

A Weapon of Mass Distraction: Misinformation in the Russian-Ukrainian Conflict

Jerome Ong

Anglo-Chinese School (Independent), 121 Dover Road, 139650, Singapore; jeromeong8888@gmail.com

ABSTRACT: Tackling the misinformation crisis is a perennial challenge that has stumped policymakers for decades. With the advent of social media, this issue has become even more pertinent for discussion. Extensive research has been conducted to broadly assess the misinformation crisis or recommend specific policies in unique circumstances. Yet there has been little investigation into specific case studies of misinformation to determine the future of misinformation regulation. As such, this research paper will predominantly use the context of the Russian-Ukrainian war to discuss the scope and magnitude of the misinformation crisis. Subsequently, it discusses the strengths and limitations of current misinformation regulation before providing a general direction for future regulation, drawing on lessons from the Russian-Ukrainian War.

KEYWORDS: Behavioral and Social Sciences, Algorithms, Russian-Ukrainian War, Misinformation.

■ Introduction

Digital misinformation has emerged as one of the most destabilising forces in the last decade, no longer confined to slow-moving rumors or hearsay.¹ Through a hyper-connected digital ecosystem, misinformation now fuels the agenda of bad actors, sows the seed of fear among the populace, and even threatens to destroy the social fabric of society as we know it today. It is uniquely characterized by its speed, reach, and integration with an indispensable part of our lives – social media.² As such, misinformation has been transformed from a peripheral risk into a systemic threat that requires immediate attention. This is particularly evident in the 2022 Russian-Ukrainian war, which has even been described as the “first social media war” in history.³ While misinformation exists on both sides of the conflict, this research will predominantly analyze pro-Kremlin misinformation, as misinformation has been used as a strategic weapon to influence public sentiment in favor of Russia in order to project military power.⁴

Existing scholarship has examined the role of digital platforms in shaping the spread and governance of misleading information. For instance, Tucker, Theocharis, Roberts, and Barberá demonstrate how social media platforms facilitate the rapid diffusion of political misinformation,⁵ while scholars such as Tarleton Gillespie⁶ and Evelyn Douek² have explored how platform governance and content moderation influence the circulation of misleading content online. However, much of this literature tends to focus either on the technological mechanisms of misinformation spread or the regulatory frameworks designed to address it. Relatively little research employs qualitative case-study analysis to examine how misinformation and regulation operate in practice within a specific geopolitical context. Therefore, this paper adopts a qualitative case-study methodology, using the Russian-Ukrainian war, to examine the dynamics of misinformation. By analyzing documented instances of misinformation throughout the conflict and sit-

uating them alongside existing scholarship on social media governance, this research paper identifies recurring patterns in the types, methods, and platforms through which misinformation can spread. Ultimately, this research paper aims to answer the central question: What does the Russian-Ukrainian war teach us about how misinformation operates across platforms, and how can regulation address it? By the end of the investigation, the research paper proves that the misinformation in modern conflicts is not merely the result of individual actors spreading false information. Instead, it is amplified by the structural features of social media platforms and the current inadequacies of self-regulation and state-run regulation. By situating these empirical observations within the broader debate over platform self-regulation and state intervention, the research paper also aims to derive broader insights from the Russian-Ukrainian conflict that can inform future approaches to regulating misinformation.

Section 1 will first define the scope of misinformation, concretizing the distinction among misinformation, disinformation, and misleading information. Subsequently, it will lay out the magnitude of harm caused by misinformation, namely through the creation of inaccurate narratives and the formation of echo chambers. By further examining the international scale and deep-rooted tensions surrounding the Russian-Ukrainian war, this section will also stress why regulation has become a pressing necessity.

Building on this foundation, Section 2 will analyze the two main types of misinformation, military-related and political-related, in the Russian-Ukrainian war. While this research has already established that misinformation is systematically amplified by social media algorithms, engineered to provoke emotional responses, and capable of influencing millions, this section seeks to evaluate the various methods and media used to spread misinformation in a contemporary geopolitical landscape. Lastly, Section 2 will introduce society's current

response to misinformation and ultimately demonstrate that the absence of social media regulation will inevitably lead to its spread.

Section 3 directly examines the contentious debate over social media regulation, evaluating the strengths and limitations of self-regulation and state regulation. Through global case studies and examples from the Russian-Ukrainian war, Section 3 will highlight the trade-off between freedom of speech and efforts to reduce harmful misinformation. This sheds light on the complex and dynamic challenge of misinformation regulation while considering the various motivations of stakeholders.

Finally, Section 4 will consolidate the findings from the first three sections and holistically analyze the nuances of the misinformation crisis in the Russian-Ukrainian war. This paper will not only evaluate current efforts to regulate misinformation but also discuss a sustainable path forward. This paper aims to inform efforts to strengthen global resilience to misinformation, thereby protecting the social fabric that underpins both democratic governance and civic stability.

SECTION 1: THE MISINFORMATION CRISIS:

To understand the importance of misinformation regulation, Section 1 will outline what constitutes misinformation, before further contextualizing it in relation to the Russian-Ukrainian war. This section aims to fulfill these objectives by elucidating the detrimental impacts and immense scale of the misinformation crisis. Coupled with the complicating factors of the Russian-Ukrainian war, the need for social media regulation will become apparent.

Section 1.1: The scope of misinformation:

Caldwell University defines misinformation as the spread of misleading information without deceptive intent. Additionally, it defines disinformation as “false or misleading information shared with the intent to mislead, confuse, or deceive people.”⁷ While intent to mislead distinguishes misinformation from disinformation, discerning intent in the real world is challenging. In the context of the Russian-Ukrainian war, for instance, isolated instances of misleading content often circulate online without clear attribution. Thus, it is difficult to determine whether the misinformation stems from genuinely misinformed individuals or from Russia’s state-sponsored disinformation campaigns. These campaigns are strategically designed to manipulate public perception on a large scale, giving them far greater reach than independent sources of misinformation. Yet, these same campaigns also disguise disinformation as the personal views of independent, concerned netizens. This further obscures the origins of misleading narratives, making it extremely difficult to distinguish between organically spreading misinformation and coordinated disinformation efforts based on intent.

Therefore, in order to facilitate discussion, this paper defines misinformation as false or inaccurate information, regardless of intent. Given that the primary focus of this paper is the mechanics of regulating misleading information on social media platforms, exploring intent would be unproductive and would require separate research. Therefore, by addressing

both unintentional and malicious fake news, this streamlines the discussion of information classification (whether it is traditionally considered misinformation or disinformation), enabling more in-depth evaluation of both forms of misinformation.

Section 1.2: The magnitude of harm caused by misinformation:

The danger of misinformation lies in its ability to remain indistinguishable from factual information. Given that it is premised on misguided or misleading information, it often relies on emotional manipulation to erode our ability to gather accurate and reliable information. Its harms manifest in two dominant ways, namely the clouding of objective facts and the creation of echo chambers.

Firstly, misinformation clouds objective facts by creating inaccurate narratives. It is significantly easier to produce misinformation than factual information, because factual information requires testing and verification, whereas misinformation relies on emotional rhetoric to appeal to the audience. This results in a greater volume of misinformation than of factual information, leading to oversaturation on social media. This information overload leads the audience to deprioritize the veracity of claims, focusing instead on features that improve audience retention, such as emotive language. This clouds objective facts and makes the audience more susceptible to believing misinformation.

