Uncontrolled Social Media Sites: Vectors For Misinformation

By Alex Robic

Social media has enabled widespread communication and connections between individuals from across the world, and has been credited with enabling some of the most influential events of the 21st century such as the Bronze Night in Estonia or the Arab Spring. US adversaries across the world have quickly picked up on how influential such platforms could be and how to weaponize them for nefarious purposes (Sanger; Solen) to the point that the US House Intelligence Committee dedicated a nearly 1000-page report to the 2016 election alone (*Report on Russian Active Measures*). Misinformation and the use of deceptive information for warfare practices is not something new that was brought on by the internet but rather was more enabled thanks to such interconnectivity. Once the Russians saw their methods' effectiveness, interference was utilized in other elections such as the 2019 Ukrainian election that brought Volodymyr Zelensky to power (Pidkuĭmukha and Kiss). The Soviets made use of "active measures" during the Cold War (Cull), sparking riots across an adversary's country aimed at destabilizing the nation, especially during the Vietnam War to demoralize the American public and during the civil rights movement to worsen race relations. Some more modern ones include the Chinese government pointing to Fort Dietrich, the Army's biological weapon study center as the origin of COVID-19, similar to how the Russians pointed to it in the 1980s as the origin of AIDS (Poster).

Misinformation campaigns appeared long before the internet but were enhanced by its invention. With online algorithms, such as YouTube's recommendation algorithm, putting forward thumbnails with shocked expressions or red arrows, adversarial nations have used our very means of entertainment and digital news feeds to feed just enough falsehoods to cause major events in modern history. Some accredit the 2016 election as being the first major misinformation campaign victory in the US while the origins for such events go further back in time, especially closer to the Russo-sphere.

1. Russian Cyber Operations in Europe

Estonia is the leading nation regarding technology and cybersecurity, with citizens being able to vote, pay their taxes, and host their ID online. In 2007, Russian agents sparked riots regarding the removal of the bronze statue in the middle of Tallinn, where Russians would come to honor the heroes of the Great Patriotic War (WW2). A symbol of the Soviet past in a country recently independent from the USSR, the Estonian government started excavations to relocate the statue from the city center when rumors were heard of Russian "diplomats" meeting with possible rioters. What ensued were three nights of rioting;

the worst event an independent Estonia had seen in its history followed by cyberattacks attributed to Russia with economic sanctions from the same actor. Today, Estonia hosts the NATO Center for Cyber Excellence and is a leading digital nation in Europe, boasting 89% of the population being internet users, but it still suffers from the long-lasting effects of the Bronze Night. Russian hybrid operations continue to broadcast propaganda at Estonia's 25% of ethnic Russians, in the hope that they become disenfranchised with their nation and create widespread riots (Jankowicz).

Figure 1 An example of a whataboutism meme, here just a Twitter screenshot but many other formats exist.

As the world woke up on February 24th, 2022, the internet discovered that Russia had acted on its threats to invade Ukraine and was now at the gates of Kyiv. As war on the ground and in the air was raging, for the first time in history, cyber war became just as important with parties exchanging attacks on major infrastructure and systems vital to the war effort. Non-technical attacks were a major part of the first days of the war as the Kremlin actively targeted Ukrainian citizens with various misinformation narratives to convince them not only to give up the fight against the Russian invaders but to start supporting them through any means possible, even including false videos aiming to expose "Ukrainian war crimes" (Sardarizadeh). Such videos are falsely labeled to confuse OSINT investigations and maliciously attribute inhumane actions to the wrong party (Fleischman).

As a direct response to Russian influence efforts, Ukrainian citizens started NAFO, the North Atlantic Fellas Organization to combat Russian misinformation through memes and directly target whataboutism, deflecting attention from a major issue by pointing out another one.

Recently, with Russian losses in Ukraine increasing, whataboutism has dramatically increased on social media with a particular focus on how much funding the Ukrainian government is getting from Western governments. This is done both directly by pointing out amounts spent as well as directly pointing out past actions of Western and Ukrainian governments and equating them to each other. Such actions involve the Waco Branch Davidian incident, MK Ultra, Ruby Ridge, and many more. Experts often accredit the misinformation branch of the Russian Intelligence industry to the Wagner Group (BBC), Russia's hybrid warfare enterprise.

2. Modern campaigns

Modern misinformation campaigns are especially aimed at the public and rely on an initial batch of fake accounts on social media to create and publish the content before being distributed by more accounts, some of them being government officials.

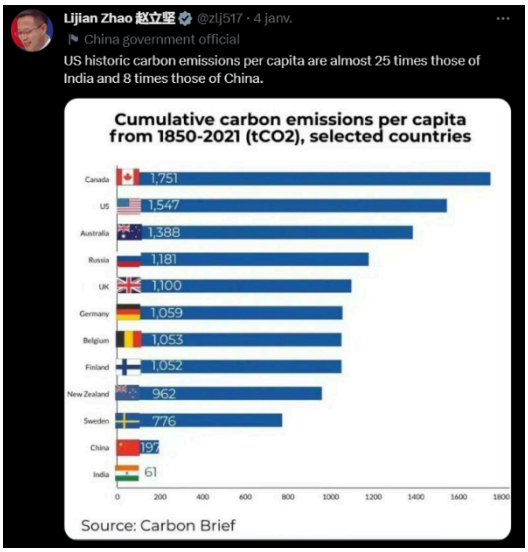


Figure 2: Twitter post from Lijian Zhao, Deputy Director-General, Department of Boundary and Ocean Affairs for China (Lijian)

