FREEDOM OF THOUGHT AND FAKE NEWS: INTERNATIONAL HUMAN RIGHTS IN THE AGE OF SOCIAL MEDIA

Allison Mitton*

Abstract: Disinformation and malinformation are spread quickly and easily on social media because anyone can post content, deepfakes and other AI-generated content are easily created, and algorithms and bots amplify this material. As a result, false information thrives with little accountability for bad actors. This raises a critical issue: using technology as a tool to manipulate how people think seems to violate the right to freedom of thought. So why is this right so often infringed upon digitally, and how can this be prevented? Although scholars have attempted to interpret the right to freedom of thought and countries have implemented their own protections, these measures fall short, as the ease with which these actions can cross international boundaries allows foreign actors to continue infringing on this right. I suggest that the underlying issue lies in the absence of treaties, conventions, or principles of customary law that link the spread of disinformation or malinformation to a human rights violation—an infringement on freedom of thought. I argue for the adoption of a Third Optional Protocol to the International Covenant on Civil and Political Rights that (1) explicitly recognizes the use of technology to manipulate thoughts as a human rights violation, (2) outlines guidelines to prevent such infringements, and (3) establishes a better way to hold states accountable. By implementing this Protocol, meaningful progress can be made toward safeguarding the right to freedom of thought as technology evolves.

Keywords: Freedom of Thought; Human Rights; International; Artificial Intelligence; Technology

* BYU Law School, US.

Table of Contents

Introduction 54

I. Background Information 55

A. Definitions and Key Terms 55

1. Mis-, Dis-, and Malinformation 55

2. Social Media as a News Source 57

3. Deepfakes 57

4. Social Media Algorithms and Bots 58

B. How Countries Use Technology to Influence Thought 59

1. Spreading Disinformation 59

2. Leaking Malinformation 61

3. Effects of Interference 62

C. Freedom of Thought 63

1. What is Freedom of Thought? 63

2. Issues with Freedom of Thought 64

3. Interpretation of Freedom of Thought 64

II. Relevant Treaties, Conventions, and Customary Law 65

A. State Sovereignty and the Non-Intervention Principle 65

1. Sources Prohibiting Intervention 66

2. Application to Technology Use 66

B. United Nations Cyber Norms 67

1. Stability and Security in Cyberspace 67

2. Respect Human Rights Online 68

C. Issues with Current Law 68

III. Proposed Solution 68

A. Process for Adopting Optional Protocols 69

B. Creating a Third Optional Protocol 69

1. Preamble 70

2. Article 1: Scope of Application 70

3. Article 2: Restrictions and Requirements 70

4. Article 3: Remedies and Enforcement Mechanisms 71

C. Policy Considerations 72

Conclusion 72

INTRODUCTION

In 2022, a video surfaced on social media and a Ukrainian news website depicting Volodymyr Zelenskiy, the country's president, "call[ing] on his soldiers to lay down their weapons and return to their families."[1] However, the video was a deepfake—it was created with generative artificial intelligence (AI).[2]

Thankfully, viewers quickly caught on to the fact that the video was fake and "immediately flagged the discrepancies between the skin tone on Zelenskiy's neck and face, the odd accent in the video, and the pixelation around his head."[3] Zelenskiy himself was also able to quickly rebut the deepfake by posting a real video of himself on his Instagram account, dismissing the video as a "childish provocation."[4]

People speculate that Russia was behind this,[5] but because tracing the origins of false content on social media is often difficult, identifying the responsible party can be challenging.[6] Although the Zelenskiy video did not appear to have any serious consequences for either party involved, it highlights a larger issue: an increasing loss of trust in social media content and news in general worldwide.[7] Because spreading false information is so rarely traceable back to them, foreign actors have increasingly taken advantage of the fact that fake news is so easy to spread and that spreading it has little to no repercussions.[8]

As one scholar notes, "[d]eliberately trying to change what people perceive, think, feel and...do appears to be at tension with the human right to form thoughts freely and make decisions based on those thoughts."[9] Although the right to freedom of thought is very broadly defined with little to no official clarification on its exact meaning,[10] one widely recognized aspect of freedom of thought is that people should have "[f]reedom from impermissible alteration of [their] thoughts."[11] When there is a

---

[1] Dan Milmo & Pjotr Sauer, *Deepfakes v Pre-Bunking: Is Russia Losing the Infowar?*, THE GUARDIAN (Mar. 19, 2022, 2:00 AM), https://www.theguardian.com/world/2022/mar/19/russia-ukraine-infowar-deepfakes.

[2] *Id.*

[3] James Pearson & Natalia Zinets, *Deepfake Footage Purports to Show Ukrainian President Capitulating*, REUTERS (Mar. 17, 2022, 6:16 AM), https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16.

[4] Milmo & Sauer, *supra* note 1.

[5] *Id.*

[6] *See* Maria Pawelec, *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*, 1 DIGIT. SOC'Y 1, 25 (2022).

[7] Simon Montlake, *Trust in the Media Has Tanked. Are We Entering a 'Post-News' Era?*, THE CHRISTIAN SCI. MONITOR (May 3, 2024, 3:25 PM), https://www.csmonitor.com/USA/Politics/2024/0503/news-media-decline-polarization-distrust.

[8] *See* Pawelec, *supra* note 6.

[9] Richard Mackenzie-Gray Scott, *Managing Misinformation on Social Media: Targeted Newsfeed Interventions and Freedom of Thought*, 21 NW. J. HUM. RTS. 109, 111 (2023).

[10] Patrick O'Callaghan, Olga Cronin, Brendan D. Kelly, Bethany Shiner, Joel Walmsley & Simon McCarthy-Jones, *The Right to Freedom of Thought: An Interdisciplinary Analysis of the UN Special Rapporteur's Report on Freedom of Thought*, 28 INT'L J. HUM. RTS. 1, 3 (2024); Simon McCarthy-Jones, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century*, 2 FRONTIERS A.I. 1, 1 (2019).

[11] Ahmed Shaheed & Wayne Martin, *Why Freedom of Thought Is So Important and How to Protect It*, ESSEX BLOGS (May 25, 2022), https://www.essex.ac.uk/blog/posts/2022/05/25/why-freedom-of-thought-is-so-important-and-how-to-protect-it.

possibility that everything one views online could be false, it disrupts individuals' ability to form their thoughts grounded in truthful information.[12]

This Note advocates for the adoption of a Third Optional Protocol to the International Covenant on Civil and Political Rights (ICCPR), affirming that technology can violate freedom of thought, as this right should include the right to base one's thoughts on truthful information. The Protocol should require states to refrain from using technology to infringe on this right and ensure that individuals within their jurisdictions who believe their right to freedom of thought has been violated through technology have access to domestic legal remedies, and, if needed, consideration by the Human Rights Committee.

Part I defines key terms, describes how countries use technology to interfere with another nation's freedom of thought, and discusses the concept of freedom of thought along with the issues associated with it. Part II examines relevant principles found in treaties, conventions, and customary law. Finally, Part III suggests potential language for a Third Optional Protocol to the ICCPR and outlines a practical approach for its implementation.

I. BACKGROUND INFORMATION

First, it is critical to understand the differences between mis-, dis-, and malinformation and to be aware of the tools that state actors can use to disseminate false information, which can affect freedom of thought. After defining these terms, the scope of the right to freedom of thought, its shortcomings, and its modern interpretation are examined, particularly considering its current lack of a clear definition.