Secondly, misinformation is problematic because it creates echo chambers. Social media algorithms work by recommending to users what they most likely want to watch. This means that the platform prioritizes content that garners user engagement – regardless of its accuracy. Consequently, users are drawn into information silos, where they are exposed only to content that already reinforces their preexisting beliefs, even when those beliefs were borne of misinformation.⁸ At the same time, echo chambers surround users with a community that shares and legitimizes the same views, making it harder for them to recognize that some content is misinformation.

Apart from the apparent problems propagated by misinformation, identifying whether a particular piece of information constitutes misinformation is also frustratingly obtrusive. One possible reason for this could be the blurry line between fact and opinion. Just as it is difficult to discern the intent behind spreading misinformation, it can be difficult to distinguish between facts and opinions. When an individual wants to present an objective fact on social media, there are still countless subjective decisions to make, such as the language used to demonstrate the point, the description of the fact, or even the balance between textual and visual elements.⁹ There is a non-zero chance that the resulting content is misinformation, especially since these subjective elements can be interpreted differently based on a viewer’s worldview or lived experiences, thereby leading to inaccurate interpretations of the information.¹⁰

Another reason for this difficulty stems from the difficulty of discerning fact from fabrication. By definition, short-form content on social media can only provide abridged explanations

or descriptions. For instance, this could look like thirty-second videos with three to four lines of speech explaining why climate change is a hoax, as well as posts featuring aesthetic infographics on healthy eating that claim to be based on “scientific research.”

Although these simplified narratives have innocuous intentions, they can cause significant issues when trying to substantiate broader, more nuanced arguments. With the information posted being highly generalized, it is hard to determine whether the content is based on semblances of the truth or pure fabrication.¹¹

Section 1.3: The scale of harm caused by misinformation:

Beyond the immense magnitude of harm it can cause, the speed and scale of misinformation’s spread underscore the need for social media regulation. According to Pew Research Center research, social media now plays a central role in news consumption. In 2024, fifty-four percent of American adults get their news from social media, underscoring how digital sources have become an essential part of individuals’ news diets.¹²

This is especially relevant when discussing misinformation stemming specifically from the Russian-Ukrainian conflict. With millions of lives displaced, global supply chain disruptions, and shifting global power dynamics, the Russian-Ukrainian conflict is one of the highest-stakes conflicts of recent history. Given that nationalistic sentiments are being fuelled by strong emotions such as pride and fear, online discourse around this conflict becomes emotionally charged. This subconsciously influences individuals to take sides in the conflict, leading to a high degree of politicization on social media platforms.¹³

Even though many other conflicts in recent history have suffered heavy casualties (such as the Yemen or Ethiopian crisis), none have had media coverage on par with that of the Russian-Ukrainian conflict. As such, the Russian-Ukrainian war has been described as the first “social media war” in history.¹⁴ Every development, atrocity, and counter-narrative is broadcast worldwide through social media, with posts and comments providing additional updates on the war. The virality of such content produced by Ukrainians, Russians, and their respective sympathizers makes the situation even more volatile, demonstrating the global reach of the Russian-Ukrainian conflict.

To the global community, the Russian-Ukrainian war is more than a regional conflict. It has the potential to escalate to an international conflagration with superpowers taking sides.¹⁵ However, one must also recognize that the tensions between the two countries have existed long before Russia’s direct invasion of Ukraine in 2022. This means that there has been, and will continue to be, exceptional scrutiny of the Russian-Ukrainian conflict, thereby increasing demand for information and the reach of misinformation surrounding this issue. Therefore, now more than ever, it is crucial to keep misinformation under control to prevent its harmful effects from affecting a significant portion of society that relies on social media for news.

However, this will be challenging given recent discoveries regarding the spread of misinformation. A study conducted by MIT reveals how false news spreads much faster than real news by a substantial margin. Across all categories of information, falsehoods diffused significantly faster, deeper, and farther than factual information.¹⁶ To rationalize these findings, this paper provides two possible explanations.

On one hand, the emotional charge of misinformation may drive audience engagement. Unlike factual information, which should always aim to make a rational point, misinformation leverages the power of strategic narratives to appeal to viewers’ emotions.¹⁷ Effectively, instead of convincing users with logic and facts, misinformation taps into emotions to blur the line between reality and fabrication, thereby facilitating easier message retention. Thus, even if viewers doubt the accuracy of such content, the emotional rhetoric leaves a lasting impression.

Alternatively, it can also be argued that platforms themselves facilitate the spread of misinformation. This is because social media companies are inherently profit-driven. Given that these companies can only achieve their goals by increasing users’ time spent on social media platforms, they would naturally prioritize user engagement. Yet this creates a perverse incentive to share content indiscriminately, regardless of quality or accuracy, because misinformation can be more engaging for the reasons stated earlier.¹⁸ Therefore, this paper posits that the spread of misinformation may also be encouraged at a systemic level.

SECTION 2: QUANTIFYING THE PROBLEM:

Building on the earlier characterization of misinformation, Section 2 explores the different types of misinformation in the status quo. It will also discuss the methods and media used to spread misinformation, as well as other difficulties in regulating the status quo. This will effectively illustrate the problem of misinformation, providing context on its various forms and the public’s response to it in the status quo.

Section 2.1: Types of misinformation:

Misinformation can vary in type and nature. From inaccurate data in viral videos to doctored images on social media posts, there are countless ways in which misinformation has emerged in the Russian-Ukrainian war. Nevertheless, for the sake of discussion, this paper has identified two dominant categories of misinformation: Military-related misinformation and political misinformation. These categories are not merely academic; they encapsulate real events.⁴

On one hand, military-related misinformation deserves particular attention in this conflict. Given the direct military confrontation taking place in Ukraine, it is only logical that misinformation regarding the war efforts starts to spread through social media. Without being physically present in Ukraine, the international community lacks firsthand knowledge about the conflict. Instead, they are forced to piece together scraps of information from social media to better understand the conflict’s developments.⁴ This information vacuum creates fertile ground for the spread of misleading claims about military events. Users may misinterpret footage,

circulate outdated images from unrelated conflicts, or fabricate battlefield developments in an attempt to contribute to online discussions about the war. Additionally, countries themselves may also exploit this uncertainty. Studies have shown that Russia has strategically used social media to amplify misleading narratives about military developments, portraying its forces as liberators while discrediting NATO and Western institutions as imperialist aggressors. In such uncertain environments, individuals often accept misleading claims rather than remain without information. As Greenhill and Oppenheim argue, the absence of direct evidence can lead audiences to adopt unverified narratives as plausible explanations. Once internalised as truth, these narratives may then be reshared by ordinary users, allowing misinformation about military developments to spread widely across social media platforms.

Political misinformation is another category for consideration. Beyond misinformation directly targeting the war efforts in Ukraine, there is also misinformation regarding the context of the conflict. Russia spreads narratives of its country being the heroic victor of World War II, and Ukraine betraying its historical roots by aligning with the West. This narrative becomes even more believable when considering that some of the accusations against Ukraine have proven true. The Azov Brigade, for example, was accused of having Nazi connections given its contentious use of Nazi-related symbols, and members of its unit committing human rights violations.¹⁹ This has led to the rampant spread of misinformation, whether intentionally by the Russian government or unintentionally by members of the public, that the rest of Ukraine is similarly fascist or has Nazi connections. Regardless of intent, this misinformation has tainted Ukrainian leaders with suspicious historical connections.⁴ The scandalous nature of such claims has further spread misinformation. Even though a claim may be supported by dubious sources or fabricated evidence, the initial outrage still garners traction on social media, depicting Ukraine as an illegitimate and morally gray country.

