO CREATE AND SPREAD DISINFORMATION, MISINFORMATION, AND MALINFORMATION

In this section, we address key technologies and tools used to create and spread MDM. Understanding information is crucial in today's world. In a steadily more digitalized world, personal data and technological practices are becoming increasingly important in all areas of life. This is true not only where technological advances are used to improve the quality of life and accelerate research and medical progress, but also where this technology is misused for personal and political gain. Since the Cambridge Analytica scandal in 2016,[60] it has become clear that technology is increasingly being used to influence democratic processes. With the rapid advancement of technology, especially the capabilities of AI, the risk of influence grows exponentially. Recently, it is rare to find an election without suspicion of (at least attempted) third-party influence. Due to the multitude of gateways for manipulation and the cloak of anonymity of the internet, it is not always possible to verify whether influence has been exerted and, more importantly, by whom.

Due to the multitude of gateways for manipulation and the cloak of anonymity of the internet, it is not always possible to verify whether influence has been exerted and, more importantly, by whom.

Definitions and Distinctions: Disinformation, Misinformation, and Malinformation

First, we want to further define and explain the differences between disinformation, misinformation, and malinformation. When separating truth from fact, there are three salient groups of bad information: disinformation, misinformation, and malinformation. Misinformation is inaccurate or misleading information unintentionally spread. Sharing an unverified news article about a scientific discovery without checking the source is an example of spreading misinformation. The intention behind the sharing of the information may be good (e.g., simply to share interesting information), but the lack of critical evaluation before sharing misinformation allows falsehoods to propagate.

Disinformation, on the other hand, involves the deliberate spreading of false or misleading information to deceive, manipulate, or harm. For example, bad actors often use fake social media accounts to influence elections through targeted propaganda. The intent is malicious and frequently aims, in pursuit of a specific agenda, to create chaos and undermine trust in institutions.

Malinformation differs from disinformation and misinformation in that it uses real information with harmful intent. Exposing someone's private information without their consent or framing a genuine news article in a manipulative way are examples of malinformation. The underlying truth adds legitimacy to the shared information, making malformation particularly effective in causing reputational or other harm.

[60] https://www.sueddeutsche.de/wirtschaft/cambridge-analytica-facebook-brittany-kaiser-1.4747594.

Distinguishing between these terms and determining intent is critical. Different responses will be required for each of these information types. Addressing disinformation demands vigilance and fact-checking. In contrast, combating misinformation requires understanding the media and how and where information is disseminated. Fighting malformation may focus on protecting privacy and battling online harassment.

The Role of Data Mining, Microtargeting, and Dark Ads in MDM Campaigns

Data Mining

Data mining allows patterns to emerge and valuable insights to be harnessed from (sometimes massive) datasets. Today, data mining is often where AI, machine learning, statistics, and database systems converge. This can unlock valuable information by transforming raw data into accessible information that can be further analyzed and used. Industries like finance, healthcare, retail, and marketing leverage this power to uncover trends, patterns, and relationships that manual analysis may miss or, at the very least, require much more time to conduct.

While the benefits are many, specific techniques—i.e., classification, clustering, and association rule learning—can be weaponized to spread disinformation or create malinformation tailored for harmful purposes such as exploiting biases or sowing discord.

In the hands of bad actors, data mining morphs from a tool into a weapon. Understanding this is crucial to harnessing the true potential of data mining while safeguarding ourselves from its pitfalls. Responsible data governance, ethical algorithms, and vigilant user awareness are essential safeguards to ensure that information is used productively and ethically.

Microtargeting

"Microtargeting" uses consumer data to identify specific audience segments and deliver tailored tailored messages to such segmented audiences. The consumer data is gathered from various sources such as social media platforms, cookies, social plugins, and tracking pixels to track online behavior and build a profile for each individual. Such profiles are used for different purposes, including for the purpose of targeted advertising. The profile is then used to predict the person's interests and intent, allowing advertisers to deliver personalized advertisements more likely to resonate with each individual. Microtargeting relies on predictive modeling and analytics to de-

termine the most appropriate content for each viewer, and it can be used to deliver particular and relevant marketing messages to individuals or small groups.[61]

Microtargeting has been around for a while. In 2012, a father angrily accused a Target manager of sending pregnancy products to his teenaged daughter without cause, only to later discover that Target's AI had figured out that his daughter was pregnant before she had told her family.[62] Barack Obama's 2012 campaign used web cookies and other data to reach interested voters, which had a beneficial effect.[63]

In 2016, microtargeting stepped up to the next level. Research regarding the 2016 election has shown that micro-targeted political ads on social media had significant effects based on geographical location, ideology, ethnicity, and gender. Exposure to these ads made individuals less likely to change their initial voting intentions, particularly among those who had expressed an intention to vote for Donald Trump. We also find that micro-targeted ads reduced turnout among targeted liberals, whereas they increased turnout and support for Trump among targeted moderates.[64] For instance, the United Kingdom's Channel 4 news reported that the Trump campaign microtargeted Black Americans with negative ads about Hillary Clinton to deter them from voting.[65]

Dark Ads

In the run-up to the 2021 German federal elections, the organization "Who Targets Me,"[66] in collaboration with journalists from the German public TV station "ZDF,"[67] analyzed a large amount of data to determine which Facebook users were being targeted with personalized advertising. Not only were all parties using microtargeting means to send these users tailored election ads, but they were also serving ads with sometimes

[61] See generally https://www.techtarget.com/searchcio/definition/microtargeting.
[62] https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&hp=&pagewanted=all.
[63] https://www.propublica.org/article/everything-we-know-so-far-about-obamas-big-data-operation; https://www.technologyreview.com/2012/12/19/114510/how-obamas-team-used-big-data-to-rally-voters/.
[64] Liberini, Federica and Redoano, Michela and Russo, Antonio and Cuevas, Ángel and Cuevas, Ruben, Politics in the Facebook Era - Evidence from the 2016 U.S. Presidential Elections (2020). CESifo Working Paper No. 8235, Available at SSRN: https://ssrn.com/abstract=3584086 or http://dx.doi.org/10.2139/ssrn.3584086.
[65] https://research.umd.edu/articles/social-medias-impact-2020-presidential-election-good-bad-and-ugly.
[66] https://whotargets.me/de/.
[67] https://noyb.eu/de/noyb-und-zdf-magazin-royal-suchen-target-leaks-datenspenderinnen.

contradictory content to different target groups. In this context, it was striking that the research revealed that the ads placed by the parties for the Bundestag election campaign were largely not to be found in the transparency database created by Facebook; they were missing from the ad library.

The green party ("Bündnis90/Die Grünen") also used dark ads.[68] Similar to microtargeting, dark ads are used to target voters. Campaign messages are tailored to selected target groups. From the totality of the data traces left on the internet, the companies create personality profiles of the possible addressees to obtain the most suitable target group for the election campaign.[69] In contrast to personalized advertising with microtargeting, dark ads are only shown to the selected target groups and are usually not published further. This lack of transparency makes it difficult to understand whether and to what extent dark ads are being used. The use of dark ads in the Green Party's 2017 election campaign is known, as the party published it. However, it can be assumed that many parties use this method.