A. Definitions and Key Terms

This section defines key terms and explains how certain technologies like deepfakes, algorithms, and bots operate, laying the foundation for understanding how countries have abused these tools to infringe on freedom of thought.

1. Mis-, Dis-, and Malinformation

"Fake news" is a term that is constantly used in the media[13] to refer to "false or inaccurate information."[14] However, because there are many different categories of information that could fall under this definition, using this term can be unhelpful or misleading.[15] As a result, the term has fallen out of favor,[16] with more precise terms—misinformation, disinformation, and malinformation (collectively referred to as

---

[12] *See* Montlake, *supra* note 7.
[13] *See generally* Daniel Funke, *Should We Stop Saying 'Fake News'?*, POYNTER (Dec. 14, 2017), https://www.poynter.org/fact-checking/2017/should-we-stop-saying-fake-news (explaining that "[i]n 2017, fake news was everywhere," and was even "named word of the year by Collins Dictionary.").
[14] *Disinformation and Human Rights Explained*, GLOB. PARTNERS DIGIT. (June 1, 2023), https://www.gp-digital.org/a-human-rights-based-approach-to-disinformation.
[15] The term "fake news" can refer to satire or parody content, misleading content, true information shared in the wrong context, entirely fabricated content, and even information that the person sharing simply does not like. *Id.*
[16] Joshua Habgood-Coote, *The Term 'Fake News' Is Doing Great Harm*, THE CONVERSATION (July 27, 2018, 8:55 AM), https://theconversation.com/the-term-fake-news-is-doing-great-harm-100406.

"MDM")—being adopted in its place.[17] Although there are no universally accepted definitions for these terms, various agencies have provided definitions of their own.[18]

Misinformation is commonly defined as false information that is "not created or shared with the intention of causing harm,"[19] as "the person who is disseminating it believes that it is true."[20] For example, someone accidentally sharing an "inaccurate photo[], quote[], or date[] online" would be considered spreading misinformation because they are unaware the information is false.[21] Though misinformation does admittedly impede on one's right to base their thoughts on truthful and accurate information, this Note focuses only on false information spread or created with malicious intent—thus, misinformation will not be further discussed.

Disinformation, on the other hand, is an "intentional falsehood[] or distortion[], often spread as news, to advance political goals such as discrediting opponents, disrupting policy debates, influencing voters, inflaming existing social conflicts, or creating a general backdrop of confusion and informational paralysis."[22] Disinformation may include things like "fabricated news stories and political propaganda."[23]

Finally, "[m]alinformation is [information] based on fact, but used out of context to mislead, harm, or manipulate."[24] For example, "editing a video to remove important context to harm or mislead" would be considered malinformation.[25] Although this type of information may arguably be considered truthful—since the information itself was not fabricated—malinformation still infringes on the right to base one's thoughts on truthful information, because by removing the important context of a sentence or selectively leaking only certain information, individuals are prevented from the ability to base their thoughts on the full context of a situation.

---

[17] *What is Mis-, Dis-, or Malinformation (MDM)?*, A MORE PERFECT UNION (June 6, 2024), https://www.jewishdemocracy.org/resources-list/what-is-mis-dis-or-malinformation-mdm.
[18] *Disinformation and Human Rights Explained*, *supra* note 14.
[19] *Mis-, Dis-, and Malinformation: Planning and Incident Response Guide for Election Officials*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY 1, https://www.cisa.gov/sites/default/files/publications/mdm-incident-response-guide_508.pdf; *Disinformation and Human Rights Explained*, *supra* note 14.
[20] *Misinformation, Disinformation and Mal-information*, MEDIA DEFENCE, https://www.mediadefence.org/ereader/publications/modules-on-litigating-freedom-of-expression-and-digital-rights-in-south-and-southeast-asia/module-8-false-news-misinformation-and-propaganda/misinformation-disinformation-and-mal-information (last visited Oct. 20, 2024).
[21] *Fake News, Hoaxes and Misinformation*, NSPCC, https://www.nspcc.org.uk/keeping-children-safe/online-safety/inappropriate-explicit-content/fake-news (last visited Oct. 22, 2024).
[22] W. Lance Bennett & Steven Livingston, *A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority*, in THE DISINFORMATION AGE: POLITICS, TECH., & DISRUPTIVE COMMC'N. IN THE U.S. 3 (W. Lance Bennett & Steven Livingstons eds., 2020).
[23] NSPCC, *supra* note 21.
[24] *Mis-, Dis-, and Malinformation: Planning and Incident Response Guide for Election Officials*, supra note 19, at 1.
[25] *Foreign Influence Operations and Disinformation*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY, https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation (last visited Oct. 22, 2024).

2. Social Media as a News Source

The rise of the internet has significantly accelerated the spread of information, which has also led to an increase in the spread of MDM.[26] Social media, in particular, has had a massive impact on escalating its spread.[27] Before the advent of social media, people around the world typically relied on news agencies held to certain "agreed-upon ethics and standards" for their news and information.[28] Today, however, much of the news is sourced from peers online.[29] In fact, studies show that "[fifty-three percent] of Americans get at least some of their news from social media,"[30] with social media being "the most common news source" among eighteen- to twenty-nine-year-olds.[31]

Social media as a news source poses an issue because little to no content is fact-checked and freedom of speech allows people to post almost anything.[32] To an extent, "social media platforms [even] inadvertently encourage users to spread misinformation" because content with more likes, comments, and shares gets pushed to more viewers.[33] While most social media content is provided on a "buyer beware" basis—with users generally under no obligation to provide accurate information[34]—not all social media users take the time to fact-check the information they read online.[35] Because many people choose not to verify the information they learn on social media, bad actors can easily exploit this and intentionally spread disinformation and malinformation.

3. Deepfakes

One method for spreading disinformation on social media and other news sources is the use of deepfakes. Deepfakes are false videos or images that are "created

---

[26] *Misinformation, Disinformation and Mal-information*, MEDIA DEFENCE, https://www.mediadefence.org/ereader/publications/modules-on-litigating-freedom-of-expression-and-digital-rights-in-south-and-southeast-asia/module-8-false-news-misinformation-and-propaganda/misinformation-disinformation-and-mal-information (last visited Oct. 22, 2024).

[27] *See, e.g.*, Anastasia Micich & R.J. Cross, *How Misinformation on Social Media Has Changed News*, U.S. PIRG (Nov. 22, 2023), https://pirg.org/edfund/articles/misinformation-on-social-media.

[28] *See Misinformation, Disinformation and Mal-information*, *supra* note 26.

[29] *See, e.g.*, *Misinformation, Disinformation and Mal-information*, *supra* note 26; *Social Media and News Fact Sheet*, PEW RSCH. CTR. (Sept. 17, 2024), https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet.

[30] Micich & Cross, *supra* note 27; PEW RSCH. CTR., *supra* note 29.

[31] Micich & Cross, *supra* note 27.

[32] *See* Olivia Kerben, *Relying on Social Media for News Is Dangerous*, THE MIAMI STUDENT (Mar. 5, 2024), https://www.miamistudent.net/article/2024/03/read-the-news?ct=content_open&cv=cbox_latest.