Both political and military misinformation pose a similar threat to society. They tend to promote pro-Kremlin narratives, conveying the consistent narrative that Russia will eventually “liberate” Ukraine, whether it is by exaggerating the overwhelming military resources of the Russians, or tainting the image of Ukrainian leaders through unsavory historical connections.⁴ Thus, these two types of misinformation will be referenced throughout the rest of the research.

Section 2.2: Methods of misinformation:

Even on social media, misinformation can be spread through various methods. Among the most common are repurposed photos and videos, fake eyewitness accounts, and internet trolls. Each of these methods contributes to misinformation on social media in its own way.⁶

Firstly, recontextualized images from unrelated sources are used to support false narratives, misdirecting viewers and leading them to form incorrect conclusions. For instance, a war photo in 2014 was falsely circulated as proof that Ukrainian soldiers injured a twelve-year-old boy in Sloviansk, thereby acting as evidence of Ukrainian aggression in Sloviansk. How-

ever, it was later revealed that the photo was actually a 2013 photo from the Syrian War, with no connection to the supposed Ukrainian aggression.⁶ This method of misinformation is particularly effective due to the well-documented Picture-Superiority Effect (PSE). First discovered by Paivio and Csapo in 1973, PSE is a cognitive bias in which images are more likely to be internalised than text on a screen.²⁰ Even for bite-sized information where both text and image are easily remembered in immediate recall, images were found to be significantly more memorable in delayed recall.²¹ At the same time, unlike other methods of misinformation, which present facts and data to viewers, images serve as visual stimuli that influence viewers' interpretation of the situation. Using contextual clues from the surrounding text, viewers form connections between text and image, under the tacit assumption that images must be relevant to the accompanying text.²² However, when images are recontextualized, their context is misrepresented. Viewers end up with an inaccurate interpretation, thereby propagating misinformation.

Additionally, fake eyewitness accounts lend a false sense of legitimacy, further propagating misinformation.⁶ In 2014, a self-proclaimed pro-Russian emergency doctor, Igor Rozovskiy, reported on Facebook that Ukrainian nationalists did not allow him to help other injured pro-Russian demonstrators after a fire at the Trade Unions House. The story was reposted thousands of times and even translated into different languages before it was realized that Rozovskiy was not an emergency doctor but a dentist. Additionally, his Facebook account was created only immediately before he published that story and was deleted soon after that truth was revealed.⁶ While the truth was eventually revealed, it was not before thousands believed the fabricated incident. What makes this method of misinformation especially problematic is the lack of readily available corroborative evidence. To correct misinformation, the correction needs to be specific enough for the public to understand which aspect of the claim is being refuted.²³ However, in the case of the Russian-Ukrainian War, verifying or refuting statements is challenging due to the conflict's isolated nature. Therefore, so long as the information seems logically congruent, individuals tend to believe the claim, regardless of its actual veracity.

Nevertheless, the most pervasive source of misinformation remains internet trolls. Internet trolls provoke social discord by expressing disagreement toward the subject matter through a combination of insults, jokes, and unhelpful comments.²⁴ Such trolls have various motivations for spreading misinformation, ranging from being excessively opinionated to being state-sponsored agents or simply wanting to seek entertainment. Russia has repeatedly denied the existence of state-sponsored internet trolls, but the discovery of a “troll factory” situated near St. Petersburg suggested otherwise. Regardless of whether the Russian government indeed paid the trolls, their disruptive effect is undeniable. The Guardian reported coordinated attacks on its Russia and Ukraine articles, facing up to forty thousand troll comments per day.⁶ Internet trolls are particularly effective as they rely on quantity, rather than the quality of information. By flooding comment sections, mass-report-

ing posts, and asserting unsubstantiated claims, internet trolls drown out legitimate, factual information, thereby enabling the rapid spread of misinformation.²⁵

Section 2.3: Media of misinformation:

At the same time, this paper acknowledges that not all social media platforms are identical: some platforms, like Telegram, are vastly different from others, such as Instagram or TikTok. Understanding these differences will be crucial, as they have significant implications for social media regulation and will inform policy decisions on how best to regulate misinformation across platforms.

One medium for misinformation is social messaging applications, such as Telegram. The platform has been widely used during the Russian-Ukrainian war to disseminate information and boost morale. Over the last few years, the Russian government has flooded the platform with over three hundred thousand inauthentic comments under the guise of seeking peace and cooperation.²⁶ These comments promote pro-Russian ideology and anti-Ukrainian narratives targeted at Ukraine's occupied territories, fuelling the government's efforts to control the narrative by framing the Russian-Ukrainian war as a "special military operation." Telegram's structure creates favorable conditions for the circulation of misleading information. This is primarily due to end-to-end encryption, which ensures that messages are unreadable by third parties, including fact-checkers who would otherwise be able to track the content being shared.²⁷ As such, the platform's decentralized communication model allows messages, images, and videos to spread rapidly across channels with minimal oversight. Nevertheless, this medium for misinformation has the lowest impact. This is because the spread of misinformation through this medium is reliant on the viewer's interest in the subject. Unlike algorithm-based platforms, Telegram requires users to actively join channels or group chats to receive information.²⁸ Therefore, if an informed individual recognizes the risk of misinformation from a specific source, they will consciously choose not to join the channel, thereby breaking the chain of misinformation.

Another medium for misinformation is media-sharing platforms like Instagram and TikTok. These platforms are highly visual and prioritize user engagement through their algorithm-driven recommendation system. Additionally, studies have shown that user engagement is prolonged by emotionally charged, shocking, or even polarising elements.²⁹ Thus, it is unsurprising that there have been numerous instances in which these platforms have been used to post aestheticized war footage with misleading captions and edited clips from interviews with government officials. Investigations by BBC Verify have revealed videos on TikTok that misleadingly portray Ukrainian President Volodymyr Zelensky as profiting from the war while remaining detached from the suffering of Ukrainian civilians.³⁰ The spread of misinformation through this medium is especially detrimental, as the algorithm-based recommendation system is highly likely to continue exposing users to misleading political content, even without seeking it

out, so long as they have watched similar content on the Russian-Ukrainian conflict.

Another medium for misinformation is discussion forums, including Twitter (now known as X) and Reddit. These platforms offer a conversation-like format for individuals to share opinions, view links, and join communities based on various topics.³¹ In the context of the Russian-Ukrainian conflict, this manifests itself in users sharing absurd claims such as the Russian-Ukrainian War being fake.³² Unlike the other two mediums, misinformation is not directly forwarded to users or recommended by the algorithm. Instead, users actively choose the topic of discussion and interact with other users on the platform. Consequently, emotionally charged or politically polarising claims may gain visibility if they resonate with a particular online community. Over time, repeated exposure to these claims within a discussion forum can contribute to the formation of collective narratives that reinforce misinformation, as long as enough community members upvoted them.³³

Section 2.4: The current response to misinformation:

Given the variety of misinformation types, methods, and media, how does society currently respond to the misinformation crisis?