How AI Deepfakes Are Used in MDM Campaigns

Days before a pivotal national election in Slovakia, an audio clip began circulating widely on social media. A voice that sounded like the country's Progressive party leader, Michal Šimečka, described a scheme to rig the vote, in part by bribing members of the country's marginalized Roma population. Two weeks later, a user posted on X (formerly Twitter), audio of the leader of the United Kingdom's Labour Party berating a staffer in a profanity-laden tirade.[70]

Last year, AI-generated media reached American politics in full force. A fake image of former president Donald Trump getting arrested went viral in 2023.[71] AI audio of Obama popped up on TikTok and other social media platforms.[72] During Chicago's February 2023 mayoral primary election, a deepfaked video surfaced of candidate Paul Vallas appearing to approve of police brutality. Vallas lost the race. Whether the deepfake ultimately affected the election, no one can say for sure.[73]

This year has brought more AI fakery. In January 2024, an AI-generated voice impersonating President Biden made robocalls to Democrats urging them not to vote in the New Hampshire primary.[74] Perhaps thinking "If you can't beat 'em, join 'em," presidential hopeful Dean Phillips created a ChatGPT version of himself that voters could interact with, and which OpenAI eventually removed.[75] The Democratic Party has experimented with AI-generated fundraising messages.[76] And both parties will likely use AI to enhance their microtargeting efforts.[77]

---

68 https://www.ndr.de/fernsehen/sendungen/zapp/medienpolitik/DarkAds-Der-geheime-Wahlkampf-im-Netz,darkads102.html.
69 https://www.br.de/nachrichten/bayern/landtagswahl-bayern-so-nutzen-parteien-social-media-im-wahlkampf,TIBrEIF.
70 https://www.washingtonpost.com/technology/2023/10/13/ai-voice-cloning-deepfakes/.
71 https://www.nytimes.com/2023/03/28/us/politics/artificial-intelligence-2024-campaigns.html?smid=nytcore-ios-share&referringSource=articleShare.
72 https://www.nytimes.com/2023/10/12/technology/tiktok-ai-generated-voices-disinformation.html.
73 https://www.cbsnews.com/chicago/news/vallas-campaign-deepfake-video/.
74 https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocall-f3469ceb6dd613079092287994663db5.
75 https://www.theguardian.com/technology/2024/jan/22/openai-bans-bot-impersonating-us-presidential-candidate-dean-phillips.
76 https://www.nytimes.com/2023/03/28/us/politics/artificial-intelligence-2024-campaigns.html?smid=nytcore-ios-share&referringSource=articleShare.
77 Id.

How Social Media Algorithms Assist in Spreading MDM

Social media algorithms weaponize human psychology by prioritizing content that is likely to generate high engagement from users; this often includes sensational, controversial, or emotionally charged content, including MDM (as the algorithms don't discern between accurate and inaccurate information).[78] The algorithms tailor content feeds to individual users' preferences and past interactions, creating echo chambers where users are repeatedly exposed to similar viewpoints, including potentially harmful MDM.

The algorithms are designed to promote content that is likely to generate strong reactions. This can lead to the amplification of extreme views, including misinformation and conspiracy theories, as these often provoke more engagement than moderate or nuanced content. The algorithms do not differentiate between credible information and MDM, as long as the content keeps users engaged.[79]

Social media algorithms exploit human psychological tendencies by amplifying types of information from which humans are biased to learn from, i.e. prestigious, in-group, moral, and emotional information (or "PRIME").

Social media algorithms exploit human psychological tendencies by amplifying types of information from which humans are biased to learn from, i.e. prestigious, in-group, moral, and emotional information (or "PRIME"). This exploitation can lead to the spread of MDM, as users are more likely to engage with and share content that aligns with their views or sparks strong emotional reactions (e.g. posts that provoke moral outrage) regardless of its veracity.[80]

How Fake Followers, Sock Puppets, and Troll Farms Are Used

Especially in social media, the proliferation of fake followers, sock puppets, and troll farms represents a significant threat to information integrity.

Fake Followers

Fake followers are artificial social media accounts, often created using bots, inflating the number of account followers. This can create the illusion of popularity or influence and spread dis, mis, and mal information or propaganda.

Team Jorge demonstrates the value of fake followers to creators of MDM. Team Jorge is a private Israeli task force specializing in cyber activities, including hacking, sabotage, and bot-farmed social media disinformation campaigns used to manipulate the results of elections. The group's activities were exposed in February 2023 after journalists conducted a secret undercover operation. The group has been active since at least 2015 and claims to have manipulated 33 presidential elections around the world, in 27 cases allegedly successfully. One of the company's most essential tools is the software program Advanced Impact Media Solutions, AIMS.[81] With AIMS, it is possible to create and automatically control thousands of fake social media profiles—some linked

[78] https://neurosciencenews.com/social-media-behavior-misinformation-23752/.
[79] https://www.fastcompany.com/90943919/the-science-behind-why-social-media-algorithms-warp-our-view-of-the-world.
[80] https://www.fastcompany.com/90943919/the-science-behind-why-social-media-algorithms-warp-our-view-of-the-world.
[81] https://www.theguardian.com/world/2023/feb/15/aims-software-avatars-team-jorge-disinformation-fake-profiles.

to Amazon, Airbnb, credit cards, or Bitcoin accounts. In addition to automatically controlling fake profiles, the team gets access to targeted accounts through direct/brute-force hacking. By taking over the accounts, the team can not only work with a created troll army but also take advantage of the popular impersonation method, making it easier to gain traction with social media targets with a leap of faith. On European soil, Team Jorge is said to have influenced the Catalan independence referendum in 2014, spread false news on French TV, and participated in the Cambridge Analytica scandal.

Sock Puppets

Similarly, sock puppet accounts, controlled by a single entity but masquerading as independent voices, manipulate public opinion by creating an illusion of consensus or support for specific agendas. These tactics amplify misleading messages and lend false credibility to viewpoints.

Troll Farms

Troll farms use both techniques on a vast scale. These operations employ numerous individuals to manage fake accounts, generate and disseminate disinformation, and engage in online harassment. Governments and political organizations may use troll farms to shift narratives, create discord, and influence decision-making, challenging the foundation of information integrity.

Research suggests that mobilization depends on the selective amplification of false or misleading tweets by influencers, the framing around those claims, and the perceived credibility of their source. These processes construct a self-reinforcing cycle wherein audiences collaborate to create a misleading version of reality, leading to offline actions that further reinforce a manufactured reality.[82]

[82] "Mobilizing manufactured reality: How participatory disinformation shaped deep stories to catalyze action during the 2020 U.S. presidential election," by Stephen Prochaska, Kayla Duskin, Zarine Kharazian, Carly Minow, Stephanie Blucker, Sylvie Venuto, Jevin D. West, Kate Starbird, published April 16 in the Proceedings of the ACM on Human-Computer Interaction (CSCW); https://dl.acm.org/doi/10.1145/3579616.

HOW GLOBAL GOVERNMENTS ARE COMBATTING MISINFORMATION, DISINFORMATION, AND MALINFORMATION

United States

As to what the United States federal government is doing to combat AI election disinformation, the answer is, regrettably, not enough.