[33] Gizem Ceylan & Susie Allen, *How Social Media Rewards Misinformation*, YALE INSIGHTS (Mar. 31, 2023), https://insights.som.yale.edu/insights/how-social-media-rewards-misinformation; Micich & Cross, *supra* note 27.

[34] *See* Jane Bambauer, *Negligent AI Speech: Some Thoughts About Duty*, 3 J. FREE SPEECH L. 343, 351 (2023). *See also Seven Best Practices for Doctors on Social Media*, FPM (Feb. 19, 2020), https://www.aafp.org/pubs/fpm/blogs/inpractice/entry/social_media_best_practices.html (explaining that "[p]hysicians on social media have a professional responsibility to ensure the accuracy of the health-related content they create or share," just one example of when social media users may be obliged to provide correct information online).

[35] *See generally* Eileen Brown, *9 Out of 10 Americans Don't Fact-check Information They Read on Social Media*, ZDNET (May 10, 2017, 2:00 PM), https://www.zdnet.com/article/nine-out-of-ten-americans-dont-fact-check-information-they-read-on-social-media (stating that "American adults spread fake news because they place a lot of trust in their social media friends and acquaintances" who share the posts that appear in their news feeds).

via deep learning,[36] a form of artificial intelligence, where a person's likeness, including their face and voice, can be realistically swapped with someone else's."[37]

Deepfakes can be created through various methods,[38] but the most common way involves the following steps. For face swapping, face images of two people are first "run through an AI algorithm" which "finds and learns similarities between the two faces, and [then] reduces them to their shared common features."[39] These two images are then fed into a second algorithm, which reconstructs the original image it was given, producing a fake yet highly realistic image.[40] For voice cloning, a similar process is followed, but audio recordings are used instead of images.[41] With enough audio input, the AI can "replicate the slight intonations of [a person's] speech with startling accuracy."[42]

Since deepfake creation software is becoming increasingly accessible and affordable,[43] it is now "possible for anyone to generate false information and fake content in vast quantities,"[44] raising concerns about potential foreign interference, widespread confusion, and other threats.[45]

4. Social Media Algorithms and Bots

Disinformation and malinformation alike can be perpetuated through manipulating social media algorithms, including by using bot accounts. Social media algorithms are "rules, signals[,] and data that govern the platform's operation . . . and determine how content is filtered, ranked, selected[,] and recommended to users."[46] These algorithms typically consider factors such as user engagement, content type,

---

[36] For more information on deep learning generally, see *What is Deep Learning?*, AMAZON WEB SERVICES, https://aws.amazon.com/what-is/deep-learning (last visited Oct. 16, 2024) ("Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain. Deep learning models can recognize complex patterns in pictures, text, sounds, and other data to produce accurate insights and predictions.").

[37] *Deepfake Technology*, ORG. FOR SOC. MEDIA SAFETY, https://www.socialmediasafety.org/advocacy/deepfake-technology (last visited Oct. 5, 2024).

[38] Deepfakes can also be created through generative adversarial networks, or GANs. The author describes: "A G[AN] pits two artificial intelligence algorithms against each other. The first algorithm, known as the generator, is fed random noise and turns it into an image. This synthetic image is then added to a stream of real images—of celebrities, say—that are fed into the second algorithm, known as the discriminator." Although the images will not look like real faces at first, repeating the process eventually results in extremely realistic faces. *See* Ian Sample, *What Are Deepfakes – And How Can You Spot Them?*, THE GUARDIAN (Mar. 3, 2023), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

[39] *Id.*

[40] *Id.*

[41] Michael Sum, *Beyond ChatGPT: How Are A.I. Voices Made?*, VOQUENT (Apr. 5, 2023), https://www.voquent.com/blog/beyond-chatgpt-how-are-ai-voices-made/?srsltid=AfmBOopAiGxz_r4LdjVVznDqL5D9ZIZfLLlKYZ1OF8_i_Hx-XUluyKd0.

[42] *Id.*

[43] Adam Satariano & Paul Mozur, *The People Onscreen Are Fake. The Disinformation Is Real*, N.Y. TIMES (Feb. 7, 2023), https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html.

[44] Julius Endert, *Generative AI Is the Ultimate Disinformation Amplifier*, DW AKADEMIE (Mar. 26, 2024), https://p.dw.com/p/4doP4.

[45] *See* discussion *infra* Part I(B)(3).

[46] Dorcas Adisa, *Everything You Need to Know About Social Media Algorithms*, SPROUT SOCIAL (Oct. 30, 2023), https://sproutsocial.com/insights/social-media-algorithms.

virality, watch time, recency, and the posting account's popularity when determining which content to push to users.[47] Although ideally, the algorithm would push content based on organic user activity, some programmers have taken advantage of the ability to influence which content the algorithm prioritizes by creating bot accounts.[48]

Social media bots are "automated programs designed to mimic human behavior on social media."[49] Bots are typically "associated with fake social media profiles, business accounts, or pages,"[50] and are created to "participate in heated debates on various social media posts or engage in spam-like activities"[51] and "boost the apparent popularity of other accounts" by following, liking, and commenting on other accounts' posts.[52] Bots are a common feature and huge concern of every social media platform.[53] In 2022, researchers estimated that bot accounts made up between twenty-five to sixty-eight percent of all users on Twitter (now known as X).[54]

Because bots are designed to generate "a large volume of posts, comments, and likes, [they] can create the illusion of widespread support or opposition for a particular viewpoint, swaying public opinion in a desired direction,"[55] and can cause the algorithm to push such content to more users due to the post's high level of engagement.[56]

B. How Countries Use Technology to Influence Thought

With technology and social media connecting countries more than ever before, new avenues have opened for countries to intentionally infringe on each other's freedom of thought. While many examples of these actions involve individuals from other countries, only some have been attributed to state actors—that is, a "public official [of another country] acting in [their] official capacity."[57] This section identifies examples where state actors have been accused or proven to have used technology to interfere with this freedom, and then discusses the broader effects of such interference.

1. Spreading Disinformation

Countries can interfere with another nation's freedom of thought by spreading disinformation on social media.[58] Although this can be accomplished in many ways,

---

[47] *Id.*
[48] *See What Is a Social Media Bot? | Social Media Bot Definition*, CLOUDFLARE, https://www.cloudflare.com/learning/bots/what-is-a-social-media-bot (last visited Oct. 26, 2024).
[49] Sanja Trajcheva, *Social Bots: How Do They Shape Public Opinion?*, CHEQ (July 12, 2023), https://cheq.ai/blog/social-bots-how-do-they-shape-public-opinion.
[50] *Id.*
[51] *Id.*
[52] *What Is a Social Media Bot? | Social Media Bot Definition*, *supra* note 48.
[53] *See, e.g.,* Trajcheva, *supra* note 49.
[54] *Id.*
[55] *Id.*
[56] *See* Dorcas Adisa, *Everything You Need to Know About Social Media Algorithms*, SPROUT SOCIAL (Oct. 30, 2023), https://sproutsocial.com/insights/social-media-algorithms.
[57] *State Actor Definition*, LAW INSIDER, https://www.lawinsider.com/dictionary/state-actor (last visited Nov. 27, 2024).
[58] *See* Maria Pawelec, *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*, 1 DIGIT. SOC'Y 1, 25 (2022).

this section focuses on two common approaches: using deepfakes to deceive viewers and utilizing bots to create the illusion of widespread support for a particular issue.