Firstly, bad actors increasingly weaponize social media to spread misinformation. These actors recognize how social media enables them to create spectacle rapidly, while drawing emphasis away from the lack of veracity of some of these claims.¹⁹ As a result, misinformation has become more pervasive than ever. These actors continue to blur the boundary between truth and fiction, creating a post-truth environment where emotional appeal dominates social media.

As for regular social media users, a growing proportion of society is cognizant of the pro-Kremlin disinformation campaign and remains largely skeptical of information from Russian sources.¹⁸ Given how this is not the first wave of misinformation, communities are already beginning to tackle misinformation through crowdsourced projects. For instance, StopFake was launched in 2014 in Kyiv to combat misinformation across various sources, including social media.⁶ Although it started as a small grassroots initiative to debunk, it has since been joined by numerous professionals, growing into a vibrant community and symbolizing a new wave of citizen accountability in digital spaces. Although efforts thus far have been largely uncoordinated, the mere presence of such crowdsourced projects indicates an increasing desire to weed out misinformation.⁶ At the same time, more vulnerable members of society continue to fall prey to misinformation. Based on research by Erlich and Garner, individuals who are less politically astute, identify with Russian ethnicity and language, and are demographically older tend to be more susceptible to misinformation. Parallels can be drawn to the status quo, with youths forwarding misleading reels on Instagram, or elders sending chain letters in their family WhatsApp chats.⁴

In all cases, it is evident that some form of intervention is necessary, whether by the government or by social media companies. When left to their own devices, social media users often fall victim to misinformation. Even within the subset

of users that recognize the importance of tackling misinformation, crowdsourcing projects are currently uncoordinated and ineffective.

SECTION 3: REGULATION IN THE STATUS QUO:

The question of social media regulation has sparked immense discussion and controversy. Striking the right balance between protecting freedom of speech and curbing harmful misinformation requires both platform self-regulation and government-led regulation. Although both approaches have yielded some results, neither is without shortcomings. Section 3 will explore the state of regulation in these two areas, drawing on case studies and regulatory policies worldwide.

Section 3.1: Self-regulation:

Social media companies are keenly aware of the misinformation crisis that plagues their platforms. As a result, companies engage in self-regulation, voluntarily identifying and removing misinformation without direct state intervention. Self-regulation often involves internal rules and regulations adopted by companies on their own accord. For example, this could include protocols for scrutinizing advertisement placements, services, and fact-checking mechanisms.³⁴ At the same time, self-regulation is often guided by international frameworks and codes of conduct. In line with the EU Code of Practice on Disinformation being revised in response to concerns about Russian propaganda during the war, social media platforms such as YouTube and TikTok committed to more accountable content moderation in line with the code.⁴⁹ This policy builds on years of past self-regulatory efforts by social media companies, streamlining such policies on an international scale.

Social media companies have an active incentive to self-regulate for various reasons. Some platforms conduct content moderation out of a sense of moral responsibility.³⁵ For other social media companies like YouTube and Twitter, establishing free-speech norms is crucial to maintaining their identities. As such, they have established community guidelines on misinformation that align with the company's values and goals.³⁶ Even for social media companies that may not have a strong orientation toward managing content quality, content moderation will be necessary to prevent backlash from allowing misinformation to spread unchecked across the platform.⁵² More compelling, however, are the economic incentives. Although corporate responsibility exists, the main reason most companies continue to regulate misinformation is the risk of losing potential profits. This is because allowing misinformation to propagate on the platform can lead to the alienation of users and advertisers, especially given the highly politicized and divisive nature of the Russian-Ukrainian war. This can lead to direct revenue losses when users leave the platform.⁵² Therefore, to retain user engagement and protect the company's profitability, social media companies have an incentive to implement an effective content moderation system.

There are many examples of self-regulation by social media companies. For instance, when companies identify misinformation, most social media algorithms actively reduce its visibility by not recommending it to users, thereby limiting its spread

and disincentivizing users from posting it.³⁷ In more serious cases, companies also have the power to swiftly and comprehensively take down malicious content.³⁸ Nevertheless, beyond reactionary measures, companies also implement preventive safeguards to reduce the spread of misinformation. Through partnerships with fact-checkers across languages, social media platforms ensure quick verification of content flagged as potentially misleading. Companies like TikTok have even launched media literacy campaigns across Europe to help users, especially younger audiences, critically assess content, identify misinformation, and access reliable sources.³⁹

Given the immense scale of misinformation on social media, platforms cannot have human moderators screen content before it is published. Instead, most companies utilize automated moderation systems to identify misinformation *ex ante*, helping prevent its spread before it reaches the public.⁴⁰ However, automated moderation runs the risk of over-censorship, especially when a post blurs the line between what is acceptable and what is not. Only the most harmful forms of misinformation are immediately removed by the platform, while less severe cases are instead made less visible to users.

From these measures, this paper posits an implicit hierarchy of misinformation in the eyes of social media companies. For outright harmful rhetoric likely to gain traction, social media companies may prioritize moderation through automated algorithms, ensuring that the post has zero reach. However, for more nuanced and contentious issues, the social media company may allow the content to circulate, relying on human moderators to sift it for misinformation.⁴¹

Unfortunately, the current system of self-regulation still has significant limitations. Across the industry, platforms often have different interpretations of the hierarchy of misinformation, leading to varying degrees of content moderation. Some companies choose to err on the side of caution, removing most content that they deem as misleading or malicious. Other companies prefer to disregard controversial cases to prevent backlash from wrongly labeling content as misinformation.⁴²

Even within the platform, content moderation can be inconsistent. One striking example is Twitter's treatment of US President Donald Trump's tweets. When President Trump made the controversial tweet "when the looting starts, the shooting starts" in 2020, the company did not restrict it, and no further action was taken against his account. This matter has already raised questions about whether Trump should be prosecuted, given that public officials should be held to a higher standard. However, when another Twitter account copied the same message, the account was promptly suspended within three days.⁴³ This double standard reveals a lack of accountability in self-regulation: Social media companies often regulate misinformation based on context, allowing it to continue spreading.

At the same time, all social media platforms are driven by profit. Social media companies earn revenue by maximizing viewer engagement. Therefore, companies have an incentive to provide users with content they want to see, regardless of whether the information is true or false. Social media algorithms are trained to feed us with information that reinforces

what we already believe, thereby increasing our satisfaction by limiting our exposure to new, disruptive viewpoints.⁴⁴ Moreover, to maximize profit, these companies have an incentive to restrict content regulation. By adding more content to the platform, the company earns more from advertising revenue. Therefore, sole reliance on social media companies to regulate misinformation is also insufficient, as these companies benefit from the spread of misinformation insofar as content gains traction and generates advertising revenue.

Section 3.2: State-run regulation:

State-run regulation, on the other hand, is vastly different. Instead of directly labeling information as misinformation, governments like the EU and UK have created frameworks to regulate misinformation from the Russian-Ukrainian war. Unlike social media companies, these governments do not have access to social media algorithms or back-end systems to remove misinformation from social media platforms. Thus, state-run regulation rarely addresses individual cases of misinformation, as it cannot directly control the content available on its platforms. Instead, they use a variety of methods to enforce regulations set by social media platforms.⁴⁵ For instance, governments may enact legislation compelling social media companies to promptly identify and remove misinformation.⁴⁶ By holding social media companies liable for misinformation on their websites, state-run regulation compels them to implement procedural measures to address it.