While most misinformation campaigns are meant solely for destruction and chaos, some are purely meant to bolster nationalism and patriotism, even if it means sacrificing a bit of truth. The example illustrated in Figure 2 shows the cumulative carbon emissions, with Canada and the US leading. This comes at the same time Joe Biden was giving a speech on the environment and the fall Unified agenda (Brugger). Looking at multiple sources, including the very article cited (Evans), it fails to mention that these carbon emissions are "Cumulative emissions per population" (European Commission. Joint Research Centre.), and with China having such a large population, the numbers are in their favor. Since Twitter's purchase by Elon Musk during the summer of 2022 and Facebook's name change to Meta, more and more people have started abandoning more traditional social media platforms in favor of newer and more specialized ones like Reddit, IFunny, and TikTok. Some of the most recent misinformation campaigns originated at the height of the coronavirus pandemic, intending to spread false facts about remedies and other ways to fight the disease, at a time when TikTok gained popularity (Basch et al.).

An often undiscussed and unresearched platform, IFunny is home to memes, images, and short videos taken from the internet with a humoristic backdrop intended to entertain the viewer and sometimes pass along a deeper, even political message. Recent events like the Chinese spy balloon and the Russian invasion of Ukraine have turned memes, a simple form of entertainment, into a means of misinformation with the objective that their facility of comprehension will entice users to share them on other platforms, thus starting the whole campaign. iFunny is a Cyprus-based meme platform that has very little censorship. Examples of censorship attempts include replacing the word "Hitler" with "loser," making his praise impossible, or allowing users to remove some of the comments and reaction images. As censorship evolves, so do the methods to circumvent censorship. The community has adapted by lightly

crossing out certain terms that might get censored on other platforms like “rape” or “genocide,” or cropping out the last sentence that might include an offensive term so that only the top parts of the letter emerge, allowing the audience to understand while tricking optical character recognition software. 4Chan sees communities gather from many different countries for different overarching themes including “Papercraft and origami,” “History and humanities,” and “politically incorrect.” Threads are posted on 10 pages worth of content and are ranked by the number of interactions, meaning unpopular threads simply disappear after a while, also allowing the site to run on minimal storage requirements. Reddit stores each thread as its separate page and uses the r/ prefix to classify each one, with the most popular pages or “subreddits” making it onto the website’s landing page. Twitter, now X, is widely known among misinformation researchers for its easily digestible 180-character format and more recently in the news for being bought by Elon Musk. Telegram has its users subscribe to channels and each owner is responsible for monitoring and encrypting every message sent, bringing it into the spotlight recently for being used as a means to spread propaganda or get news from the frontline of the Russian-Ukrainian conflict.

With 2.8 million subreddits, and 430 million Monthly Active Users (MAUs), Reddit might seem like the largest National Security concern, however, that spot would go to 4Chan. Although it has only 22M unique users, with more than half from the USA, 4Chan users are more likely to translate their actions into the real world. As 4chan brings together people from different backgrounds with a pervasive knowledge of the internet, they can do some incredible things for the sole purpose of trolling individuals. Some noteworthy contributions of 4Chan’s “weaponized autism” include Shia LeBeouf’s flag live stream hijacking and contributing to the 2016 election events. By utilizing the stars in the sky and airline trajectories, 4chan users were able to find the location where actor Shia LeBeouf was live streaming a flag saying, “He will not divide us” and replace it with one depicting Pepe the frog, 4chan’s unofficial mascot. 4chan also contributed to making memes about Donald Trump, depicting him as a divine savior and remixing his physical attributes, mainly the blonde hair and MAGA hat.

With enough motivation and a cause interesting enough that the mainstream media would pick up on it, internet communities will jump into action in real life, something our adversaries have exploited, and continue to exploit to this day. The Internet Research Agency (IRA), Russia’s online hybrid warfare division, can cause riots from across the world. The IRA has already utilized online divisions and meme-making as a weapon, with different floors targeting opposing groups of people (Sanger).

As the war in Ukraine continues, so does the coverage of it on the internet, both through mainstream channels and backchannels. iFunny, a Cyprus-based platform dedicated to memes, has seen an increase in memes supporting the war in Ukraine as a Russian intervention. The narrative pushed is the suppression of bioweapons labs funded by America and Fauci in particular, allying both covid-deniers/skeptics and pro-Russians around a common cause.

3. Technical Research

Figure 3. An attempt at justifying the invasion of Ukraine through past conspiracies

Most of the research currently done in computer science seems to use a similar dataset to that of the 2018 US Midterm Elections and 2016 US Presidential Elections. While excellent datasets, using a static dataset for this application will result in a model that might expire rather quickly since formats, features, and styles quickly evolve and become obsolete. For this reason, models must be constantly evolving and use up-to-date memes and data to be relevant. With more important events happening every day like the Chinese spy balloon or upcoming events such as the 2024 election, utilizing data from past events will likely result in rapidly expiring research. The dataset origins used also do not reflect where the most problematic memes are coming from, with some of the more reaction-causing ones coming from unknown platforms like iFunny or uncontrolled ones like 4Chan. As a lot of previous work has been done looking at both misinformation campaigns and memes separately, it is important to note that very little research has tried to combine both aspects of studying social media. One key factor missing from the current computer science literature is the attribution of misinformation to a specific actor. Since most of the research focuses on classifying whether a certain meme is offensive or is misinformation, the question needs to be expanded to look at the issue of attribution and "propaganda memes, especially during an election" (Afridi et al.).