First, deepfakes can be posted online to mislead viewers. For example, just "[d]ays before a pivotal [presidential] election in Slovakia, . . . a damning audio recording spread online in which one of the top candidates seemingly boasted about how he [had] rigged the election."[59] However, the recording was generated by AI.[60] Although the source of the deepfake has not been officially confirmed, the post is believed to be "part of a wider influence campaign by Russia."[61]

Thankfully, many deepfakes can currently be detected due to their low quality, because creating high-quality deepfakes can be expensive and time-consuming.[62] However, "deepfake technology is quickly maturing and increasingly becoming more difficult to detect," demonstrating an urgent need to implement safeguards for it now.[63]

Social media bots can also be used to spread disinformation—and perhaps in a more believable way.[64] Although deepfakes can be identified somewhat easily (depending on their quality),[65] bots can be harder to detect because they make up such a large portion of social media accounts and are designed to act as a human would online.[66]

For instance, the Russian government recently operated a social media bot farm as part of a Kremlin-approved and funded project run by a Russian intelligence officer.[67] For this project, "elements of AI [were used] to create fictitious social media profiles—often purporting to belong to individuals in the United States—which the operators then used to promote messages in support of Russian government objectives."[68] Most of the posts were made on X, formerly known as Twitter, and showed support for Russia's war in Ukraine and other pro-Kremlin narratives.[69]

---

[59] Curt Devine, Donie O'Sullivan & Sean Lyngaas, *A Fake Recording of a Candidate Saying He'd Rigged the Election Went Viral. Experts Say It's Only the Beginning*, CNN (Feb. 1, 2024, 6:09 AM), https://www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html.
[60] *Id.*
[61] *Id.*
[62] Catherine Bernaciak & Dominic A. Ross, *How Easy Is It to Make and Detect a Deepfake?*, CARNEGIE MELLON UNIV. (Mar. 14, 2022), https://insights.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake.
[63] Don Philmlee, *Practice Innovations: Seeing Is No Longer Believing—The Rise of Deepfakes*, THOMSON REUTERS (July 18, 2023), https://www.thomsonreuters.com/en-us/posts/technology/practice-innovations-deepfakes/#:~:text=Right%20now%2C%20deepfake%20technology%20is,becoming%20more%20difficult%20to%20detect.
[64] *See* Emily Harding, *A Russian Bot Farm Used AI to Lie to Americans. What Now?*, CTR. FOR STRATEGIC & INT'L STUD. (July 16, 2024), https://www.csis.org/analysis/russian-bot-farm-used-ai-lie-americans-what-now.
[65] Bernaciak & Ross, *supra* note 62.
[66] *See* Harding, *supra* note 64.
[67] Shannon Bond, *U.S. Says Russian Bot Farm Used AI to Impersonate Americans*, NPR (July 9, 2024, 5:32 PM), https://www.npr.org/2024/07/09/g-s1-9010/russia-bot-farm-ai-disinformation.
[68] Press Release, Office of Public Affairs, Justice Department Leads Efforts Among Federal, International, and Private Sector Partners to Disrupt Covert Russian Government-Operated Social Media Bot Farm (July 9, 2024), https://www.justice.gov/opa/pr/justice-department-leads-efforts-among-federal-international-and-private-sector-partners.
[69] Bond, *supra* note 67.

Not only were U.S. citizens tricked by these realistic accounts—social media platforms also believed that these false "personas [were] not bots at all."[70] Alarmingly, the operation "had been underway for two years" before it was noticed and eventually shut down.[71] Experts worry that in the near future, "the state of the art in AI will be such that a bot can identify the messages that resonate best with a micropopulation and then feed that population what they want to hear," making bots "feel as local and genuine as a conversation over the fence with a neighbor,"[72] and thus, even more difficult to detect.

2. Leaking Malinformation

Not all false information online is entirely fabricated, however. Countries can also infringe on freedom of thought through spreading malinformation on social media to affect other countries.[73] Malinformation can include "selectively edited videos or true information exaggerated to appear more sinister,"[74] and typically involves hacking or leaking sensitive information to create a desired narrative.[75] By strategically leaking truthful information taken out of context, countries can expose vulnerabilities of other countries and sway public opinion among their citizens.[76]

For example, in 2016, Russia "hack[ed] the Democratic National Committee and leak[ed] documents aimed at undercutting Hillary Clinton's presidential campaign."[77] More recently, the United States faced another "major Russian government-backed effort to influence the 2024 U.S. presidential election"[78] when Vladimir Putin, the country's president, directed "three Russian companies [to] use[] fake profiles to promote false narratives on social media,"[79] with "one of the goals... [being] to boost the candidacy of Donald Trump or whoever emerged as the Republican nominee for president."[80]

The United States was involved in a similar scheme in 2022, when "an influence campaign pushing U.S. interests abroad" was discovered and taken down from

---

[70] Harding, *supra* note 64.
[71] *Id.*
[72] *Id.*
[73] *See, e.g.*, Patrick O'Callaghan, Olga Cronin, Brendan D. Kelly, Bethany Shiner, Joel Walmsley & Simon McCarthy-Jones, *The Right to Freedom of Thought: An Interdisciplinary Analysis of the UN Special Rapporteur's Report on Freedom of Thought,* 28 INT'L J. HUM. RTS. 1, 12 (2024).
[74] *What is Mis-, Dis-, or Malinformation (MDM)?,* A MORE PERFECT UNION (June 6, 2024), https://www.jewishdemocracy.org/resources-list/what-is-mis-dis-or-malinformation-mdm.
[75] *See* Luca Tancredi, *Understanding Misinformation, Disinformation and Malinformation: Definitions and Real-world Examples*, SUSTAINABLE COOP. FOR PEACE & SEC. (July 20, 2024), https://sustainablepeace.eu/understanding-misinformation-disinformation-and-malinformation-definitions-and-real-world-examples.
[76] *See id.*
[77] Sean Lyngaas, Evan Perez, Kylie Atwood, Zachary Cohen & Jennifer Hansler, *Biden Administration Announces Major Actions to Tackle Russian Efforts to Influence 2024 Election*, CNN: POLITICS (Sept. 4, 2024), https://www.cnn.com/2024/09/04/politics/biden-administration-accuse-russia-election-influence-efforts/index.html; Simon Ostrovsky & Yegor Troyanovsky, *How Russia is Using Artificial Intelligence to Interfere in Elections*, PBS NEWS (Sept. 4, 2024, 6:35 PM), https://www.pbs.org/newshour/show/how-russia-is-using-artificial-intelligence-to-interfere-in-elections.
[78] Lyngaas et al., *supra* note 77.
[79] *Id.*
[80] *Id.*

Facebook, Instagram, WhatsApp, and Twitter.[81] The campaign used accounts "posed as news outlets or... people who didn't exist" to promote American views and attack the interests of Russia, China, and Iran.[82] Although Russia receives plenty of attention for spreading disinformation and malinformation using technology, they are evidently not the only country that is guilty—this is a widespread issue that needs to be addressed.[83]

3. Effects of Interference

Undoubtedly, these types of online content can mislead people and have tangible effects on the world.[84] On a societal level, disinformation and malinformation can impact a country's "media, politics, science, and economics."[85] It can cause people to "disengage[] with news media in general and [lose] trust in traditional news institutions,"[86] "lead[] voters toward advocating policies that are counter to their own interests,"[87] influence "public understanding of science,"[88] and result in financial costs due to efforts "to debunk misinformation," raise awareness, and regulate it.[89]

On an individual level, disinformation and malinformation can have psychological effects.[90] It is often challenging for people to revise their "beliefs when false claims are retracted,"[91] and "frequent exposure to misinformation is claimed to hinder people's general ability to distinguish between true and false information."[92] Most importantly, using technology to spread disinformation and malinformation is

---

[81] Sheera Frenkel & Tiffany Hsu, *Facebook, Twitter and Others Remove Pro-U.S. Influence Campaign*, N.Y. TIMES (Aug. 24, 2022), https://www.nytimes.com/2022/08/24/technology/facebook-twitter-influence-campaign.html.