The government has an active incentive to implement state-run regulation for several reasons. From a principle perspective, the government has a duty to ensure democratic participation through all media, including cyberspace. Therefore, especially when misinformation directly targets other individuals and inhibits their ability to share their views freely online, the government has a moral responsibility to intervene.⁴⁷ From a practical perspective, current attempts at self-regulation are insufficient. Even though each social media platform regulates misinformation independently, state-run regulation is still necessary to establish a common standard for platforms to work toward. This standardization helps social media platforms to regulate misinformation within a set framework, rather than continuously treading the line between over-censorship and allowing misinformation to run rampant.⁴⁷ The government is also incentivized to regulate due to the unique context of the Russian-Ukrainian War. As the first “social media war” in history,⁴⁸ Ukraine recognizes the need to keep its citizens well-informed by combating the threat of misinformation. As much as the Russian-Ukrainian War is a physical battle of arms, it is also a battle of narratives.

State-run regulation is one of the most effective ways to tackle Pro-Kremlin misinformation. This is because government policies, unlike self-regulation, are legally binding. For instance, when Russia first invaded Ukraine in 2022, the EU and the United Kingdom took drastic action by banning the Russian state broadcaster RT in their respective territories.¹ This limited European citizens’ exposure to Russian propaganda, thereby reducing the spread of pro-Russian misinformation. This demonstrates that the government has the legal mandate

to regulate misinformation by any means necessary, even if that means complete censorship of information channels. When direct government intervention is not required, state-run regulation can also provide an impetus for social media companies to act. Through advisories or direct legislation, state-run regulation acts as a strong signal of the government’s attitude toward this serious issue. This can push social media companies to review user content more regularly and thoroughly, in line with the guidelines and direction provided by the state-run regulation.⁴⁹

However, implementing broad-level state-run regulation alone is insufficient. By using frameworks to shape expectations, governments can only offer implementation suggestions.⁵⁰ This includes examples such as the European Commission’s voluntary Code of Practice on Disinformation, which sets out commitments by social media platforms to combat fake news.⁵¹ While such frameworks are a step in the right direction, they lack legally binding enforcement mechanisms. Thus, they are unable to hold social media companies to account effectively.

On the other hand, the government can wield its executive power and implement direct state-run regulation. This typically manifests as governments compelling social media platforms to remove content deemed misleading, though the process remains subject to judicial review. However, this raises concerns about excessive control over the flow of information. When governments are granted the authority to determine what constitutes misinformation, critics argue that such powers may risk suppressing legitimate political expression. This emphasizes the importance of judicial oversight, helping ensure that state-run regulation remains accountable. Ultimately, even though stronger regulatory frameworks may help curb the spread of misleading information, excessive intervention risks suppressing legitimate political discourse. As Stefan Theil argues, even in this extreme case of state-run regulation, governments must navigate a balance between safeguarding against misinformation while preserving the democratic principle of free speech.⁵²

Most countries currently fall somewhere between the two opposing approaches: broad-level regulation and direct state-run regulation. Vietnam, for instance, has reached an agreement with Facebook to expedite content-removal requests from Vietnamese government ministries. Turkey has sought to compel social media platforms like Twitter and YouTube to remove defamatory content about Mustafa Kemal Atatürk, a renowned national figure often hailed as the founder of modern Turkey, but with little success.⁵³ Additionally, the United States is even proposing new legislation to crack down on social media bots and online political falsehoods.⁵⁴

In the case of the Russian-Ukrainian war, both broad-level and direct state-run regulation have been employed to combat misinformation. On a broader level, the European External Action Service’s East StratCom Task Force launched the EU-vsDisinfo initiative, intending to issue guidelines and reports in response to the Russian government’s disinformation campaign. This influences platform behavior and public awareness, helping to reduce the spread of pro-Kremlin disinformation without legal compulsion.⁵⁵ At the same time, the EU has also taken concrete steps to directly remove pro-Kremlin misinform-

mation sources. Following the invasion of Ukraine in 2022, the EU banned pro-Kremlin news outlets RT and Sputnik across all member states, regardless of content.⁵⁶

This further broadens the concern about state-run regulation. Beyond accounting for differing treatments toward the gray area of misinformation, we must also consider the possibility of a government outright censoring clearly valid information. Some sacrifices should be made when considering freedom of speech, but it must definitely not become a weapon of oppression in the hands of the government. Going one step further, governments themselves can serve as sources of misinformation. In the Russian-Ukrainian war, Russia has repeatedly framed the invasion of Ukraine as a “special military operation” aimed at protecting Russian-speaking populations. In such cases, state-run regulation is ineffective, since the government can freely spread false narratives and conveniently decide what constitutes misinformation. State-run regulation provides a legitimate means for bad actors to censor information that contradicts their narrative, labeling it “misinformation.” This makes it difficult to hold governments accountable, especially when they have a vested interest in spreading misinformation.

SECTION 4: CONCLUSION:

With a better understanding of the misinformation crisis and the current global response, it is apparent that more must be done to prevent the spread of harmful misinformation. Although misinformation continues to plague the international community, current efforts to regulate misinformation have been disorganized and ineffective. Section 4 aims to outline a path forward, offering preliminary guidance on improving the current status quo.

Section 4.1: Lessons from the past:

Based on Sections 1 to 3, it is apparent that the misinformation crisis is multifaceted. As evident in the Russian-Ukrainian war, misinformation has taken many forms, communication methods, and dissemination platforms.

Ranging from doctored battlefield videos being uploaded onto Telegram to misleading political narratives spread through Reddit forums, there are endless ways for misinformation to manifest.⁵⁷ At the same time, misinformation spreads farther, deeper, and more widely than the truth. The culmination of these factors has created a systemic misinformation epidemic. Platforms have implemented algorithmic filters and fact-checkers, but they continue to lag behind the viral falsehoods that plague the internet.¹⁷ State-run regulation is also being explored, but its heavy-handed nature makes it difficult for countries to implement effectively without being accused of abusing their executive power. To move forward, society cannot simply repeat ineffective policies. We must create a robust strategy that addresses the multifaceted nature of the issue.

There are two main ways to address the misinformation crisis surrounding the Russia-Ukraine war. The first approach places the blame for misinformation squarely on platform users. This approach treats the spread of misinformation on social media as a reflection of society and, by extension, of human behavior. During the Russian-Ukrainian war, some users

shared misinformation to gain attention; others chose not to fact-check before sharing. Therefore, it does not matter if the system is changed, because people will continue to spread misinformation.⁵⁹ The second approach involves treating social media as the source of the issue. Given that social media algorithms are built for sensationalist headlines and short-form content, they naturally favored emotionally charged narratives about the death of innocent civilians or the pain of wounded soldiers, rather than logical reasoning and detailed explanations. As such, social media platforms are mainly responsible for creating a systemic framework that prioritizes misinformation. Hence, without greater algorithmic transparency, efforts to change people's mindset will be ineffective in curbing the spread of misinformation.¹²

However, neither approach alone tells the whole story. The former approach ignores algorithms that systematically recommend misinformation.⁷⁶ Without social media algorithms, malicious users would be able to amass only hundreds of views, thereby exponentially slowing the spread of misinformation. On the other hand, the latter approach fails to account for why individuals find misinformation memorable. Without users clicking on misinformation that pops up in their feeds, social media algorithms would also be unable to catalyze its spread.⁶⁰ As such, treating the misinformation crisis as purely a matter of human behavior or algorithmic bias will lead to a myopic outcome. Instead, this paper purports that the proper solution could stem from a combination of both approaches. Even though some users have malicious intent, social media places these perspectives on a pedestal, amplifying misinformation that does not represent society as a whole.