Congress

The answer on the federal legislative front is even simpler: Congress has done little to address the issue of deceptive AI use in elections. Federal election law currently prohibits campaigns from "fraudulently" misrepresenting themselves "as speaking or writing . . . on behalf of any other candidate or political party . . . on a matter which is damaging" to that candidate or party.[83] This statute does not even purport to apply to political action committees or overzealous advocates.

Federal Election Commission

On July 13, 2023, the United States Federal Election Commission received a Petition[84] for Rulemaking ("Petition") from Public Citizen, a nonprofit advocacy organization. The Petition asked the Commission to amend its regulation on "fraudulent misrepresentation" to clarify that "the restrictions and penalties of the law and the Code of Regulations are applicable" should "candidates or their agents fraudulently misrepresent other candidates or political parties through deliberately false [Artificial Intelligence]-generated content in campaign ads or other communications." The Petition asserted that generative AI and deepfake technology, is being "used to create convincing images, audio and video hoaxes."[85] The Petition asserted that, although the technology cannot yet fool viewers, if the use of the "technology continues to improve, it will become increasingly

[83] https://www.ecfr.gov/current/title-11/chapter-I/subchapter-A/part-110/section-110.16.
[84] https://www.citizen.org/article/petition-for-rulemaking-to-clarify-that-the-law-against-fraudulent-misrepresentation-applies-to-deceptive-ai-campaign-communications/.
[85] Petition at 2.

difficult, and perhaps, nearly impossible for an average person to distinguish deepfake videos and audio clips from authentic media."[86] The FEC declined to issue a rule.

Department of Homeland Security

The Department of Homeland Security's (DHS) efforts to combat MDM are largely undertaken by the Cybersecurity & Infrastructure Security Agency (CISA). Initially formed as the National Protection and Programs Directorate in 2007, CISA's focus is on the protection of the U.S.'s critical infrastructure from both physical and cyber threats.[87]

In 2017, DHS officially designated election systems as critical infrastructure, bringing their protection under the purview of CISA.[88] As concerns regarding MDM assume particular importance in the context of elections, CISA established an MDM team and has taken a leading role in addressing MDM.[89] CISA is supported by another section of DHS, the Office of Intelligence and Analysis (I&A), which has a branch focused on creating analytical products about disinformation known as the Foreign Influence and Interference Branch.[90]

CISA's efforts to combat MDM focus largely on education and awareness of MDM campaigns. These efforts include the publications of a number of web-based graphic novels. The graphic novels are part of CISA's "Resilience Series" which "communicate the dangers and risks associated with dis- and misinformation through fictional stories that are inspired by real-world events."[91] These stories include "Real Fake," about an individual who is deceived into working at a troll farm run by Russia under the guise of a social media operation for a news organization, and "Bug Bytes," about a COVID disinformation campaign created with the goal to inspire physical attacks on 5G infrastructure. In a similarly light-hearted approach to raise awareness of MDM campaigns, CISA has also put out an infographic on "The War on Pineapple," explaining how foreign influence campaigns sow discord by amplifying divisive issues, like whether pineapple is a legitimate pizza topping,[92] with extreme language.[93] CISA's infographics also address new and emerging threats. These include infographics regarding social media bots[94] and deepfakes.[95] While CISA's efforts have focused mostly on MDM as a threat to trust and confidence in elections, CISA has also tried to increase awareness of MDM regarding COVID-19.[96]

CISA's election focus includes publishing an MDM "Planning and Incident Response Guide for Election Officials," the focus of which is preparing election officials to handle MDM related issues before, during, and after elections.[97] It has also put out issue-specific advice for election officials, including on "Generative A.I. and the 2024 Election Cycle," highlighting the need to prepare for more abundant and more sophisticated MDM campaigns, enabled by generative AI capabilities, during the 2024 election.[98] During the 2020 election cycle, CISA operated a "rumor control" site

---

[86] See generally 88 FR 55606.
[87] https://www.cisa.gov/news-events/alerts/2018/11/19/cybersecurity-and-infrastructure-security-agency.
[88] https://www.cisa.gov/sites/default/files/publications/19_0531_cisa_election-security-resources-guide-may-2019.pdf.
[89] https://www.cisa.gov/sites/default/files/publications/CSAC_MDM_September_2022_Final_Recommendations_09132022-508.pdf.
[90] https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation/resilience-series-graphic-novels.
[91] https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation/resilience-series-graphic-novels.
[92] It is.
[93] https://www.cisa.gov/sites/default/files/publications/19_1008_cisa_the-war-on-pineapple-understanding-foreign-interference-in-5-steps.pdf.
[94] https://www.cisa.gov/sites/default/files/2023-01/social_media_bots_infographic_set_508.pdf.
[95] https://www.cisa.gov/sites/default/files/publications/mdm-inauthentic-content-product-english_508.pdf.
[96] https://www.cisa.gov/sites/default/files/publications/CISAInsights-COVID-19_Disinformation_Activity_508.pdf.
[97] https://www.cisa.gov/sites/default/files/2022-11/mdm-incident-response-guide_508.pdf.
[98] https://www.cisa.gov/sites/default/files/2024-01/Consolidated_Risk_in_Focus_Gen_AI_ElectionsV2_508c.pdf.

dedicated to addressing common MDM narratives by:

- Preemptively debunking likely MDM attacks.
- Being fact-forward. CISA focused on providing short, easily understandable statements of truth before listing the MDM it was debunking, flipping the more common practice of listing what is being debunked first.
- Providing independent sources to bolster credibility.[99]

CISA plans to run similar rumor control sites for the 2024 and future elections.[100]

Department of State

In 2011, the Center for Strategic Counterterrorism Communications (CSCC) was established within the U.S. Department of State with the initial purpose of "supporting agencies in Government-wide public communications activities targeted against violent extremism and terrorist organizations."[101] The CSCC became the Global Engagement Center (GEC) in 2016, and, in 2017, its mission expanded to include addressing foreign state and non-state propaganda and disinformation.[102] The original focus on terrorism has now entirely disappeared, The GEC's current mission is to "direct, lead, synchronize, integrate, and coordinate U.S. Federal Government efforts to recognize, understand, expose, and counter foreign state and non-state propaganda and disinformation efforts aimed at undermining or influencing the policies, security, or stability of the United States, its allies, and partner nations."[103]

The GEC's public-facing efforts fall into the "exposé" category. The GEC regularly publishes reports regarding how various state actors are using MDM to influence the U.S. and its allies.[104] One such report, entitled "How the People's Republic of China Seeks to Reshape the Global Information Environment," explains the GEC's approach:

> The immediate goal of this report is to shed light on the tactics, techniques, and processes by which the PRC endeavors to use the information environment to its advantage. By publishing this report, we hope to inform the audiences targeted by the PRC and to empower

99 https://www.cisa.gov/sites/default/files/publications/rumor-control-startup-guide_508.pdf.
100 https://thehill.com/policy/cybersecurity/574491-cisa-to-continue-rumor-control-site-to-counter-election-disinformation/.
101 https://www.state.gov/about-us-global-engagement-center-2/.
102 Id.
103 Id.
104 https://www.state.gov/bureaus-archive/global-engagement-center/.