[82] *Id.*

[83] For examples of countries interfering in other countries' elections, see generally Chen-Ling Hung, Wen-Cheng Fu, Chang-Ce Liu, & Hui-Ju Tsai, *AI Disinformation Attacks and Taiwan's Responses During the 2024 Presidential Election*, THOMSON FOUND., https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf (last visited Oct. 19, 2024) (describing how China used "AI-fabricated content, misleading narratives[,] and coordinated behavior" to influence the outcome of Taiwan's 2024 election) (internal quotation marks omitted); Press Release, U.S. Department of Justice: Office of Public Affairs, Two Iranian Nationals Charged for Cyber-Enabled Disinformation and Threat Campaign Designed to Influence the 2020 U.S. Presidential Election (Nov. 18, 2021), https://www.justice.gov/opa/pr/two-iranian-nationals-charged-cyber-enabled-disinformation-and-threat-campaign-designed (describing how "two Iranian nationals [were charged] for their involvement in a cyber-enabled campaign to intimidate and influence American voters" by disseminating a video with disinformation, among other things); and Bill Bigelow, *The Overthrow of Democracy in Chile—A Timeline*, ZINN EDUC. PROJECT (2023), https://www.zinnedproject.org/materials/chile-coup-timeline (last visited Oct. 19, 2024) (explaining that in 1964, "[t]he United States Central Intelligence Agency spen[t] $20 million in Chilean election to defeat the Marxist candidate, Salvador Allende" using technology available at the time).

[84] Zoë Adams, Magda Osman, Christos Bechlivanidis & Björn Meder, *(Why) Is Misinformation a Problem?*, 18(6) PERSPS. ON PSYCH. SCI. 1436, 1438–45 (2023).

[85] *Id.* at 1438.

[86] *Id.* at 1438–40.

[87] *Id.* at 1440.

[88] *Id.* at 1441.

[89] *Id.*

[90] *Id.* at 1442.

[91] *Id.*

[92] *Id.* at 1443.

problematic because it hinders people from being able to form their own thoughts based on truthful information, and arguably infringes on their right to freedom of thought.[93]

C. Freedom of Thought

Freedom of thought is an internationally recognized and fundamental human right.[94] It was first legally recognized in 1948 with the passage of the Universal Declaration of Human Rights (UDHR), and became formally recognized as international law in 1976 when the International Covenant on Civil and Political Rights (ICCPR) came into force.[95] Since its establishment, the Human Rights Council has clarified that the right to freedom of thought is protected unconditionally, permitting no limitations on this right.[96]

1. What is Freedom of Thought?

A preliminary issue is defining what "thought" entails. Since Article 18 of the ICCPR groups together the rights to freedom of thought, conscience, and religion,[97] and given that there are several other human rights (including, among others, the freedoms of conscience, belief, expression, and privacy) which "already safeguard the interests in question, even if they do so indirectly,"[98] it is hard to parse out where one freedom ends and the other begins.

Due to this ambiguity, some have characterized the relationship between freedom of thought and its other closely related freedoms as having two parts.[99] Freedom of thought itself can be considered "a freedom to think and believe whatever [one] choose[s]," while the other rights pertain to "a freedom to demonstrate [one's] thoughts or beliefs publicly."[100] Thus, at its most basic level, freedom of thought can

---

[93] *See* Richard Mackenzie-Gray Scott, *Managing Misinformation on Social Media: Targeted Newsfeed Interventions and Freedom of Thought*, 21 NW. J. HUM. RTS. 109, 111 (2023); Simon Montlake, *Trust in the Media Has Tanked. Are We Entering a 'Post-News' Era?*, THE CHRISTIAN SCI. MONITOR (May 3, 2024, 3:25 PM), https://www.csmonitor.com/USA/Politics/2024/0503/news-media-decline-polarization-distrust.

[94] *See, e.g.*, Patrick O'Callaghan, Olga Cronin, Brendan D. Kelly, Bethany Shiner, Joel Walmsley & Simon McCarthy-Jones, *The Right to Freedom of Thought: An Interdisciplinary Analysis of the UN Special Rapporteur's Report on Freedom of Thought*, 28 INT'L J. HUM. RTS. 1, 3 (2024); Simon McCarthy-Jones, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century*, 2 FRONTIERS A.I. 1, 1 (2019).

[95] *See, e.g.*, Simon McCarthy-Jones, *supra* note 94, at 5; *FAQ: The Covenant on Civil & Political Rights (ICCPR)*, ACLU (July 11, 2013), https://www.aclu.org/documents/faq-covenant-civil-political-rights-iccpr#:~:text=The%20Covenant%20was%20adopted%20by,to%20comply%20with%20the%20ICCPR%3F.

[96] Human Rights Council, *Report of the Special Rapporteur on Freedom of Religion or Belief*, ¶ II(A)(2), A/HRC/31/18 (Dec. 23, 2015).

[97] G.A. Res. 2200A (XXI), International Covenant on Civil and Political Rights (Dec. 16, 1966); O'Callaghan et al., *supra* note 94.

[98] O'Callaghan et al., *supra* note 94.

[99] *Right to Freedom of Thought, Conscience, Religion and Belief*, QUEENSL. HUM. RTS. COMM'N, https://www.qhrc.qld.gov.au/your-rights/human-rights-law/right-to-freedom-of-thought,-conscience,-religion-and-belief (last visited Nov. 27, 2024).

[100] *Id.*

be viewed as a preliminary right for individuals to think for themselves, preceding any right to share or demonstrate those thoughts.[101]

2. Issues with Freedom of Thought

Although both the UDHR and the ICCPR state that "[e]veryone has the right to freedom of thought," neither document elaborates on what this actually means.[102] There is no indication of "what counts as thought, what qualifies as a violation of the right, whether the right should be absolute and, if not, what would justify its violation."[103] Scholars also "note that there is no legal authority and little scholarly analysis on what, precisely, is meant by the 'right to freedom of thought,'" including "what the right protects[] or how it can be used."[104] Additionally, there is no explanation on what the term "thought" means, or even what "freedom" entails in this context.[105]

For this reason, many people and organizations believe the right to freedom of thought needs to be clarified.[106] For example, "the UN Special Rapporteur has encouraged clarification of the content and scope of the [right to freedom of thought] both by individual [s]tates and through a General Comment of the Human Rights Committee."[107]

3. Interpretation of Freedom of Thought

Given the broad scope of the right to freedom of thought, individuals have taken it upon themselves to interpret its meaning.[108] In a 2021 report to the UN General Assembly, "legal scholars, philosophers, ethics experts[,] and neuroscientists" came together to propose four attributes of the right to freedom of thought: (1) "[f]reedom not to disclose one's thoughts"; (2) "[f]reedom from punishment for one's thoughts"; (3) "[f]reedom from impermissible alteration of thoughts"; and (4) "[t]he State obligation to foster an enabling environment for thought."[109] Although this definition is a good start to interpreting the meaning of freedom of thought, the third prong—freedom from impermissible alteration of thoughts—remains unclear as to what this would look like or what a violation would entail.