Some of the challenges that will arise include the fact that memes, by their nature, are meant to be humorous and this can translate into an added difficulty for Natural Language Processing (Smitha et al.). There therefore needs to be more attention to the human aspect behind the problem while "common misinformation strategies […] tend to focus on the technical […] aspect" (Fernandez and Alani) while utilizing technical solutions.

4. Memes as a vector for misinformation

Moving away from simple images and videos, certain images or situations have become ingrained in pop culture, especially thanks to the Vine social media platform. Vine first introduced the world to short videos, up to 7 seconds long, to entertain a population of teens and young adults who grew up with short attention spans due to being raised with screens. When Vine was taken down by its parent company Twitter in 2017, the internet longed for a replacement. With TikTok and Musical.ly combined, it found refuge in Tiktok, "the app's spiritual successor" (Stokel-Walker). With much of Vine's source code being written over a decade ago, a possible revival with Elon Musk would face uphill challenges in compatibility and marketing, having to reconquer users from TikTok's current base.

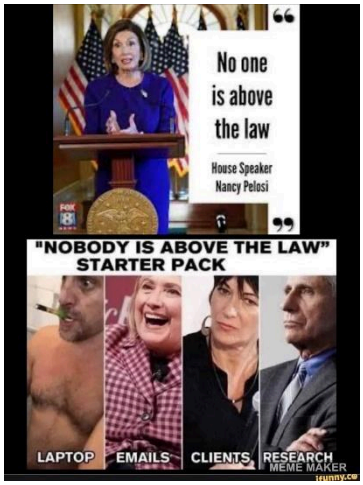


Figure 4. Whataboutism in action, note the key terms used on the bottom line

The fact TikTok is foreign-owned did not go unnoticed by the US government, with Donald Trump trying to ban it in 2020 and Texas banning it from state-owned devices, including university-owned devices. On top of being a threat to national security, because many in the military have it installed on their devices, TikTok also is completely different in the US and China, where it originated. Users on "Douyin", the CCP-approved version of TikTok, are presented with "science, educational and historical content," while American teens are shown "stupid dance videos with the main goal of making us imbeciles" (Schlott). TikTok mainly saw its growth during the COVID-19 lockdowns throughout the world as the population needed a way to connect but was also a vector for medical misinformation (Basch et al.).

Tensions in the South China Sea and lessons learned from the conflict in Ukraine lead the misinformation analysis community to believe similar tactics of hybrid warfare will be used preemptively. With China having access to TikTok as a weapon of mass influence, campaigns that support the Chinese forceful reunification with Taiwan or spam whataboutism memes on the internet are very likely to appear shortly before the invasion.

Figure 5. A tweet about Finland's accession into NATO. Note the jump from Finland joining the alliance to the US being involved in sabotage

As the invasion draws near, whataboutism campaigns will appear on the American internet to remind citizens of past American

government mistakes or refuel their current motivations by echoing the current problems faced by our nation.

A recent example involves Finland's recent accession to NATO and the reactions that have emerged on the internet. iFunny reacted by posting memes (sometimes including screenshots from Twitter) while 4Chan hosted one-sided discussions (see Appendix for full-sized screenshot). In both cases, past controversies were brought up with the wording recalling controversial events like the Nord Stream pipeline explosion or the pullout from Afghanistan.

With Russia focusing its misinformation campaign efforts on using text-based and image-based approaches to pass along its messages, China uses video-based approaches, banking on younger users' weaker attention spans. A good illustration is Nancy Pelosi's trip to Taiwan in the summer of 2022 which sparked Chinese citizens to post videos of themselves crying on Douyin after the government threatened to shoot down her plane but did not act. Seeing this, the Chinese internet would react similarly if they were promised a reunification of Taiwan and were not given it (Yu et al.). An interesting parallel to make with the events in Ukraine is that posts that openly showed opposition to the special military operation were automatically censored by Russia's decentralized censorship network, running through the ISPs. In the case of a possible invasion of Taiwan, China would utilize its extensive firewall censorship network to block out any forms of criticism through deep packet analysis, policies already in place on Chinese social media sites. A large number of people are employed by the state to do just that as well as monitor any forms of outside communication.

Having a large number of people to rely on who will always be loyal to their nation, China has created an impressive online militia with the sole objective of discrediting online opposition and flooding the internet with pro-China messages. This can be seen as an internet equivalent of the police stations the Chinese government runs overseas tasked with harassing citizens living overseas until they return to the mainland, usually to face trial for political opposition (dos Santos). The people hired as part of this operation fabricate up to 448 Million comments in a year (King et al.). Instead of arguing with people on the internet, the strategy involves cheering on China and distracting the public by changing the subject, contributing to Russia's strategy of whataboutism. Such accounts will cultivate an online persona, often with several of them being attributed to the same operator and are referred to as "wumao" or part of the "50c army" a term popularized by YouTuber China Uncensored referring to their alleged salary per hour.