governments, civil society, academia, the press, the private sector, and publics around the world to more effectively collaborate in their efforts to protect the integrity of the information space.[105]

The GEC's focus is generally not on a country-level, but instead addresses specific, individual MDM operations, increasingly before they are well-developed.[106]

On January 18, 2024, the GEC published its "Framework to Counter Foreign State Information Manipulation," which "seeks to develop a common understanding of this threat and establish a common set of action areas from which the United States, with its allies and partners, can develop coordinated responses to foreign information manipulation and protect free and open societies."[107] In it, the GEC identified five key action areas: "(1) national strategies and policies; (2) governance structures and institutions; (3) human and technical capacity; (4) civil society, independent media, and academia; and (5) multilateral engagement."[108] Overall, the framework seeks to address what the U.S. and its allies can do to directly counter MDM. The framework acknowledges that "monitor-and-report" approaches are insufficient, and provides recommendations concerning what can be done to promote and support the private sector's capabilities to identify and counteract MDM campaigns.

Federal Law Enforcement and Intelligence

Although generally far less public facing than CISA or the GEC, law enforcement and intelligence agencies in the U.S. play crucial roles in monitoring, identifying, and combatting foreign MDM campaigns.

The Federal Bureau of Investigation (FBI) established the Foreign Influence Task Force (FITF) in 2017 to "identify and counteract malign foreign influence operations targeting the United States."[109] FITF focuses on investigations and operations, information and intelligence sharing, and private sector partnerships.[110] Through its partnerships in both the public and private sectors, FITF shares threat indicators in an effort to combat MDM generally and as targeted at elections.[111]

The National Security Agency (NSA) has directly, and in partnerships with other agencies and institutions, worked to combat MDM. One such partnership with two universities focuses on researching methods to detect adversary manipulation techniques, including the spread of MDM.[112] The NSA has also partnered with the FBI and CISA in publishing information regarding the threat of deepfakes and synthetic media generally as a tool to propagate MDM.[113] The report prepared by the NSA and other agencies includes identification of the problem, examples of malicious use of synthetic media, recommendations for addressing the threat posed by synthetic media-based MDM, and resources that can be consulted.

A lesser known and far more recent agency, the Foreign Malign Influence Center (FMIC), creat-

---

[105] https://www.state.gov/wp-content/uploads/2023/10/HOW-THE-PEOPLES-REPUBLIC-OF-CHINA-SEEKS-TO-RESHAPE-THE-GLOBAL-INFORMATION-ENVIRONMENT_508.pdf.
[106] https://www.nytimes.com/2023/10/26/technology/russian-disinformation-us-state-department-campaign.html.
[107] https://www.state.gov/the-framework-to-counter-foreign-state-information-manipulation/.
[108] Id.
[109] https://www.fbi.gov/investigate/counterintelligence/foreign-influence.
[110] Id.
[111] Id.
[112] https://www.nsa.gov/Research/Technology-Transfer-Program/Success-Stories/Article/3340208/nsa-and-minority-serving-institutions-detecting-adversary-misinformation-with-h/.
[113] https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF.

ed on September 23, 2022, was charged with "strengthening Intelligence Community efforts to counter the enduring threat posed by hostile foreign actors seeking to influence the U.S. Government, state and local governments, or public opinions through overt or covert means."[114] FMIC is the primary organization for "integrating intelligence analysis and reporting pertaining to foreign malign influence, including election security."[115] Essentially, the FMIC taps on the intelligence resources of the various government agencies to identify MDM and then acts as a central hub for that MDM intelligence.

States

In light of the dearth of federal government action, states are filling the vacuum. California, Texas, Washington, Minnesota, and Michigan have enacted laws regulating AI in elections. Washington, Minnesota, and Michigan require campaigns to disclose whether the campaign used AI in an ad's creation.[116] A number of other states are also considering similar AI-related legislation to combat voter MDM campaigns.[117] The Voting Rights Lab, a U.S. voting rights organization, recently reported that over 100 bills are pending in 40 state legislatures.[118] But this patchwork approach cannot fully address a national problem.

Canada

Federal Agencies

Elections Canada, the agency responsible for administering federal elections, works with security agencies like CSE and CSIS to monitor and investigate foreign interference. They have established protocols to inform the Prime Minister, political party officials, and the public if interference attempts are detected during an election.[119] In early 2024, Elections Canada released an online disinformation tool to assist voters with factual information, debunking common misconceptions and disinformation in the next federal election.

The Government of Canada has taken steps to counter foreign interference, including investigating and laying criminal charges, conducting national security reviews of foreign investments, and coordinating diplomatic responses with allies.[120] The Public Inquiry into Foreign Interference in Federal Electoral Processes and Democratic Institutions was established to examine the issue further and is expected to submit a final report by the end of 2024.[121]

For more detailed information on the government's response to foreign interference, the Public Safety Canada website provides resources and guidance.[122] The inquiry into foreign interference will begin hearings in 2024, focusing on whether foreign actors tried to influence the 2019 and 2021 federal elections and Canada's ability to detect and counter such interference.[123] The Communications Security Establishment has also released updates on cyber threats to Canada's democratic process.[124]

---

[114] https://www.dni.gov/index.php/nctc-who-we-are/organization/340-about/organization/foreign-malign-influence-center.
[115] Id.
[116] https://www.nytimes.com/2024/01/11/us/ai-election-ads-state-legislators.html.
[117] Id.
[118] https://votingrightslab.org/2024/03/26/new-state-legislative-efforts-to-stem-the-tide-of-ai-generated-election-disinformation/.
[119] https://www.elections.ca/content.aspx?dir=int&document=index&lang=e&section=vot.
[120] https://www.publicsafety.gc.ca/cnt/ntnl-scrt/frgn-ntrfrnc/fi-en.aspx.
[121] https://www.asiapacific.ca/publication/briefing-note-canadas-public-inquiry-foreign-interference.
[122] https://www.publicsafety.gc.ca/cnt/ntnl-scrt/frgn-ntrfrnc/pyfi-en.aspx; https://www.canada.ca/en/public-safety-canada/news/2023/11/foreign-interference-and-canada.html.
[123] https://www.cbc.ca/news/politics/foreign-interference-inquiry-hogue-1.7016148.
[124] https://www.canada.ca/en/communications-security/news/2023/12/cyber-threats-to-canadas-democratic-process-2023-update.html.