---

[101] *See id.*
[102] G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948), Art. 18.
[103] Simon McCarthy-Jones, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century*, 2 FRONTIERS A.I. 1, 1 (2019).
[104] Patrick O'Callaghan, Olga Cronin, Brendan D. Kelly, Bethany Shiner, Joel Walmsley & Simon McCarthy-Jones, *The Right to Freedom of Thought: An Interdisciplinary Analysis of the UN Special Rapporteur's Report on Freedom of Thought*, 28 INT'L J. HUM. RTS. 1, 6 (2024).
[105] *Id.*
[106] *See, e.g.*, Sjors Ligthart, Christoph Bublitz, Thomas Douglas, Lisa Forsberg & Gerben Meynen, *Rethinking the Right to Freedom of Thought: A Multidisciplinary Analysis*, 22 HUM. RTS. L. REV. 1, 2 (2022).
[107] *Id.*
[108] *See* Ahmed Shaheed & Wayne Martin, *Why Freedom of Thought Is So Important and How to Protect It*, ESSEX BLOGS (May 25, 2022), https://www.essex.ac.uk/blog/posts/2022/05/25/why-freedom-of-thought-is-so-important-and-how-to-protect-it.
[109] *Id.*

Different countries around the world also have adopted protections for the right to freedom of thought in their respective jurisdictions.[110] For example, in the United States, "[t]he Constitution protects the right . . . to be generally free from governmental intrusions into one's privacy and control of one's thoughts."[111] Other nations hold similar views, with Europe "protect[ing] a space where one can be free from unwanted intrusions," with that space being one's head, home, or even one's mind.[112]

Though different countries have implemented their own protections for the right to freedom of thought—which is certainly a step in the right direction—these measures may only prevent people within their own borders from violating this right. Due to the online nature of social media and the ease with which these actions can cross international boundaries, foreign actors still infringe on this right, demonstrating a crucial need for international law to recognize and address such issues.

II. RELEVANT TREATIES, CONVENTIONS, AND CUSTOMARY LAW

Many nations have domestic laws to protect freedom of thought in their respective countries,[113] but international law must govern international actions.[114] Though there is no treaty currently prohibiting the use of technology to influence freedom of thought, this section examines the non-intervention principle as outlined in international treaties, conventions, and customary law, as well as other cyber norms that may apply when countries take actions that arguably violate the right to freedom of thought.

A. State Sovereignty and the Non-Intervention Principle

The principle of non-intervention stems from the broader principle of sovereignty, which is defined as "the right of a state to rule itself and those who live within its territory...unfettered by outside dictates or interference."[115] Broadly speaking, the non-intervention principle can be found in many international legal documents and "involves the right of every sovereign [s]tate to conduct its [domestic] affairs without outside interference,"[116] though specific language may vary. The principle is a well-accepted "customary rule of international law,"[117] meaning that it is widely recognized as binding law—even without a treaty.[118]

---

[110] *See* Simon McCarthy-Jones, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century*, 2 Frontiers A.I. 1, 5–6 (2019).
[111] *Id.* at 5 (quoting Stanley v. Georgia, 394 U.S. 557, 564 (1969)).
[112] *Id.* at 5–6.
[113] For example, in the United States, the Bill of Rights protects freedom of thought. *Id.* at 2.
[114] *What Is International Law?*, COUNCIL ON FOREIGN REL. (July 25, 2023), https://education.cfr.org/learn/reading/what-international-law#:~:text=Source:%20Reuters,to%20hold%20their%20governments%20accountable.
[115] Douglass Cassel, *A Framework of Norms: International Human-Rights Law and Sovereignty*, 22 HARV. INT'L R. 60, 60 (2001).
[116] Jianming Shen, *The Non-Intervention Principle and Humanitarian Interventions Under International Law*, 7(1) INT'L LEGAL THEORY 1, 6 (2001).
[117] *Id.*
[118] *Customary IHL*, INT'L COMM. OF THE RED CROSS, https://www.icrc.org/en/law-and-policy/customary-ihl#:~:text=Why%20is%20customary%20international%20law,%2C%20as%20such%2C%20are%20binding (last visited Oct. 27, 2024).

1. Sources Prohibiting Intervention

First, the non-intervention principle can be found in Articles 2(4) and 2(7) of the United Nations (UN) Charter,[119] as well as in the Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations, a UN General Assembly document.[120] The non-intervention principle states that "[a]ll Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the [p]urposes of the United Nations."[121] Although the principle focuses on conventional military threats and force, some scholars argue that cyber actions are considered a use of force, since there is a "possibility that massive harm [can also] be perpetuated in cyberspace, rather than physical space."[122]

The Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty, which is another UN General Assembly Declaration, as well as Principle VI of the Helsinki Final Act,[123] adjusts this language and explicitly prohibits all forms of intervention "in the internal and external affairs of any other [s]tate."[124] It also clarifies that all states must respect the "independence of peoples and nations" with "absolute respect for human rights and fundamental freedoms."[125]

2. Application to Technology Use

Challenges arise in applying these sources, though, as none explicitly mention whether they apply to the use of technology. Fortunately, a subsequent document—the Tallinn Manual—was drafted to clarify that the principle of non-intervention and the broader principle of sovereignty apply to contexts involving technology.[126] The Tallinn Manual is "an influential resource for legal advisers and policy experts dealing with cyber issues."[127] It explains that "[t]he [p]rinciple of [s]overeignty applies to

---

[119] U.N. Charter art. 2.
[120] G.A. Res. 25/2625, Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations (Oct. 24, 1970). The language in this document states that "[e]very [s]tate has the duty to refrain in its international relations from the threat or use of force against the territorial integrity or political independence of any [s]tate."
[121] U.N. Charter art. 2, ¶ 4.
[122] *See* Matthew C. Waxman, *Cyber Attacks as "Force" Under UN Charter Article 2(4)*, 87 INT'L LAW STUD. 43–45 (2011); *Use of Force in Cyberspace*, CONG. RSCH. SERV. (June 25, 2024), https://sgp.fas.org/crs/natsec/IF11995.pdf.
[123] The only difference between the two sources is that the Helsinki Final Act adds the words "falling within the domestic jurisdiction" after "affairs." Conference on Security and Co-Operation in Europe Final Act (1975), art. 6, ¶ 1.
[124] G.A. Res. 20/2131, Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty, art. 1 (Dec. 21, 1965).
[125] *Id.* at art. 6.
[126] *See The Tallinn Manual*, CCDCOE, https://ccdcoe.org/research/tallinn-manual (last visited Oct. 27, 2024).
[127] *Id.*

cyberspace,"[128] and that a "[s]tate must not conduct cyber operations that violate the sovereignty of another [s]tate."[129]

Notably, intervention only applies to states, not non-state actors or individuals.[130] Thus, if a state actor used deepfakes, bots, or social media algorithms to disseminate disinformation or malinformation, this could be seen as a form of intervention. While there is no precise definition of what qualifies as an internal affair that cannot be intervened with, the "conduct of national elections" is generally regarded as one.[131] Because using technology in this manipulative way could sway another country's public opinion, lead to distrust in government, or influence a country's elections or economy, this could be viewed as undermining state sovereignty.[132]

B. United Nations Cyber Norms

In addition to the non-intervention principle, non-binding UN Cyber Norms establish other guidelines for international conduct when using technology. In 2021, the UN Open-Ended Working Group (OEWG) established "a framework for responsible state behavior in cyberspace," formalizing norms developed earlier by the UN Group of Governmental Experts (GGE) in 2015.[133] Of the eleven norms they established, two are particularly relevant.