5. Possible solutions

In the face of all of these political problems facing social media, especially foreign governments' abuse of America's freedom of speech to propagate their messages, several solutions are possible: full censorship, reverse engineering, or creating an American 50c army.

Figure 6. An example of a message generated in the style of the original album cover

Full censorship of content is impossible, mainly due to the First Amendment, but also due to the nature of the internet itself and the fact everything can easily be saved and distributed elsewhere. A main problem to look at is the internet's constant ability to innovate and change the way the content is posted as well as new ways to post offensive content. Known and proven methods outlined earlier involve certifying text (replacing letters with numbers, which is already treated by specially trained models), adding extra letters, or changing the spelling so that the pronunciation is different enough to still be understood as the original message.

Images are much harder to censor than text, making memes a high-value target. Their nature as remixed or modified images makes it nearly impossible to recognize, both from a technical and philosophical standpoint. As certain songs, extracts from shows, images, and many more pieces of content have made their way into memes by various online communities, creating a technical censorship solution might lead to the end of pop culture. Technical censorship of memes would be limited to only feature recognition and basic Optical Character Recognition (OCR). Between Imgflip and many other sites, memes are easier than ever to generate, and modern OCR techniques cannot catch every occurrence of a word. The iFunny community has recently discovered that the Metallica logo generator can generate a custom 72 Seasons album poster and the text in that poster will not get caught by current technical means. This has led many to post very offensive comments, some of them racially motivated, that will not be detected due to the distortion of the letters. While the First Amendment does not apply to social media companies as they are only "providers" of content and can freely "censor what people post on their websites as they see fit" (Nott and Peters), censorship could be enacted by the government by asking for certain stories, like the Hunter Biden laptop controversy, or posts to be suppressed as the Twitter files have suggested (Bond).

Reverse engineering misinformation campaigns, from a technical point of view, has already worked and led to several common methods of detecting misleading articles and posts on the internet. An analysis of the currently available computer science literature regarding the analysis and classification of content as misinformation shows the major following contributions:

Linguistic-based analyses (Moura et al.; Schuster et al.; Bright et al.; de Oliveira et al.) have been used on articles to look at the way they were written to determine if they were generated with the intent to create an emotional response. Since fake news is meant to trigger an emotional response, researchers have focused mainly on the idea of novelty in misinformation. This concept works by introducing the reader to something so revolutionary that it entices people to click on it (Kumari et al.). Multi-level voting using a Term Frequency Inverse Document Frequency (TFIDF) vector for this purpose was done to look at India's 2019 election misinformation and try to automate its detection (Kaur et al.).

Content blind analyses (Szanto; Bo et al.; Jasser; Luo et al.; Murayama et al.) focus on the misinformation after it has been generated, once it is time to start sharing it and interacting with it. Bo et al focus mainly on predicting how likely a user is going to click on a misinformation article and was tested on a few known topics that have generated such content. Szanto et al focus mainly on the sharing aspect of misinformation content and whether it corresponded to previous patterns of misinformation sharing based on the network representation using graph-based machine learning. Jasser et al focus on sharing campaigns from a non-technical point of view and look mainly at the different stages of misinformation campaigns. By looking at how misinformation campaigns emerge and the dynamics behind each "information burst," the public can better be protected from them.

As Large Language Models (LLM) like ChatGPT have emerged and been opened up to the public all over the internet, many adversaries have already started using this technology to create code and sometimes malware (Mijwil and Aljanabi) but cannot seem to do it perfectly (Starks and Schaffer). Machine-generated misinformation has also been a serious question that the computer science community has started to worry about. This has implications for other human interactions over the internet like product reviews (Jawahar et al.). According to the Ohio State University (Bhat and Parthasarathy), LLMs have made "the detection of fake news […] challenging". The Allen Institute for AI's research on the topic includes the GROVER model, able to both detect fake news articles and generate them from a simple headline. Its detection though is limited and only works best with the examples provided. Attempts at detecting machine-generated articles often result in false positives and classify the

articles as “written by a machine” when the real author was human, just not a native English speaker (Zellers et al.). Researchers from the French Superior Institute of Electronics (Wang et al.) expand on the concept of machine-generated misinformation by adding the ability to have bots share the content and detect such accounts.

With "67% of adults" getting their daily updates from social media, news companies have adapted their content to fit such platforms. This has directly contributed to social media's "growth and popularity" over the years and led to the innovation of formats easier to digest like TikTok's 30-second videos (Khan et al.).

Adversarial parties rely on America choosing not to engage in disinformation operations, which introduces the question of creating our active measure operations to counteract what is being done to us. The resulting operation would not be spreading false information but could closely resemble what is already done by Russian, Chinese, and Iranian internet trolls, focusing the attention on other events, both past and present as well as flooding chatrooms with huge amounts of political satires and memes depicting the current leaders. What America lacks in this scenario would be people. With a population 3 times the size of America and a school system that engrains patriotism in its citizens from a young age with every private enterprise linked to the government, the Chinese have been very active on the internet to spread their influence. As the American internet is opened to the rest of the world to enable commerce and trade, many have sought to misuse Western social media’s openness. Foreign social media is heavily surveilled, scrutinized, and censored, with some nations even forcing their citizens to tie identification to their online persona, thus eliminating anonymity. An American influence operation would therefore have to recruit foreign nationals, spoof identities or steal credentials to gain access to a foreign network.