The Critical Election Incident Public Protocol is a government mechanism to alert the public in the event of an incident that threatens the integrity of an election.[125]

To prevent foreign interference in Canadian elections, the government and various agencies have implemented a comprehensive set of measures aimed at safeguarding the country's democratic processes. These efforts span legislative changes, enhanced security protocols, public awareness campaigns, and international cooperation. Here is a detailed overview of the key initiatives:

Legislative and Regulatory Measures

Modernization of Laws

The Canadian government has been urged to modernize its legal framework to better address the evolving threats of foreign interference. This includes updating the Canadian Security Intelligence Service Act and the Security of Information Act to ensure they are equipped to handle modern challenges.[126] The Elections Modernization Act, Bill C-76, introduced new campaign finance regulations to limit foreign influence in elections.[127] Most significantly, the federal government has introduced Bill C-70, An Act respecting countering foreign interference. Bill C-70 proposes changes to existing federal legislation (i.e. the Security of Information Act, the Canadian Security Intelligence Service Act, and the sabotage offence in the Criminal Code) intended to "detect, disrupt, and protect against foreign interference threats against all people in Canada, including members of diaspora, marginalized or otherwise vulnerable communities."[128]

Criminal Code Enhancements

Proposals have been made to amend the Criminal Code to create new offences that specifically address foreign interference, ensuring that activities detrimental to Canada's interests by foreign entities can be effectively prosecuted.[129] Bill C-70 would also see amendments to the Code's sabotage provisions to encompass activities related to election interference.[130]

Foreign Influence Transparency Registry

Public consultations have been launched to guide the creation of a Foreign Influence Transparency Registry. This registry aims to create registration requirements for individuals or entities undertaking non-transparent influence activities targeting Canada.[131]

Security and Intelligence Measures

The Security and Intelligence Threats to Elections (SITE) Task Force, established as part of the government's Plan to Protect Canada's Democracy, is composed of officials from key security agencies. It works to identify and prevent covert, clandestine, or criminal activities from influencing or interfering with the electoral process.[132]

The SITE Task Force provides enhanced monitoring and assessment of foreign interference

---

[125] https://www.publicsafety.gc.ca/cnt/trnsprnc/brfng-mtrls/prlmntry-bndrs/20230629/07-en.aspx.

[126] https://www.cigionline.org/articles/how-to-curb-foreign-interference-in-canadian-elections-here-are-five-fixes/; https://www.justice.gc.ca/eng/cons/fi-ie/pdf/Addressing_foreign_interference.pdf.

[127] https://tspace.library.utoronto.ca/bitstream/1807/128613/1/Combatting%20Foreign%20Election%20Interference.pdf.

[128] https://www.canada.ca/en/public-safety-canada/news/2024/05/government-introduces-legislation-to-counter-foreign-interference.html.

[129] https://www.justice.gc.ca/eng/cons/fi-ie/pdf/Addressing_foreign_interference.pdf.

[130] https://www.canada.ca/en/public-safety-canada/news/2024/05/government-introduces-legislation-to-counter-foreign-interference.html.

[131] https://www.canada.ca/en/public-safety-canada/news/2023/11/foreign-interference-and-canada.html; https://www.pm.gc.ca/en/news/news-releases/2023/03/06/taking-further-action-foreign-interference-and-strengthening.

[132] https://www.canada.ca/en/democratic-institutions/news/2024/01/government-of-canada-announces-measures-to-protect-durham-by-election-from-foreign-interference.html; https://www.canada.ca/en/democratic-institutions/news/2023/05/government-of-canada-announces-measures-to-protect-upcoming-by-elections-from-foreign-interference.html.

threats. This includes producing classified and unclassified reports on any attempts at foreign interference identified during elections.[133]

Additionally, a new position has been established within Public Safety Canada to coordinate efforts to combat foreign interference. This role involves developing and implementing strategies to protect Canada's democracy and national security.[134]

Public Awareness and Education

Elections Canada has undertaken initiatives to educate the public about the electoral process and the risks of foreign interference. This includes delivering comprehensive voter information campaigns and working with stakeholders to spread accurate voter information.[135]

Efforts are also being made to promote fact-checking services and encourage vigilance among the public. This aims to bolster the resilience of new Canadians and enhance their understanding of the electoral process.[136]

International Cooperation

Canada is also working closely with international partners to share intelligence and best practices for countering foreign interference. This includes cooperation within the Five Eyes alliance and other multilateral forums.[137]

European Union

Actions by the European Parliament

To combat disinformation, misinformation, and malinformation, the European Parliament created a special committee (ING2) to address those threats.[138] ING2 has released reports on the history and status of attempts to interfere with elections and proposing specific actions. The reports were published on March 9, 2022[139] (updated in May, 2023[140]), and June 1, 2023,[141] in the form of Resolutions.[142]

Members of the European Parliament joined the fight against disinformation and have called for a coordinated strategy to increase EU resistance to foreign interference and manipulation and to protect the 2024 EU elections.[143] The ING2 provided a report with further details about what such a strategy should include.[144] The European Parliament also drafted a legal framework in 2018 to address "fake news" and create a uniform EU policy.[145] In addition to the legal framework, the European Parliament and the Council adopted various regulations and directives to address the

---

[133] https://www.canada.ca/en/democratic-institutions/news/2024/01/government-of-canada-announces-measures-to-protect-durham-by-election-from-foreign-interference.html; https://www.canada.ca/en/democratic-institutions/news/2023/05/government-of-canada-announces-measures-to-protect-upcoming-by-elections-from-foreign-interference.html.
[134] https://www.pm.gc.ca/en/news/news-releases/2023/03/06/taking-further-action-foreign-interference-and-strengthening.
[135] https://www.elections.ca/content.aspx?dir=int&document=fint&lang=e&section=vot.
[136] https://www.elections.ca/content.aspx?dir=int&document=fint&lang=e&section=vot.
[137] https://www.canada.ca/en/democratic-institutions/services/reports/countering-evolving-threat.html.
[138] https://www.europarl.europa.eu/committees/de/ing2/home/highlights.
[139] https://www.europarl.europa.eu/doceo/document/TA-9-2022-0064_EN.html.
[140] https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/747908/EPRS_ATA(2023)747908_EN.pdf.
[141] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0219_EN.html.
[142] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0219_EN.html.
[143] https://www.europarl.europa.eu/news/en/press-room/20230524IPR91908/foreign-interference-meps-call-for-urgent-protection-of-2024-european-elections.
[144] https://www.europarl.europa.eu/news/en/press-room/20230601IPR93601/ing2-report-takeaways.
[145] https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA(2018)619013_EN.pdf.

disinformation problems from multiple directions. Some of those acts include the following:

The Directive on Measures for a High Common Level of Cybersecurity Across the Union (NIS 2)

NIS 2[146] is a follow-up regulation to its predecessor, NIS 1, and was introduced as part of the "Program to Ensure Regulatory Efficiency and Performance." NIS 2 addresses many of the criticisms of NIS 1 and is much broader in scope. In particular, NIS 2 now includes more sectors, such as public administration. In addition, educational institutions may also be classified as critical institutions, under certain circumstances, provided they carry out essential research activities. Depending on the criticality of the sector, a distinction is now made between 'essential' and 'important' institutions. These include all institutions that employ more than 50 people or have an annual turnover or balance sheet total of more than €10 million. Irrespective of whether the threshold is met, public administration bodies involved in defense, national security, public safety, or law enforcement are explicitly excluded.