As evident in Section 3, self-regulation and state-regulation are both limited in isolation but undeniably crucial for curbing the spread of misinformation. To address this issue, a hybrid approach that accounts for the unique strengths and limitations of both actors is necessary. Self-regulation can be used as the first line of defense against misinformation. This is because social media companies have the exclusive ability to implement technical interventions, such as deprioritising misleading posts and flagging suspicious content.⁶¹ With complete access to the algorithm, social media companies will be able to adapt misinformation regulations to swiftly overcome platform-specific challenges. This can be complemented by state-run regulation in more critical situations, where the potential for public harm is immediate and severe. In these cases, the government must rely on its executive powers to regulate misinformation related to high-stakes issues, such as false claims about public health data or emergency-response news. Nevertheless, the government should also exercise restraint in cases involving national sensitivities. As evidenced through the case study of Mustafa Kemal Atatürk in Turkey, governments must not compel social media companies to remove uncomfortable content from their platforms. Instead, it must be keenly aware of the risk that state intervention crosses the line into authoritarian overreach.⁶²

However, the isolated efforts of individual countries may be insufficient to solve the misinformation crisis. With its international reach and extensive viewership, social media platforms have effectively become superstate actors who transcend geo-

graphical boundaries.⁶³ As such, no single nation can exercise complete control over content posted on social media platforms, nor can it unilaterally enforce accountability on platforms beyond its jurisdiction. To address this power imbalance, the next logical step is for the international community to move toward multilateral cooperation and international treaties. By establishing global standards, countries can create a standardized regulatory framework that all social media companies must adhere to. This allows for more straightforward guidelines on the scope of regulation and on how to hold social media companies accountable through self-regulation agreements and state-regulated directives. Even if multilateral forums seem unproductive, they still offer an essential avenue for discourse.⁶⁴ This allows countries to voice regulation concerns in a cordial setting, preventing disagreements from escalating into global tensions or censorship wars.

Section 4.2: Envisioning the future:

Going one step further, this paper argues that a long-term solution requires not just an effective response to misinformation after it has spread, but also an emphasis on preventing its spread on social media platforms in the first place. This is because it is exponentially harder to reverse the effects of misinformation after it has gone viral. Therefore, fact-checking and takedown efforts, while necessary, are too late to fully neutralize the harms caused by misinformation. Incorporating a preventive framework in the future can reduce the burden on reactive regulation systems.⁵² There are a few key characteristics of this proposed long-term solution.

Firstly, social media companies must prioritize accuracy over sensationalism. While it is unavoidable for social media algorithms to promote content based on popularity, companies must also take steps to verify content accuracy. For instance, social media companies could implement a tiered verification system that involves expedited fact-checking for viral posts. This means that, even as social media algorithms prioritize engaging content, there will be a parallel mechanism to ensure content accuracy. This prevents misinformation from spreading by ensuring that a heightened sense of responsibility over content accuracy accompanies the extensive reach of social media.⁶⁵

Secondly, users must evolve from passive consumers to active stakeholders. While the government and social media companies may enact joint regulation through state-level and international-level treaties, users must also collaborate with both stakeholders. On the one hand, social media users must work with companies to effectively moderate social media platforms. One good example of this is TikTok's efforts to collaborate with researchers and fact-checkers. By partnering with users in all official EU languages, the company ensures swift responses to international misinformation reports.⁶⁶ On the other hand, social media users must also be receptive toward the government's misinformation education efforts. Through which, users will be incentivized to recognize and call out potential instances of misinformation.

Lastly, this paper also reveals the need for greater transparency and accountability across all stakeholders. At its core,

the misinformation crisis cannot be solved without genuine collaboration between stakeholders. Social media companies must strive for complete transparency about their algorithms. Through regular accountability reports, third parties can hold social media companies to account, ensuring that they take adequate measures to combat harmful content on their platforms. Additionally, the need for governmental transparency is twofold. While the government should be transparent about how it chooses to regulate and enforce platform obligations, it must also be transparent about its decision-making process regarding government censorship.⁶⁷

The long-term solution described in this section is merely a guide to the path forward. It serves as guidance for governments, given the current landscape of the misinformation crisis. More importantly, however, social media companies must remain open to new developments in the misinformation crisis while remaining unwavering in their support for regulation. Ultimately, the fight against misinformation cannot be merely a series of reactive measures against falsehoods. Instead, the collective efforts of all stakeholders must be geared toward cultivating a proactive digital environment that calls out misinformation and embraces the truth.

■ Acknowledgments

I would like to express my sincere gratitude to Fernando Loayza-Jordan, J.S.D. candidate at Yale Law School, for his patient mentorship and invaluable guidance throughout my research journey.

■ References

1. Dean Jackson and Jon Bateman, "Countering Disinformation Effectively: An Evidence-Based Policy Guide," Carnegie Endowment for International Peace, accessed October 21, 2025, <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en>.
2. Evelyn Douek, "Content Moderation as Systems Thinking," *Harvard Law Review* 136, no. 2 (2022).
3. Dan Ciuriak, "The Role of Social Media in Russia's War on Ukraine," SSRN Scholarly Paper no. 4078863 (Social Science Research Network, April 8, 2022), <https://doi.org/10.2139/ssrn.4078863>.
4. Aaron Erlich and Calvin Garner, "Is Pro-Kremlin Disinformation Effective? Evidence from Ukraine," *The International Journal of Press/Politics* 28, no. 1 (2023): 5–28, <https://doi.org/10.1177/19401612211045221>.
5. Joshua A. Tucker *et al.*, "From Liberation to Turmoil: Social Media and Democracy," *Journal of Democracy* 28, no. 4 (2017): 46–59, <https://doi.org/10.1353/jod.2017.0064>.
6. Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press, 2018), <https://doi.org/10.12987/9780300235029>.
7. Kelcy Mathurin, "Misinformation & Disinformation," EN 101: The Process of Writing, Caldwell University Jennings Library, February 26, 2026, <https://libguides.caldwell.edu/EN101/misinformation-disinformation>.
8. Megan Kurten, "Why We Fight for Fractured Truths – How Misinformation Fuels Political Violence in Democracies," *MediaWell*, January 9, 2025, <https://mediawell.ssrc.org/research-reviews/why-we-fight-forfractured-truths-how-misinformation-fuels-political-violence-in-democracies/>.