1. Stability and Security in Cyberspace

First, states are recommended to "[c]ooperate to increase stability and security in cyberspace."[134] Although the framework does not specify exactly what this should entail, it emphasizes that "practices that are acknowledged to be harmful or that may pose threats to international peace and security" should be avoided.[135] State actors spreading disinformation and malinformation online would likely violate this principle, as attempting to influence the views of another country's citizens are unlikely to lead to world peace. However, because this norm's language is so vague, states could easily fail to identify their malicious use of technology as something as extreme as a threat to international security, making them unlikely to cease their actions.

---

[128] TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 11 (Michael N. Schmitt ed., 2013) [hereinafter TALLINN MANUAL]; Eric Talbot Jensen, *The Tallinn Manual 2.0: Highlights and Insights*, GEO. J. INT'L LAW 735, 741 (2017).
[129] TALLINN MANUAL, *supra* note 128, at 17 r. 4; Jensen, *supra* note 128.
[130] *See* Jianming Shen, *The Non-Intervention Principle and Humanitarian Interventions Under International Law*, 7(1) INT'L LEGAL THEORY 1, 6 (2001).
[131] *Prohibition of Intervention*, CYBER LAW TOOLKIT, https://cyberlaw.ccdcoe.org/wiki/Prohibition_of_intervention#cite_note-:PI0-2 (last visited Nov. 30, 2024).
[132] *See* Michael Wood, *Non-Intervention (Non-interference in Domestic Affairs)*, PRINCETON UNIV.: ENCYCLOPEDIA PRINCETONIENSIS, https://pesd.princeton.edu/node/551 (last visited Nov. 27, 2024).
[133] James Andrew Lewis, *Creating Accountability for Global Cyber Norms*, CTR. FOR STRATEGIC & INT'L STUD. (Feb. 23, 2022), https://www.csis.org/analysis/creating-accountability-global-cyber-norms.
[134] Bart Hogeveen, *The UN Cyber Norms: How Do They Guide the Responsible Development and Use of Offensive Cyber Capabilities?*, 7 CYBER DEF. REV. 123, 124 (2022).
[135] G.A. Res. 70/237, Developments in the Field of Information and Telecommunications in the Context of International Security (2015).

2. Respect Human Rights Online

States are also recommended to "[r]espect human rights...online."[136] This norm specifically mentions that the rights to privacy and freedom of expression should be protected, yet it does not mention the right to freedom of thought.[137] Although acknowledging that human rights can be infringed upon through technology demonstrates progress, a major shortcoming becomes clear upon closer examination. Spreading disinformation and malinformation arguably does not violate the rights to privacy or expression, as these actions do not prevent individuals from acting freely online, but rather expose them to deceptive content and propaganda. This suggests the need for a new document to address the issue in a more effective way.

C. Issues with Current Law

Evidently, these treaties, conventions, and principles of customary law each fall short in addressing the problem of disinformation and malinformation being spread online. First, most sources focusing on principles of non-intervention or sovereignty fail to address their application to modern-day technology.[138] Although the Tallinn Manual does eventually address this,[139] a major issue remains unresolved—it does not link these actions to the human rights violation it constitutes: an infringement on freedom of thought.

Additionally, UN cyber norms serve merely as a framework, offering suggestions rather than binding international law.[140] Even beyond this limitation, their language is overly vague, lacking specificity regarding what would constitute a violation of the norms or what actions states can take to implement the norms.[141] Moreover, these sources provide no mechanism for individuals to seek recourse if their rights are violated and impose on states no accountability to uphold these principles. Since they are not required to publicly sign or ratify these documents, states lack a tangible obligation to abide by these principles.

Considering these issues, it is apparent that a new document is needed to address these concerns—ideally, this document would be a binding agreement that (1) explicitly recognizes the use of technology to manipulate or influence one's thoughts as a violation of the right to freedom of thought, (2) outlines what states should do to prevent such infringements from occurring, and (3) implements a better way to hold states accountable.

III. PROPOSED SOLUTION

Although many scholars and organizations have attempted to clarify the right to freedom of thought and establish guidelines for the online conduct of state actors, their efforts have not brought about sufficient meaningful change—perhaps due to their

---

136 Hogeveen, *supra* note 134, at 124.
137 G.A. Res. 70/237, *supra* note 135.
138 *See* discussion *supra* Part II(A)(1).
139 *See* discussion *supra* Part II(A)(2).
140 *See* discussion *supra* Part II(B).
141 *Id.*

failure to connect the two.[142] To remedy this, a Third Optional Protocol to the ICCPR should be adopted, declaring that technology can be used to violate freedom of thought, as this right includes the right to base one's thoughts on truthful information. The Protocol should require each state party to refrain from using technology to infringe on the right to freedom of thought and ensure that individuals within their jurisdictions who believe their right to freedom of thought has been violated through technology have access to domestic legal remedies, and, if necessary, consideration by the Human Rights Committee.

A. Process for Adopting Optional Protocols

Optional Protocols are "additional treaties that provide further rights or processes" to the ICCPR.[143] As its name suggests, these Protocols are not mandatory and are only binding upon the states who "become[] a party to the Protocol."[144] If a state party to the Protocol violates its requirements, "any person subject to the jurisdiction of the [s]tate party may lodge a complaint with the Human Rights Committee."[145]

There are currently two Optional Protocols to the ICCPR.[146] The First Optional Protocol, which was adopted in 1976, "sets out a system by which the Human Rights Committee can receive and consider complaints from individuals who allege that their human rights have been violated."[147] The Second Optional Protocol was adopted in 1991 and aims to abolish the death penalty for states who adopt the Protocol.[148]

Optional Protocols are added by the same lengthy process any international treaty would go through.[149] The treaty must first be drafted, and then opened for signature by UN member states.[150] When a state signs, it "express[es] its consent to be bound by the treaty" until its ratification,[151] which typically involves approval by each state's legislative body.[152] After this occurs, states that have ratified the treaty are required to implement its provisions in good faith.[153]

B. Creating a Third Optional Protocol

To ensure the effectiveness of a Third Optional Protocol to the ICCPR, the following language should be included.