NIS 2 imposes comprehensive measures and obligations, including technical, operational, and organizational requirements. The measures aim to identify and reduce risks to the security of the network and IT systems. The aim is to reduce the impact of security incidents on third parties. Examples of measures include:

- The creation of risk management plans
- Measures to manage security incidents,
- Measures to ensure the security of the supply chain,
- Measures to maintain business operations,
- Crisis management measures, and
- Holding company directors accountable

Digital Services Act and Member State Laws

The Digital Services Act (DSA)[147] is the cornerstone of the European Union's digital security architecture. Its purpose is to define and harmonize the rules governing providers' activities of various digital services in the EU. The DSA also imposes obligations on platforms that act as intermediaries and connect consumers to goods, services, and content. The obligations are graduated according to the size of the companies—the most stringent obligations apply to massive online platforms (VLOP) and enormous online search engines (VLOSE) with more than 45 million average monthly active users. The primary obligations arising from the DSA are to create a safe digital environment free of illegal content, to improve transparency and accountability of digital interme-

[146] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555&qid=1707423907665.
[147] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065&from=DE.

diary services and to improve and strengthen the protection of European fundamental and consumer rights, to facilitate and promote competition and innovation in the European digital single market and to enhance enforcement of the obligations imposed.

A key element of the DSA, particularly to protect against interference in European elections, is a stricter approach to illegal content. Whether content is illegal is not determined by the DSA itself, but by the applicable law of each EU Member State (Art. 3 lit. h). Many European countries have already adapted their laws to combat disinformation campaigns. In Malta, for example, it is a criminal offense to maliciously disseminate false information likely to alarm public opinion, disturb public order or peace, or incite the population or certain sections of the population (Art. 9 I Press Act). Similar legislation exists in Greece, France, Romania, the Czech Republic, Croatia and several other European countries.

Providers are not only obliged in their own interest to quickly identify and remove illegal content; they can now also be informed of such content by courts or authorities and must then take the necessary measures (Art. 9, 10 DSA). Predefined reporting and removal procedures for such content must be followed and, if and as required, remedied.

Finally, a rapid response mechanism for the very large online platforms (VLOPs) and very large online search engines ("VLOSEs" has been introduced as a last resort for particularly severe crises. In the event of an exceptional crisis—a serious threat to public security or health in the EU—the European Commission may require service providers to cooperate and take defensive measures, such as adapting content moderation measures. The measures could be as drastic as those taken by the EU in 2022 when the rising tide of disinformation accompanying the war of aggression on Ukraine was followed by a comprehensive ban on the Russian state media Russia Today[148] and Sputnik, the Russian state-funded international news agency and radio network, for the entire EU.

EU AI Act

On April 21, 2023, the European Commission published the first draft of a regulation on the use of artificial intelligence ("EU AI Act"),[149] which aims to increase society's trust in AI systems without blocking the potential of this technology. The EU AI Act sets out harmonized rules for developing, marketing, and using AI systems in the European Union. The EU AI Act was adopted by the EU Parliament on March 13, 2024, and approved by the EU Council on May 21, 2024. The EU AI Act entered into force on August 1, 2024. It will generally be effective beginning August 2, 2026, with the exception of certain provisions. For example, the prohibitions on unacceptable risk AI will apply beginning February 2, 2025.

Because of these increasingly far-reaching attempts to influence elections in the European area, the prohibition of behavioral manipulation in the EU AI Act is particularly relevant. The EU AU Act prohibits AI systems that deploy subliminal techniques to influence a person unconsciously or purposefully use manipulative or deceptive techniques to materially distort people's behavior by impairing their ability to make informed decisions, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person, or a group of persons significant harm;[150] as well as AI systems that exploit weaknesses or vulnerabilities of a particular group of persons due to their age or a particular social or economic situation, with the purpose or effect of substantially influencing the behavior of persons in a way that may cause them or others significant harm.[151]

[148] https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rtrussia-today-and-sputnik-s-broadcasting-in-the-eu/.
[149] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.
[150] EU AI Act, Article 5(1)(a).
[151] EU AI Act, Article 5(1)(c).]

In light of the EU AI Act's broad definition of an "AI system,"[152] a prohibited AI system could include a variety of social networks. Thus, these provisions could be used as tools to combat disinformation campaigns in social networks.

Regulation on Transparency and Targeting of Political Advertising

To combat political influence through microtargeting (dark ads), the European Union is currently planning the so-called Regulation on Transparency and Targeting of Political Advertising.[153] The regulation aims not to ban targeted advertising but to ensure that specific user data cannot be used and to make the process more transparent.

The primary tool to achieve this is the transparency label. Paid political advertising would be required to be clearly labeled. This label must include a range of information, such as the donor's name (politician or organization), the amount spent, and the ad's link to the election. The information must be prominently displayed and easily visible. In addition, the use of particularly sensitive data for advertisement placement should be prohibited. These include ethnic origin, religion, and sexual orientation.

National data protection authorities will enforce the regulation. To this end, they should be given the power to impose substantial fines for infringements.

Cybersecurity Enforcement and Monitoring Institutions

To ensure that the planned European law will be enforced and monitored regularly, national institutions and the European Union Agency for Cybersecurity (ENISA)[154] will be the competent authorities.

The Member States of the European Union have their national security agencies responsible for cybersecurity. Although these are national responsibilities, there is a lively exchange between the institutions—coordinated by the European Union Agency for Cybersecurity. The Federal Office for Information Security (BSI)[155] has been established in Germany. In France, there is the General Secretariat for Defence and National Security (SGDSN); in Austria, it is called the National Security Authority (ISK).

In Germany, the BSI's primary goal as the federal government's cybersecurity authority is to shape information security in digitalization through prevention, detection, and response for the state, the economy, and society. As a result of the increasingly far-reaching threats to cybersecurity, the BSI has been continuously expanded in terms of personnel and infrastructure. Similarly, its rights and powers have been and continue to be extended through new legislation such as the IT Security Act, the DSA, and the AI Resolution. In addition, the BSI is the head of the National Cyber Defense Centre, which is a cooperation of several authorities and was established in 2011 to defend against cyber-attacks on the IT infrastructures of the Federal Republic of Germany.

At the European level, a common authority for cybersecurity, the European Union Agency for Cybersecurity (ENISA), was established in 2005. The Agency aims to achieve the highest possible level of cybersecurity in all relevant areas within the EU. It has been given increasingly broad powers through various pieces of legislation (see above). Its tasks include improving the trustworthiness of ICT products, services, and processes through cybersecurity certification schemes, working with Member States and EU institutions, and helping Europe and its Member States prepare

[152] See EU AI Act, Article 3(1).
[153] https://eur-lex.europa.eu/resource.html?uri=cellar:9cec62db-4dcb-11ec-91ac-01aa75ed71a1.0001.02/DOC_1&format=PDF.
[154] https://www.enisa.europa.eu/.
[155] https://www.bsi.bund.de/DE/Themen/KRITIS-und-regulierte-Unternehmen/Kritische-Infrastrukturen/KRITIS-aktuell/kritis-aktuell_node.html.

for tomorrow's cyber challenges. Through knowledge sharing, capacity building, and awareness raising, the Agency works with its key stakeholders to build trust and confidence in the networked economy, increase the Union's infrastructure's resilience, and ensure the digital security of European society and citizens.