9. Stefania Farace *et al.*, "Standing Out While Fitting In: Visual Design of Text Overlays in Social Media Communication," *Journal of Marketing*, February 12, 2025, <https://doi.org/10.1177/00222429251322773>.
10. Shiqi Yi *et al.*, "Is Information Evaluated Subjectively? Social Media Has Changed the Way Users Search for Medical Information," *Digital Health* 10 (September 2024), <https://doi.org/10.1177/20552076241259039>.
11. Sierra L. Lawson and Steven W. Ramey, "Sourcing Stereotypes: Constructing and Challenging Simplified Knowledge," *Culture and Religion* 19, no. 4 (2018): 416–34, <https://doi.org/10.1080/14755610.2018.1505759>.
12. Pew Research Center, "Social Media and News Fact Sheet," *Pew Research Center*, September 17, 2024, <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>.
13. Gianella Pena Barbery, "The Role of Communication in 21st Century Nationalism: A Tool of Influence and Manipulation," *Identity Hunters*, April 30, 2025, <https://identityhunters.org/2025/03/30/the-role-of-communication-in-21st-century-nationalism-a-tool-of-influence-and-manipulation>.
14. Dan Ciuriak, "Social Media Warfare Is Being Invented in Ukraine," *Centre for International Governance Innovation*, n.d., accessed April 21, 2025, <https://www.cigionline.org/articles/social-media-warfare-is-being-invented-in-ukraine/>.
15. Malaka Gharib, "Not Every War Gets the Same Coverage as Russia's Invasion — and That Has Consequences," *Ukraine Invasion — Explained, NPR*, March 4, 2022, <https://www.npr.org/sections/goatsandsoda/2022/03/04/1084230259/not-every-war-gets-the-same-coverage-as-russias-invasion-and-that-has-consequenc>.
16. Peter Dizikes, "Study: On Twitter, False News Travels Faster than True Stories," *MIT News*, March 8, 2018, <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>.
17. Irina Khaldarova and Mervi Pantti, "Fake News: The Narrative Battle over the Ukrainian Conflict," *Journalism Practice* 10, no. 7 (2016): 891–901, <https://doi.org/10.1080/17512786.2016.1163237>.
18. Gizem Ceylan, "How Social Media Rewards Misinformation," *Yale Insights*, March 31, 2023, <https://insights.som.yale.edu/insights/how-social-media-rewards-misinformation>.
19. Al Jazeera, "Profile: Who Are Ukraine's Far-Right Azov Regiment?," *Al Jazeera*, March 1, 2022, accessed October 5, 2025, <https://www.aljazeera.com/news/2022/3/1/who-are-the-azov-regiment>.
20. Abd Allah Aljalabneh, "Visual Media Literacy: Educational Strategies to Combat Image and Video Disinformation on Social Media," *Frontiers in Communication* 9 (November 2024), <https://doi.org/10.3389/fcomm.2024.1490798>.
21. Terry L. Childers and Michael J. Houston, "Conditions for a Picture-Superiority Effect on Consumer Memory," *Journal of Consumer Research* 11, no. 2 (1984): 643–54, <https://doi.org/10.1086/209001>.
22. Eryn J. Newman and Norbert Schwarz, "Misinformed by Images: How Images Influence Perceptions of Truth and What Can Be Done about It," *Current Opinion in Psychology* 56 (April 2024): 101778, <https://doi.org/10.1016/j.copsyc.2023.101778>.
23. Toby Prike and Ullrich K. H. Ecker, "Effective Correction of Misinformation," *Current Opinion in Psychology* 54 (December 2023): 101712, <https://doi.org/10.1016/j.copsyc.2023.101712>.
24. Evita March and Jessica Marrington, "A Qualitative Analysis of Internet Trolling," *Cyberpsychology, Behavior, and Social Networking* 22, no. 3 (2019): 192–97, <https://doi.org/10.1089/cyber.2018.0210>.
25. Flemming Splidsboel Hansen, "Trolls in Your Feed: Russian Disinformation," *DIIS*, 2019, accessed October 5, 2025, <https://www.diis.dk/en/trolls-in-your-feed-russian-disinformation>.
26. Atlantic Council, "Digital Occupation: Inside Russia's Telegram Battle in Ukraine," July 16, 2025, <https://www.atlanticcouncil.org/event/digital-occupation-inside-russias-telegram-battle-in-ukraine/>.
27. BBC, "Ukraine Bans Telegram Use on State-Issued Devices," September 21, 2024, <https://www.bbc.com/news/articles/c78dwepw95do>.
28. Telegram, "Telegram Channels," accessed October 6, 2025, <https://telegram.org/tour/channels>.
29. Zicheng Cheng and Yanlin Li, "Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos," *Social Science Computer Review* 42, no. 1 (2024): 201–23, <https://doi.org/10.1177/08944393231178603>.
30. BBC, "Ukraine War: How TikTok Fakes Pushed Russian Lies to Millions," December 15, 2023, <https://www.bbc.com/news/world-europe-67687449>.
31. Katie Elson Anderson, "Ask Me Anything: What Is Reddit?," *Library Hi Tech News* 32, no. 5 (2015): 8–11, <https://doi.org/10.1108/LHTN-03-2015-0018>.
32. Alistair Coleman and Shayan Sardarizadeh, "Ukraine War: Viral Conspiracy Theories Falsely Claim the War Is Fake," *BBC*, February 27, 2023, <https://www.bbc.com/news/world-europe-64789737>.
33. Holly Haney, "Systems of Upvoting and Their Implications (Term Paper)," *Medium*, December 16, 2015, <https://hollyhaney.medium.com/systems-of-upvoting-and-their-implications-term-paper-1954a004e916>.
34. Ekin Balkan and Sinan Ülgen, "The Role of Self-Regulation and Regulation in Combating Disinformation," *EDAM*, 2023, https://edam.org.tr/uploads/yukleme_resim/Report-The-Role-of-Self-Regulation-and-Regulation-in-Combating-pdf.
35. Jeffrey Howard, "The Ethics of Social Media: Why Content Moderation Is a Moral Duty," *Journal of Practical Ethics* 11, no. 2 (2024), <https://doi.org/10.3998/jpe.6195>.
36. Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech," *Harvard Law Review* 131, no. 6 (2018): 1599–670.
37. Tarleton Gillespie, "Do Not Recommend? Reduction as a Form of Content Moderation," *Social Media + Society* 8, no. 3 (2022): 20563051221117552, <https://doi.org/10.1177/20563051221117552>.
38. Paul M. Barrett, "The Supreme Court Protects Social Media Platforms' Ability to Filter Out Harmful Content," *NYU Stern Center for Business & Human Rights*, n.d., accessed October 6, 2025, <https://bhr.stern.nyu.edu/quick-take/the-supreme-court-protects-social-media-platforms-ability-to-filter-outharmful-content/>.
39. TikTok, "Protecting the Integrity of TikTok during the Czech Parliamentary Elections," *TikTok Newsroom*, n.d., accessed October 6, 2025, <https://newsroom.tiktok.com/protecting-the-integrity-of-tiktok-during-the-czech-parliamentary-elections?lang=en-150>.
40. Alessia Zornetta, "Online Misinformation: Improving Transparency in Content Moderation Practices of Social Media Companies" (LL.M. thesis, McGill University, 2022), accessed October 21, 2025, <https://escholarship.mcgill.ca/concern/theses/rf55zd782>.
41. Zornetta, "Online misinformation."
42. Thomas James Vaughan Williams *et al.*, "Policy vs Reality: Comparing the Policies of Social Media Sites and Users' Experiences, in the Context of Exposure to Extremist Content," *Behavioral Sci-*