With the trend of online content pointing towards memes and easily digestible formats like short videos (TikTok, Youtube Shorts, Instagram Reels), foreign adversaries will move their misinformation efforts to accommodate these formats. The 2016 election introduced memes as a form of misinformation on the internet, and the 2020 election showed that it would be the future of hybrid warfare. With future electoral cycles, world events, and pop culture events influencing the atmosphere and current trends, foreign adversaries will learn to adapt their messages to the online communities that will actively go out and spread their message, similar to how 4Chan users were active in real life as well as online. Censorship is not a valid solution, both for legal purposes and due to the technical challenges faced. The

American IC will need to find new and innovative ways to both predict the next misinformation campaigns but also attribute them to the actors responsible. Current computer science research can be used as a starting point but will need to be expanded as it only focuses on the binary question of whether or not something is misinformation, not which campaign it belongs to. The greatest challenge with attribution will of course always be online anonymity and the speed and range at which popular posts of all natures will spread, especially if the message broadcasted is one that will generate a lot of animosity.

6. The 2024 election season

As we begin the seasons leading up to the 2024 presidential election, it is important to realize that we are not living in the same environment as we were back in 2016 or even 2020. Covid-19 has changed the playing field, bringing along with it a rise in mail-in voting, a return to pandemic-time politics, and a wide variety of conspiracies accompanying it. Technology has also evolved significantly, to a point where the line between real and fake images is getting blurrier by the day, especially with large image models such as DALL-E and Stable Diffusion.

With the democratization of large image models, their applications and the prompts people have suggested have illustrated the versatility of such innovations. Users have generated anything from Pixar-style movie posters to oil paintings depicting abstract topics, while some have used them to propagate hurtful stereotypes. As we inch closer to another election, both in person at the polls and digitally on social media, it will be interesting to see what images will make their way onto the internet to disrupt electoral security. The first time such a situation arose was in March 2023, at a time when Donald Trump was facing charges related to business fraud, and the rumor that he had been arrested started to sweep the internet with images of him getting chased or handled by police officers (Arijeta Lajka) or even him in prison lifting weights. The images were of course proven to be fake rather quickly. An attentive eye can quickly discern minute details like blurriness of faces, missing fingers in fake pictures, or impossible postures, but adversaries eager to sow chaos into our democracy and agitate the political sphere do not differentiate. What started as a simple experiment by “Eliot Higgins, the founder of Bellingcat, a Netherlands-based investigative journalism collective” quickly turned into people sharing images with no context and certain forums getting ahold of them to distort their true context. Research has luckily started to catch up with the popularization of such images, looking at extremely detailed features such as the “lack of explicit 3D modeling of objects and surfaces [that] causes asymmetries in

shadows and reflected images" (Corvi et al.). While many argue that AI models must be regulated, it is important to note that the images created were generated from a prompt written by a user, and are therefore not to be blamed for the rise of false content.

With a growing concern that the 2024 election will see a sharp rise in disinformation content on all platforms, especially the more popular ones like Twitter/X or TikTok, which allow for rapid digestion of content, we can start to look at possible policies to put in place to help mitigate or eliminate such risks. While addressing conspiracies and falsehoods right as they emerge might be seen as the obvious answer, online communities will be actively monitoring responses to ongoing events. Reacting to them too soon might trigger an adverse response, with users claiming a cover-up. This would especially be true if the reaction immediately came from a presidential candidate. The mainstream media would immediately report on it, raising the issue of giving online disinformation posters a wider and more mainstream platform or legitimizing their content. One of the pitfalls to avoid would be to directly reply online or have teams of people dedicated to replying to the content posted, to prevent the idea of a "thought police" from being popularized.

Since the candidates themselves are hesitant to be involved with the issue due to the fear of the online conversation veering off towards more conspiracies, the federal government might be tempted to respond. As already outlined in the second recommendation of the Cyberspace Solarium Commission white paper on countering disinformation in the US, funding can be attributed to "non-governmental disinformation researchers," which could lead to further breakthroughs, especially regarding emerging technologies like large-image models. As the intelligence community most likely already has a presence on non-mainstream social media, a public report could be published periodically outlining the "current state of conspiracies on the internet," similar to that of a State of the Union but for online disinformation. As previously outlined, you would be giving disinformation posters not only a platform, but an objective of making it onto the list, and online communities would see it as a game, trying to be the highest on the list, like a scoreboard. Adding a legal layer to this issue would only create a bigger challenge for prosecution. Unless an individual uses a social media site based in the US, allowing law enforcement to execute warrants for the information shared and personal information, there is no good way of tracking individuals who post disinformation on less mainstream sites. A good case study to look at is Operation FIREWALL, involving the arrests of 28 individuals across eight states and six foreign countries (United States Secret Service). This operation was a success since the Secret Service asset, who became one of the administrators of an online marketplace selling stolen information, encouraged users

to tunnel their traffic through a server controlled by the Secret Service, which completely de-anonymized their information Since then, VPNs (Virtual Private Networks) like the ones used for operation FIREWALL, encryption standards, and an overall general understanding of technology would make such an operation impossible in today's cyberspace.