Latin American Countries

Fact-Checking Agencies

In June 2019, Mexican President Andres Manuel Lopez Obrador started Verificado, a fact-checking organization that operates in conjunction with government-operated newswire Notimex. However, Verificado seems to have engaged in negligible activity; despite this, it has aroused concerns over politicization and censorship.[156] Similar criticisms have been made regarding Argentina's Observatory of Disinformation and Symbolic Violence, instituted in October 2020.[157]

Legislation

Multiple South American companies are considering or have enacted legislation intended to combat disinformation. One such effort is Brazil's Internet Freedom, Responsibility, and Transparency Bill, which seeks to hold social media companies accountable for combating disinformation.[158] In Uruguay, the congress signed an Ethical Pact Against Disinformation, in which the country's political parties promised not to share disinformation or promote false narratives on social media.[159]

Collaboration with Social Media Companies

Latin American governments are choosing to work directly with social media companies to limit the circulation of disinformation. Mexico and Argentina have worked with Facebook and other social media platforms in recent years in these efforts.

156 https://americasquarterly.org/article/how-latin-american-governments-are-fighting-fake-news/.
157 https://americasquarterly.org/article/how-latin-american-governments-are-fighting-fake-news/.
158 https://www.csis.org/analysis/ensuring-information-integrity-electoral-processes-americas.
159 https://www.csis.org/analysis/ensuring-information-integrity-electoral-processes-americas.

ACTIONS TAKEN BY THE PRIVATE SECTOR TO COMBAT DISINFORMATION, MISINFORMATION, AND MALINFORMATION RISKS

Governments are often hobbled by their inability to react quickly or nimbly in addressing MDM risks. Additionally, many governments, including the U.S. Congress, are divided on how to implement legislation that may meaningfully tackle AI and other MDM risks. Even if Congress were to pass legislation, such legislation would almost certainly raise First Amendment concerns. Although chatbots, as non-humans, do not have First Amendment rights, the First Amendment does implicate the rights of people to repost texts generated by a chatbot. And despite the efforts that are being undertaken by governments, challenges do and will remain in effectively countering foreign interference and other MDM risks. The complexity of the threat landscape and the covert nature of many MDM activities make detection and response difficult.

The complexity of the threat landscape and the covert nature of many MDM activities make detection and response difficult.

Ongoing efforts must continue to focus on enhancing the legal framework, improving interagency coordination, and raising public awareness about the risks of foreign interference.

As noted at the beginning of our white paper, the World Economic Forum 2024 Global Risks Report identified disinformation and misinformation as the most severe short-term risk the world faces in 2024. The Report warned: "Synthetic content will manipulate individuals, damage economies and fracture societies in numerous ways over the next two years. Falsified information could be deployed in pursuit of diverse goals, from climate activism to conflict escalation."[160] The combination of the high level of risk and governmental inability to timely, sufficiently regulate this risk, make it critical for news organizations, private sector companies and law firms, lawyers, educators, and society at large to step up and take action.

Private Industry Actions

Private industry is paying attention to the problem of AI misinformation. For example, on January 15, 2024, OpenAI, the creator of ChatGPT, issued a blog post entitled, "How OpenAI is Approaching 2024 Worldwide Elections." In that post, OpenAI reaffirmed that it would not allow political campaigns to create chatbots with its technology. This section sets forth additional actions that should be implemented (or further implemented) by the private sector to combat the serious MDM risks we are facing in the 2024 elections and beyond.

AI companies have explored ways to keep deceptive AI from being created or disseminated, but these methods are not yet perfected. One AI creator, ElevenLabs, designed software that successfully detected AI audio—until music was added. Open AI has announced that it was "experimenting with a provenance classifier, a new tool for detecting images generated by its image generator, DALL·E."[161]

[160] World Economic Forum 2024 Global Risks Report at 18, available at https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf.
[161] https://openai.com/dall-e-3.

More promisingly, other AI companies are considering adding a "watermark" to content that their software generates, allowing easy detection of fake AI images.[162] On October 30, 2023, President Biden issued an "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."[163] Among other things, the Order sought to develop guidance regarding AI watermarking—that is, embedding a recognizable, unique signal into an AI text or image that identifies that content as AI generated.[164] The Order directed the Department of Commerce to develop guidance for content authentication and watermarking of AI-generated content.[165]

Private industry appears to broadly support adopting AI watermarking. Adobe—which, through the notorious use of its Photoshop product, has long sat in the epicenter of falsified images—supports AI-content labeling, as do OpenAI and Microsoft.[166] Microsoft, Adobe, Intel, and other companies are members of the Coalition for Content Provenance and Authenticity ("C2PA"), which "addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content."[167] And on February 6, 2024, Facebook's owner, Meta, proposed to promote watermarking for photo, video and audio material that would signal that the content was generated using artificial intelligence.[168]

Among major social media platforms, Chinese-owned TikTok has—perhaps surprisingly, taken an early lead on watermarking. TikTok asserts that it requires users to be highly transparent when AI tools and effects are used within their content, and it encourages users to be comfortable with sharing the role that AI plays in altering their videos or photos in fun ways.[169] Specifically, TikTok shows viewers the specific AI effects that its content creators use.[170] Further, TikTok's policy provides: "Synthetic or manipulated media that shows realistic scenes must be disclosed." The policy prohibits and it prohibits "synthetic media of public figures if the content is used for endorsements or violates any other policy."[171] This prohibition includes "[m]aterial that has been edited, spliced, or combined (such as video and audio) in a way that may mislead a person about real-world events."[172]

However, experts caution that watermarking is not a panacea. OpenAI's watermarking is easily cropped, for example.[173] Google uses a more sophisticated technique that inserts a watermark imperceptibly into the pixels of an image, but Google warns that this technology "isn't foolproof."[174]

Occasionally, watermarks issue "false positives," whereby human-created content is identified as AI-generated content.[175] And bad actors can remove watermarks with relative ease.[176] Faced with these obstacles, one AI researcher helpfully explained that, with respect to AI watermarking that could resist removal, "There's no hope."[177]

---

162 https://www.nytimes.com/2023/10/12/technology/tiktok-ai-generated-voices-disinformation.html.
163 https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.
164 https://www.techtarget.com/searchenterpriseai/definition/AI-watermarking.
165 October 30, 2023 Executive Order § 4.5(c).
166 https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/.
167 https://c2pa.org/.
168 https://www.nytimes.com/2024/02/06/technology/meta-ai-standards-labels.html.
169 https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/.
170 Id.
171 https://www.tiktok.com/community-guidelines/en/integrity-authenticity/#3.
172 Id.
173 https://www.theverge.com/2023/10/31/23940626/artificial-intelligence-ai-digital-watermarks-biden-executive-order.
174 Id.
175 https://www.techtarget.com/searchenterpriseai/definition/AI-watermarking.
176 https://www.wired.com/story/artificial-intelligence-watermarking-issues/.
177 Id.

Strengthening public–private partnerships will prove critical in fighting MDM.