- ences of Terrorism and Political Aggression* 17, no. 1 (2025): 110–27, <https://doi.org/10.1080/19434472.2023.2195466>.
43. Donie O’Sullivan, “Mark Zuckerberg Silent as Trump Uses Facebook and Instagram to Threaten ‘Looting’ Will Lead to ‘Shooting,’” CNN, May 29, 2020, <https://www.cnn.com/2020/05/29/tech/facebook-violence-trump>.
 44. Harvard Law School, “The Changing Ecosystem of News and Challenges for Freedom of the Press,” *Harvard Law School*, 2018, accessed October 21, 2025, <https://hls.harvard.edu/bibliography/the-changingecosystem-of-news-and-challenges-for-freedom-of-the-press/>.
 45. Ben Sperry, *Knowledge and Decisions in the Information Age: The Law & Economics of Regulating Misinformation on Social-Media Platforms* (International Center for Law & Economics, 2024), <https://laweconcenter.org/resources/knowledge-and-decisions-in-the-information-age-the-law-economics-ofregulating-misinformation-on-social-media-platforms/>.
 46. Howard Tumber and Silvio Waisbord, eds., *The Routledge Companion to Media Disinformation and Populism*, (Routledge, 2021), accessed October 21, 2025, <https://www.routledge.com/The-Routledge-Companion-to-Media-Disinformation-and-Populism/Tumber-Waisbord/p/book/9780367704919>.
 47. Robyn Sanders and Ashleigh Maciolek, “The Government Shouldn’t Be Barred from Countering False Information,” Brennan Center for Justice, September 12, 2025, <https://www.brennancenter.org/ourwork/analysis-opinion/government-shouldnt-be-barred-countering-false-information>.
 48. Ciuriak, “Social Media Warfare Is Being Invented in Ukraine.”
 49. Amélie Heldt, “Reading between the Lines and the Numbers: An Analysis of the First NetzDG Reports,” *Internet Policy Review* 8, no. 2 (2019), published June 12, 2019, <https://doi.org/10.14763/2019.2.1398>.
 50. Paul Henman, “The Big Question,” UQ Contact Magazine, 2019, accessed October 21, 2025, <https://stories.uq.edu.au/contact-magazine/2019/big-question-social-media-regulation/>.
 51. European Commission, “The 2022 Code of Practice on Disinformation | Shaping Europe’s Digital Future,” accessed October 21, 2025, <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>.
 52. Stefan Theil, “Private Censorship and Structural Dominance: Why Social Media Platforms Should Have Obligations to Their Users under Freedom of Expression,” *The Cambridge Law Journal* 81, no.3 (2022): 645–672, <https://doi.org/10.1017/S0008197322000484>.
 53. Daphne Keller, “Internet Platforms: Observations on Speech, Danger, and Money,” Hoover Institution, accessed October 21, 2025, <https://www.hoover.org/research/internet-platforms-observations-speech-dangerand-money>.
 54. Ariel Shapiro, “Democratic Sen. Warner Has a New Policy Paper with Proposals to Regulate Big Tech Companies,” CNBC, July 30, 2018, <https://www.cnbc.com/2018/07/30/sen-warner-proposes-20-ways-toregulate-big-tech-and-radically-change.html>.
 55. EEAS, “EUvsDisinfo,” European External Action Service, n.d., accessed November 23, 2025, https://www.eeas.europa.eu/delegations/china/euvsdisinfo_en?s=166.
 56. Gergely Ferenc Lendvai, “Media in War: An Overview of the European Restrictions on Russian Media,” *European Papers - A Journal on Law and Integration* 2023 8, no. 3 (2024): 1235–45, <https://doi.org/10.15166/2499-8249/715>.
 57. Clothilde Goujard, “YouTube’s Algorithm Pushes Hateful Content and Misinformation: Report,” *Politico*, July 7, 2021, <https://www.politico.eu/article/mozilla-firefox-report-youtube-algorithm-pushes-hateful-contentmisinformation/>.
 58. Robin Ramcharan, *Fake News and Elections in Southeast Asia: Impact on Democracy and Human Rights*, 1st ed. (Routledge, 2022), <https://www.routledge.com/Fake-News-and-Elections-in-Southeast-Asia-Impact-on-Democracy-and-Human-Rights/Gomez-Ramcharan/p/book/9781032379821>.
 59. Tucker *et al.*, “From Liberation to Turmoil.”
 60. Franziska Zimmer *et al.*, “Fake News in Social Media: Bad Algorithms or Biased Users?,” Korea Institute of Science and Technology Information, June 30, 2019, <https://doi.org/10.1633/JISTaP.2019.7.2.4>.
 61. Corinne Tan, “Regulating Disinformation on Twitter and Facebook,” *Griffith Law Review* 31, no. 4 (2022): 513–36, <https://doi.org/10.1080/10383441.2022.2138140>.
 62. Tech Against Terrorism, “The Online Regulation Series | Turkey,” October 23, 2020, accessed November 7, 2025, <https://techagainstterrorism.org/news/2020/10/23/the-online-regulation-series-turkey>.
 63. Suvojit Bandopadhyaya, “Transcendence through Social Media,” *Journal of Media and Communication Studies* 8, no. 3 (2016): 25–30, <https://doi.org/10.5897/JMCS2015.0469>.
 64. UNESCO, “Online Disinformation: UNESCO Unveils Action Plan to Regulate Social Media Platforms,” November 6, 2023, accessed November 7, 2025, <https://www.unesco.org/en/articles/online-disinformationunesco-unveils-action-plan-regulate-social-media-platforms>.
 65. Otávio Vinhas and Marco Bastos, “Fact-Checking Misinformation: Eight Notes on Consensus Reality,” *Journalism Studies* 23, no. 4 (2022): 448–68, <https://doi.org/10.1080/1461670X.2022.2031259>.
 66. Caroline Greer, “TikTok’s Continued Commitment to Combating Disinformation Under the EU Code of Practice,” September 24, 2024, <https://newsroom.tiktok.com/tiktoks-continued-commitment-to-combatingdisinformation-under-the-eu-code-of-practice?lang=en-150>.
 67. Mark MacCarthy, “Transparency Is Essential for Effective Social Media Regulation,” *Brookings*, January 11, 2022, <https://www.brookings.edu/articles/transparency-is-essential-for-effective-social-media-regulation/>.

■ Author

Jerome Ong is a student at Anglo-Chinese School (Independent) in Singapore, where he is enrolled in the six-year Integrated Programme culminating in the International Baccalaureate Diploma (2021–2026), with a focus on Humanities. He has demonstrated consistent academic excellence, achieving predicted IB scores of 44–45 points. In recognition of both his scholastic achievements and personal character, he was awarded the Ministry of Education Pre-University Scholarship.

Jerome has a strong interest in governance and public policy. This is reflected in his active participation in Model United Nations conferences, where he has received distinctions including Best Delegate and Outstanding Delegate. He currently serves as President of the school’s debate club and has earned multiple accolades that highlight his strengths in critical thinking, argumentation, and public advocacy.

Beyond academics, Jerome has gained practical exposure through legal internships at leading law firms, including Allen & Gledhill, and through grassroots community service. These experiences have deepened his understanding of the interplay between law, policymaking, and societal outcomes. In addition, he authored an anthology exploring public discourse, rhetoric, and policy influence, which was published on 4 February 2026.