Since there are no concrete policy solutions the federal government could implement that would have immediate effects, one might be tempted to turn towards social media companies as they would have better control over the content posted on their platform. As previously mentioned, one of the newer trends that have emerged recently has been to make use of a large image model to generate content that only a subpopulation of users would be able to pick up on. The driving force behind the use of such software is the inability of social media websites to properly detect double-meaning images, with either a hidden image or hidden text, giving way to a completely new meaning. Hiring in-house researchers to tackle the issue or funding universities to run the projects could be the first step to addressing this issue. As reading some of these dual-meaning images sometimes requires a user to squint their eyes, a mild amount of blur could be applied to a picture to simulate the physical motion. To detect if text is present, multiple techniques of Optical Character Recognition (OCR) could be applied to try to find the hidden message. Other research projects could focus on looking at the most popular n-grams of words to try to figure out if a certain narrative is being repeated all over the website, whether that may be information or misinformation, and identify known conspiracies. The issue then arises of a true story, thus a repeated set of words, that a social media site would want to censor or cover-up. One could argue this would give a site more accurate data as to which posts to censor or how popular the story has already received. As a case study, the Hunter Biden Laptop comes to mind.

Countering disinformation is an impossible game to win for governments, social media sites, political candidates, and public figures alike. With users constantly innovating new ways to sneak censorable messages into content, technological advances enabling it, and world events happening faster than journalism can report them accurately, the disinformation war is one we will never win. Not winning does not mean losing though, we can hold out for as long as possible, but every card put down would be unique and impossible to pick back up; each technological rollout becoming public after a time (likely due to leaks and reverse engineering) and every policy gaining unwanted attention from the communities posting online. As the American public is very attached to its First Amendment rights, enacting rules or laws that directly interfere with an individual's ability to post their ideas will always cause controversy and encourage online communities to circumvent them.

References:

18 U.S. Code § 35 - Imparting or Conveying False Information. Cornell Law School, https://www.law.cornell.edu/uscode/text/18/35.

Afridi, Tariq Habib, et al. "A Multimodal Memes Classification: A Survey and Open Research Issues." *Innovations in Smart Cities Applications Volume 4*, edited by Mohamed Ben Ahmed et al., vol. 183, Springer International Publishing, 2021, pp. 1451–66. *DOI.org (Crossref)*, https://doi.org/10.1007/978-3-030-66840-2_109.

Arijeta Lajka. "Trump Arrested? Putin Jailed? Fake AI Images Flood the Internet, Increasing 'Cynicism Level.'" *Syndey Morning Herald*, 24 Mar. 2023, https://www.smh.com.au/world/north-america/trump-arrested-putin-jailed-fake-ai-images-flood-the-internet-increasing-cynicism-level-20230324-p5cuup.html.

Basch, Corey H., et al. "A Global Pandemic in the Time of Viral Memes: COVID-19 Vaccine Misinformation and Disinformation on TikTok." *Human Vaccines & Immunotherapeutics*, vol. 17, no. 8, Aug. 2021, pp. 2373–77. *DOI.org (Crossref)*, https://doi.org/10.1080/21645515.2021.1894896.

BBC. *What Is Russia's Wagner Group of Mercenaries in Ukraine?* 23 Jan. 2023, https://www.bbc.com/news/world-60947877.

Bhat, Meghana Moorthy, and Srinivasan Parthasarathy. "How Effectively Can Machines Defend Against Machine-Generated Fake News? An Empirical Study." *Proceedings of the First Workshop on Insights from Negative Results in NLP*, Association for Computational Linguistics, 2020, pp. 48–53. *DOI.org (Crossref)*, https://doi.org/10.18653/v1/2020.insights-1.7.

Bo, Hongbo, et al. *Ego-Graph Replay Based Continual Learning for Misinformation Engagement Prediction*. no. arXiv:2207.12105, arXiv:2207.12105, arXiv, 25 July 2022. *arXiv.org*, http://arxiv.org/abs/2207.12105.

Bond, Shannon. “Elon Musk Is Using the Twitter Files to Discredit Foes and Push Conspiracy Theories.” *NPR*, 14 Dec. 2022, https://www.npr.org/2022/12/14/1142666067/elon-musk-is-using-the-twitter-files-to-discredit-foes-and-push-conspiracy-theor.

Bright, Laura F., et al. “A Deeper Look at the 2020 Facebook Boycott and Related Themes of Misinformation: A Text Mining Analysis of Topics, Emotion, and Sentiment.” *Journal of Brand Strategy*, vol. 11, no. 1, 2022, pp. 65–79.

Brugger, Kelsey. “White House Releases Latest Regulatory Plans.” *E&E News*, 4 Jan. 2023, https://www.eenews.net/articles/white-house-releases-latest-regulatory-plans.

Corvi, Riccardo, et al. *On the Detection of Synthetic Images Generated by Diffusion Models*. arXiv:2211.00680, arXiv, 1 Nov. 2022. *arXiv.org*, http://arxiv.org/abs/2211.00680.

Cull, Nicholas. *America’s Countering Soviet Disinformation in the 1980s*. European Network Remembrance and Solidarity, 15 Feb. 2021, https://hi-storylessons.eu/wp-content/uploads/2021/02/15_N.Cull_Americas-Countering-Soviet-Disinformation-in-the-1980s_EN.pdf.

de Oliveira, Nicollas R., et al. “Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges.” *Information*, vol. 12, no. 1, Jan. 2021, p. 38. *DOI.org (Crossref)*, https://doi.org/10.3390/info12010038.

dos Santos, Nina. “Exclusive: China Operating over 100 Police Stations across the World with the Help of Some Host Nations, Report Claims.” *CNN*, 4 Dec. 2022, https://www.cnn.com/2022/12/04/world/china-overseas-police-stations-intl-cmd/index.html.