Strengthening public–private partnerships will prove critical in fighting MDM. Governments should strengthen their public–private partnerships with social media companies and companies that develop messaging apps. These companies have access to the data that would allow for more real-time analysis of MDM efforts and are also uniquely able to limit their effect by throttling the reach of keywords, trending topics, and links. Governments should increase their efforts to educate these companies on detection methodology and current trends and keep them engaged as partners in addressing combatting MDM before it spreads.

Governments also should engage open-source intelligence communities and allow them to assume some of the work in analyzing and publishing information about emerging MDM threats. To the extent governments are able to make public more raw data, rather than merely disclosing fully analyzed and reviewed reports, open-source intelligence communities could take on some of the work of identifying and publicizing MDM efforts. Additionally, to the extent that MDM has particular effectiveness against those already distrusting of government, open-source communities provide a potentially more friendly and credible source for those individuals.

At minimum, governments should engage in "no-strings-attached," one-way information sharing to the private sector. Governments can also encourage private-to-private data sharing among social media and messaging companies to allow for trends first observed on one platform to be monitored by another before MDM spreads.

Actions by Educators

Educators, while not having the same power to implement policy or develop technology, serve an important role in combatting MDM. And given a 2021 survey finding that 84 percent of 18- to 24-year-olds are not sure that they can distinguish true from false content on social media, educators may in fact have the most important role of all.[178] As government efforts make clear, inoculation is one of the strongest weapons against MDM, and inoculation is the result of understanding how to evaluate a source and how to research an issue. To fulfill this role, educators should:

- Prioritize media literacy. As more and more children and teenagers get their information from social media sites rather than traditional news media, it is more important than ever for educators to teach students how to determine the credibility of a source.
- Teach students research skills aimed at them. Although students should learn how to do primary source research at some point in their educational career, educators should understand that students are most likely to turn to Google or social media to investigate whether something they hear or see is true or not. Accepting that and teaching students how best to engage with those tools to truly investigate the veracity of something they encounter online will be far more useful to them on a daily basis.
- Challenge students to find MDM in their social media feeds or in videos they see. Tasking students with identifying MDM and sharing what they learned will both allow them to put in practice the topics discussed above but will also make them more keenly aware of the prevalence of MDM both through their own efforts to identify MDM and through the MDM their classmates identify.

As social media continues to be a crucial part of young people's lives, ensuring that students are able to discern what is true and what is not must be part of every school's curriculum—at every

[178] https://www.scientificamerican.com/article/young-people-tell-us-they-need-help-identifying-misinformation/.

level of education.

Individual Actions

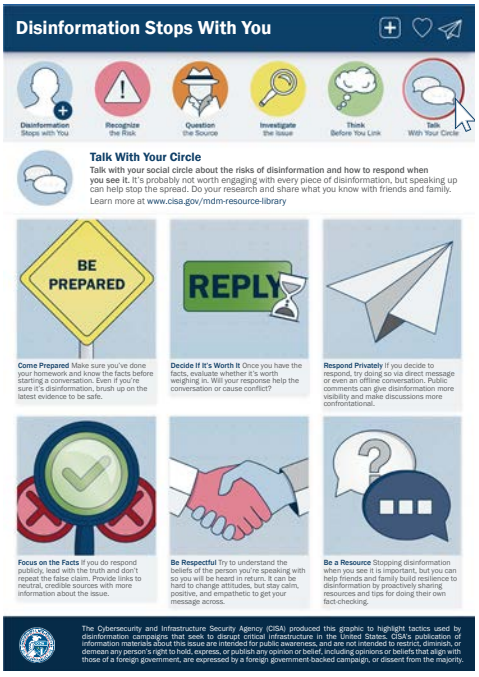
The consequences of MDM campaigns are far-reaching. Erosion of trust in institutions, social division, and even violence can arise from propagation of these information types. Cultivating critical thinking skills, verifying information before sharing, and being mindful of the potential harm of our online actions are essential steps in navigating this complex information landscape.

Cultivating critical thinking skills, verifying information before sharing, and being mindful of the potential harm of our online actions are essential steps in navigating this complex information landscape.

The Canadian Centre for Cybersecurity suggests asking the following questions to identify MDM:

- Does it provoke an emotional response?
- Does it make a bold statement on a controversial issue?
- Is it an extraordinary claim?
- Does it contain clickbait?
- Does it have topical information that is within context?
- Does it use small pieces of valid information that are exaggerated or distorted?
- Has it spread virally on unvetted or loosely vetted platforms?

The Canadian Centre for Cybersecurity recommends that individuals take the following actions to investigate content and help protect themselves from MDM:



- Look for out of place design elements such as unprofessional logos, colors, spacing, and animated gifs;
- Verify domain names to ensure they match the organization. The domain name may have typos or use a different Top Level Domain (TLD) such as .net or .org;
- Check that the organization has contact information listed, a physical address, and an 'About Us' page;
- Perform a WHOIS lookup on the domain to see who owns it and verify that it belongs to a trustworthy organization. WHOIS is a database of domain names and has details about the owner of the domain, when the domain was registered, and when it expires;
- Conduct a reverse image search to ensure images are not copied from a legitimate website or organization;
- Use a fact-checking site to ensure the information you are reading has not already been proven false;
- Do not automatically assume information you receive is correct, even if it comes from a valid source (such as a friend or family member); and
- Ensure the information is not out of date.

CONCLUSION: HOW WE CAN ADDRESS INCREASING RISKS FROM EMERGING TECHNOLOGIES–AND WHY WE MUST MOVE FAST

Emerging AI technologies are rapidly reshaping our world and opening up a host of wonderful possibilities for advancing society. Of course, emerging AI technologies can also be used to cause a host of MDM-related harms, including harms caused by cyber threat actors using MDM to steal money from businesses or individuals; by litigants introducing deep fake evidence during trial; and by adverse nation states and bad actors spreading MDM in efforts to influence elections and threaten the rule of law.

As evidenced by the MDM campaigns covered in the body of this white paper, MDM has posed a significant risk to global elections for some time. But these risks have been significantly amplified because of recent AI advances, which have made it possible for a host of individuals with malicious intent to "produce vast volumes of text peppered with falsehoods; generate convincing deceptive images, video, and audio; and distort public figures' words and actions on a previously unseen scale."[179] These threats will only grow.

It is incumbent upon us, as lawyers, to uphold the rule of law. We should stay on top of evolving election-related MDM threats, learn to identify MDM, participate in efforts to combat election-related MDM, and otherwise do our part to uphold the rule of law in a nonpartisan nature. Lawyers should stay on top of evolving MDM risks in the context of cyber threats to our firms and our clients' businesses and in the context of deep fake and purported deepfake evidence.

Lawyers should stay on top of evolving MDM risks in the context of cyber threats to our firms and our clients' businesses and in the context of deep fake and purported deepfake evidence.

Evolving MDM threats merit a whole-of-society response. Lawyers can, and should, play a leading role in mitigating MDM risks. In furtherance of assisting DRI members in addressing MDM risks, the DRI Center for Law and Public Policy's Data Privacy and Security Working Group will continue to provide white papers and other educational offerings addressing MDM risks and risk-reduction opportunities in areas we anticipate will be of interest to DRI members.

[179] https://www.brennancenter.org/our-work/policy-solutions/securing-2024-election.