---

[142] *See* Sjors Ligthart, Christoph Bublitz, Thomas Douglas, Lisa Forsberg & Gerben Meynen, *Rethinking the Right to Freedom of Thought: A Multidisciplinary Analysis*, 22 HUM. RTS. L. REV. 1, 2 (2022).
[143] *International Covenant on Civil and Political Rights (ICCPR)*, EQUAL. & HUM. RTS. COMM'N (Apr. 10, 2024), https://humanrightstracker.com/en/un-treaty/iccpr.
[144] U.N. Human Rights Committee, Background to the International Covenant on Civil and Political Rights and Optional Protocols, https://www.ohchr.org/en/treaty-bodies/ccpr/background-international-covenant-civil-and-political-rights-and-optional-protocols (last visited Nov. 27, 2024).
[145] *Id.*
[146] *Id.*
[147] *Id.*
[148] *Id.*
[149] *See* U.N. OFFICE OF LEGAL AFFAIRS, TREATY HANDBOOK, U.N. Sales No. E.12.V1 (2013).
[150] *See id.*
[151] *Id.* at 5.
[152] *See id.* at 5, 9.
[153] *See, e.g.*, Vienna Convention on the Law of Treaties art. 26, May 23, 1969, 1155 U.N.T.S. 331.

1. Preamble

> The States Parties to the present Protocol,
> Recognizing that advancements to technology and the rapid spread of information have impacted individuals' ability to form their thoughts based on accurate information,
> Realizing that the right to freedom of thought should include the right to base one's thoughts on truthful information,
> Recalling article 18 of the Universal Declaration of Human Rights, adopted on 10 December 1948, and article 18 of the International Covenant on Civil and Political Rights, adopted on 16 December 1966,
> Desirous to undertake hereby an international commitment to put an end to manipulation of one's thoughts through technology,
> Have agreed as follows:

The Third Optional Protocol would begin with a preamble clearly outlining the underlying principles and reasons for adopting this treaty. Highlighting how technological advancements can impact an individual's ability to form independent thoughts before detailing the Protocol's requirements would help establish the connection between freedom of thought and technology. Furthermore, explicitly confirming that the right to freedom of thought includes the right to base one's thoughts on truthful information would define the right's scope more clearly, providing it a tangible requirement and noting that the right can be violated when one's thoughts are subject to manipulation.

2. Article 1: Scope of Application

> This Optional Protocol applies to the use of digital technologies, including artificial intelligence, algorithms, automated bots, and other behavioral targeting mechanisms, that affect freedom of thought as protected under article 18 of the International Covenant on Civil and Political Rights.

The first article should define the Optional Protocol's scope of application. By specifying what "digital technologies" may include, it would ensure states are aware that any technology affecting freedom of thought falls under its scope. Additionally, providing examples of the different types of technologies that are likely to infringe on freedom of thought would help states understand what the Protocol may apply to. This article intentionally does not provide a comprehensive list of every qualifying technology, with hopes that in the future, it can adapt to new technological advancements that may also threaten the right to freedom of thought.

3. Article 2: Restrictions and Requirements

a. No State Party shall use digital technologies to infringe on their citizens' freedom of thought or the freedom of thought of citizens of other nations.
b. Each State Party shall take appropriate measures to prevent digital technologies from being used to infringe on their citizens' freedom of thought or the freedom of thought of other nations.

The first paragraph of the second article should prohibit states party to the Protocol from using technology to violate the freedom of thought of their own citizens

or those of other nations. While it may seem some states could be reluctant to adopt the Protocol if it restricts their ability to spread disinformation and malinformation to citizens of other countries in addition to their own citizens, this is likely not the case. Countries that spread such propaganda typically desire favorable opinions from everyone—especially their own citizens.[154] Thus, these states would likely not be further deterred by the seemingly additional requirement of refraining to use technology to affect the freedom of thought of individuals in other nations.

The second paragraph should expound on the previous one, not only requiring states themselves to refrain from using technology to infringe on freedom of thought, but also requiring them to take measures to prevent others from using technology to infringe on this right. This would prevent states who spread disinformation and malinformation from passing the blame or responsibility to other parties, such as non-state actors or tech companies.

Although only requiring "appropriate measures" may seem too vague, this flexibility is deliberate, as it allows nations to decide for themselves the actions they will take to best comply. Doing so would ideally deter nations from resisting adoption of the Protocol simply because it has overly strict or costly requirements, much like the approach other international treaties seem to have taken.[155]

4. Article 3: Remedies and Enforcement Mechanisms

a. Individuals who believe their right to freedom of thought has been infringed upon through technology must have access to legal remedies within their country.
b. With respect to the States Parties to the first Optional Protocol to the International Covenant on Civil and Political Rights adopted on 16 December 1966, the competence of the Human Rights Committee to receive and consider communications from individuals subject to its jurisdiction shall extend to the provisions of the present Protocol, unless the State Party concerned has made a statement to the contrary at the moment of ratification or accession.

Finally, the third article should outline the remedies available to individuals who believe their rights have been violated and detail how this would be enforced. The Protocol would be ineffective if it did not ensure that victims have access to domestic legal remedies, as this would prevent individuals from receiving help when their right to freedom of thought is infringed upon.

Affirming that victims may submit complaints to the Human Rights Committee if legal remedies within their countries are insufficient would ensure the UN is aware of states that may be violating this right. Notably, this paragraph borrows the exact language from Article 5 of the Second Optional Protocol.[156] This approach would

---

[154] *See* Bianca Ortiz-Miskimen, *Online Political Manipulation Efforts Persist in Russia, China, and the Middle East*, PRINCETON SCH. OF PUB. & INT'L AFFS. (Aug. 10, 2020), https://spia.princeton.edu/news/online-political-manipulation-efforts-persist-russia-china-and-middle-east.

[155] There are many examples of "soft language" in international treaties. *See, e.g.*, G.A. Res. 2200A (XXI), International Covenant on Civil and Political Rights, art. 2, ¶ 2 (Dec. 16, 1966).

[156] G.A. Res. 44/128, Second Optional Protocol to the International Covenant on Civil and Political Rights, aiming at the abolition of the death penalty, art. 5 (Dec. 15, 1989).

ensure consistency across the Optional Protocols and enhance clarity for a state's citizens and governments.

C. Policy Considerations

By drafting and adopting this Third Optional Protocol, international law governing the use of technology in relation to freedom of thought would be consolidated in one binding document, resolving the current fragmentation of treaties, customary law, and other sources that fail to address the entire problem on their own. Moreover, by linking the right to freedom of thought with disinformation and malinformation, countries would be encouraged to recognize their actions as a violation of freedom of thought and, ideally, reconsider their behavior.

Though adopting another Optional Protocol—something that has not been done since 1989[157]—may seem like an unimportant, time-consuming, or daunting task, it is crucial that action is taken now. Doing so will ensure that as technology continues to become more advanced, future generations will be protected from its misuse by state actors, other individuals, or companies to influence freedom of thought.

CONCLUSION

The intentional spread of disinformation and malinformation by foreign state actors poses a major problem because countries fail to recognize it as the infringement it is—a violation of the right to freedom of thought. A Third Optional Protocol, as suggested, would hold states accountable for their spread of disinformation and malinformation and provide individuals with solutions when they believe these rights have been infringed on. By implementing this Protocol, the world can make significant progress toward safeguarding the right to freedom of thought as technology becomes increasingly more advanced.

---

[157] *Id.*