European Commission. Joint Research Centre. *CO2 Emissions of All World Countries :JRC/IEA/PBL 2022 Report.* Publications Office, 2022. *DOI.org (CSL JSON)*, https://data.europa.eu/doi/10.2760/730164.

Evans, Simon. “Analysis: Which Countries Are Historically Responsible for Climate Change?” *Carbon Brief*, 5 Oct. 2021, https://www.carbonbrief.org/analysis-which-countries-are-historically-responsible-for-climate-change/.

Fernandez, Miriam, and Harith Alani. “Online Misinformation: Challenges and Future Directions.” *Companion of The Web Conference 2018 on The Web Conference 2018 - WWW ’18*, ACM Press, 2018, pp. 595–602. *DOI.org (Crossref)*, https://doi.org/10.1145/3184558.3188730.

Fleischman, Amir. “Effective Use of OSINT in theRusso-Ukrainian War:Collection and Disclosure of Reliable Information Along with Refuting False Information.” *American Intelligence Journal*, vol. 32, no. 2, 2022, pp. 66–79.

Jankowicz, Nina. *How to Lose the Information War: Russia, Fake News, and the Future of Conflict*. I.B. Tauris, 2020.

Jasser, Jasser. “Dynamics of Misinformation Cascades.” *Companion Proceedings of The 2019 World Wide Web Conference*, ACM, 2019, pp. 33–36. *DOI.org (Crossref)*, https://doi.org/10.1145/3308560.3314194.

Jawahar, Ganesh, et al. *Automatic Detection of Machine Generated Text: A Critical Survey*. no. arXiv:2011.01314, arXiv:2011.01314, arXiv, 2 Nov. 2020. *arXiv.org*, http://arxiv.org/abs/2011.01314.

Kaur, Sawinder, et al. “Automating Fake News Detection System Using Multi-Level Voting Model.” *Soft Computing*, vol. 24, no. 12, 12, June 2020, pp. 9049–69. *DOI.org (Crossref)*, https://doi.org/10.1007/s00500-019-04436-y.

Khan, Sayeed Ahsan, et al. “The Use and Abuse of Social Media for Spreading Fake News.” *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, IEEE, 2019, pp. 145–48. *DOI.org (Crossref)*, https://doi.org/10.1109/I2CACIS.2019.8825029.

King, Gary, et al. “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument.” *American Political Science Review*, vol. 111, no. 3, 2017, pp. 484–501.

Kumari, Rina, et al. “Misinformation Detection Using Multitask Learning with Mutual Learning for Novelty Detection and Emotion Recognition.” *Information Processing & Management*, vol. 58, no. 5, 5, Sept. 2021, p. 102631. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ipm.2021.102631.

Lijian, Zhao. “Lijian Zhao’s Twitter Profile.” *Twitter*, https://twitter.com/zlj517?

Luo, Han, et al. “Spread of Misinformation in Social Networks: Analysis Based on Weibo Tweets.” *Security and Communication Networks*, edited by Chenquan Gan, vol. 2021, Dec. 2021, pp. 1–23. *DOI.org (Crossref)*, https://doi.org/10.1155/2021/7999760.

Mijwil, M., and Mohammad Aljanabi. “Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime.” *Iraqi Journal for Computer Science and Mathematics*, Jan. 2023, pp. 65–70. *DOI.org (Crossref)*, https://doi.org/10.52866/ijcsm.2023.01.01.0019.

Moura, Ricardo, et al. “Automated Fake News Detection Using Computational Forensic Linguistics.” *Progress in Artificial Intelligence*, edited by Goreti Marreiros et al., vol. 12981, Springer International Publishing, 2021, pp. 788–800. *DOI.org (Crossref)*, https://doi.org/10.1007/978-3-030-86230-5_62.

Murayama, Taichi, et al. “Modeling the Spread of Fake News on Twitter.” *PLOS ONE*, edited by Kazutoshi Sasahara, vol. 16, no. 4, Apr. 2021, p. e0250419. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0250419.

Nott, Lata, and Brian Peters. “Free Speech on Social Media: The Complete Guide.” *Freedom Forum*, https://www.freedomforum.org/free-speech-on-social-media/.

Pidkuĭmukha, Liudmyla, and Nadiya Kiss. “Battle of Narratives: Political Memes During the 2019 Ukrainian Presidential Election.” *Cognitive Studies | Études Cognitives*, no. 20, Dec. 2020. *DOI.org (Crossref)*, https://doi.org/10.11649/cs.2246.

Poster, Alexander. “The Russian ‘fake News’ Campaign That Damaged the United States - in the 1980s.” *Washington Post*, 12 Mar. 2018, https://www.washingtonpost.com/news/made-by-history/wp/2018/03/12/the-russian-fake-news-campaign-that-damaged-the-united-states-in-the-1980s.

Report on Russian Active Measures. House Permanent Select Committee on Intelligence, 22 Mar. 2018, https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures.

Sanger, David E. *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age*. First paperback edition, Broadway Books, 2019.

Sardarizadeh, Shayan. “Ukraine War: False TikTok Videos Draw Millions of Views.” *BBC*, 25 Apr. 2022, https://www.bbc.com/news/60867414.

Schlott, Rikki. “China Is Hurting Our Kids with TikTok but Protecting Its Own Youth with Douyin.” *New York Post*, 25 Feb. 2023, https://nypost.com/2023/02/25/china-is-hurting-us-kids-with-tiktok-but-protecting-its-own/.

Schuster, Tal, et al. “The Limitations of Stylometry for Detecting Machine-Generated Fake News.” *Computational Linguistics*, vol. 46, no. 2, 2, June 2020, pp. 499–510. *DOI.org (Crossref)*, https://doi.org/10.1162/coli_a_00380.

Smitha, E. S., et al. “Meme Classification Using Textual and Visual Features.” *Computational Vision and Bio Inspired Computing*, edited by D. Jude Hemanth and S. Smys, vol. 28, Springer International Publishing, 2018, pp. 1015–31. *DOI.org (Crossref)*, https://doi.org/10.1007/978-3-319-71767-8_87.

Solen, Derek. “Fight Fire with Fire: The PLA Studies Hybrid Warfare.” *Air University, China Aerospace Studies Institute*, Mar. 2022, https://www.airuniversity.af.edu/CASI/Display/Article/2975035/fight-fire-with-fire-the-pla-studies-hybrid-warfare/.

Starks, Tim, and Aaron Schaffer. “Yes, ChatGPT Can Write Malicious Code - but Not Well.” *Washington Post*, 26 Jan. 2023, https://www.washingtonpost.com/politics/2023/01/26/yes-chatgpt-can-write-malware-code-not-well/.

Stokel-Walker, Chris. “Elon Musk’s Plans to Revive Vine Face One Big Problem: The Reason It Closed Originally.” *MIT Technology Review*, Oct. 2022, https://www.technologyreview.com/2022/10/31/1062465/elon-musks-plans-to-revive-vine-face-one-big-problem-the-reason-it-closed-originally/.

Szanto, Aron. *Defuse the News: Predicting Misinformation and Bias in News on Social Networks via Content-Blind Learning*. 2018. Harvard College, https://dash.harvard.edu/handle/1/38811538.

United States Secret Service. *Operation FIREWALL*. https://www.secretservice.gov/operationfirewall.

Wang, Patrick, et al. “Is This the Era of Misinformation yet: Combining Social Bots and Fake News to Deceive the Masses.” *Companion of The Web Conference 2018 on The Web Conference 2018 - WWW ’18*, ACM Press, 2018, pp. 1557–61. *DOI.org (Crossref)*, https://doi.org/10.1145/3184558.3191610.

Yu, Sun, et al. “‘Take down Pelosi’s Plane’: Chinese Social Media Users React to Taiwan Visit.” *Financial Times*, 3 Aug. 2022, https://www.ft.com/content/78c12485-3dcc-45f8-b779-7dd581cd66b9.

Zellers, Rowan, et al. *Defending Against Neural Fake News*. arXiv:1905.12616, arXiv, 11 Dec. 2020. *arXiv.org*, http://arxiv.org/abs/1905.12616.

Appendix:

File: skynews-ukraine-finland-m(...).png (143 KB, 768x432)

What can NATO even do when Russia invades Finland? Anonymous (ID: nVa7IJLb) 04/16/23(Sun)12:49:40 No.423763660 [Reply] ▶ >>423778520

All of NATO's forces and equipment are tied up in ukraine

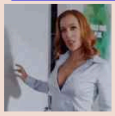
52 replies and 8 images omitted. Click here to view.

Anonymous (ID: 5XiD3uiC) 04/16/23(Sun)15:15:18 No.423777319 ▶

If Russia attacks Finland, we are no longer fighting a proxy war and can bring the fight to Russia directly. Also, most of our equipment is not in Ukraine. If you haven't noticed, we haven't sent them any fighter jets or naval ships. Did you think our bread and butter was artillery? Bombers are our artillery you dumbass. We don't fight ground wars.

Anonymous (ID: MVOegX5H) 04/16/23(Sun)15:27:45 No.423778520 ▶ >>423778761

File: 1680500260061822.gif (475 KB, 220x220)



>>423763660 (OP)

>What can NATO even do

Nato is a suicide pact
What appears to be a coalition of dozens of nations
is in reality a dysfunctional web of quasi-territories
which all have a non-functional economic structure
propping up the entire rotten house of cards.

The US isn't your "friend". Not even an ally.
NATO isn't there to guarantee your security, it's there to secure the US' dominance over Europe.
The US/NATO is unironically the worst enemy of Europe today.
Nato is LITERALLY an extra-national expeditionary force the US masquerades as on European soil for 'taste' and 'optics' purposes, nothing more.

Europe literally cant do anything.

Comment too long. Click here to view the full text.

Anonymous (ID: RE5mCTlo) 04/16/23(Sun)15:30:08 No.423778761 ▶

File: 1674672308167905.png (693 KB, 1196x1508)

>>423778520

This shit is beyond based. Spot on anon

Anonymous (ID: Dgl5d/oU) 04/16/23(Sun)15:46:16 No.423780393 ▶

>>423771461

ahh your grand dad in pic.
ironically 90% of russian soldiers in winter war were from jewkrain

Anonymous (ID: TIm6vZDb) 04/16/23(Sun)15:48:12 No.423780602 ▶

>>423772597

some variant of manchu