December 2021

AI and the Future of Disinformation Campaigns

Part 1: The RICHDATA Framework

CSET Policy Brief

CSET
CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

AUTHORS
Katerina Sedova
Christine McNeill
Aurora Johnson
Aditi Joshi
Ido Wulkan

Executive Summary

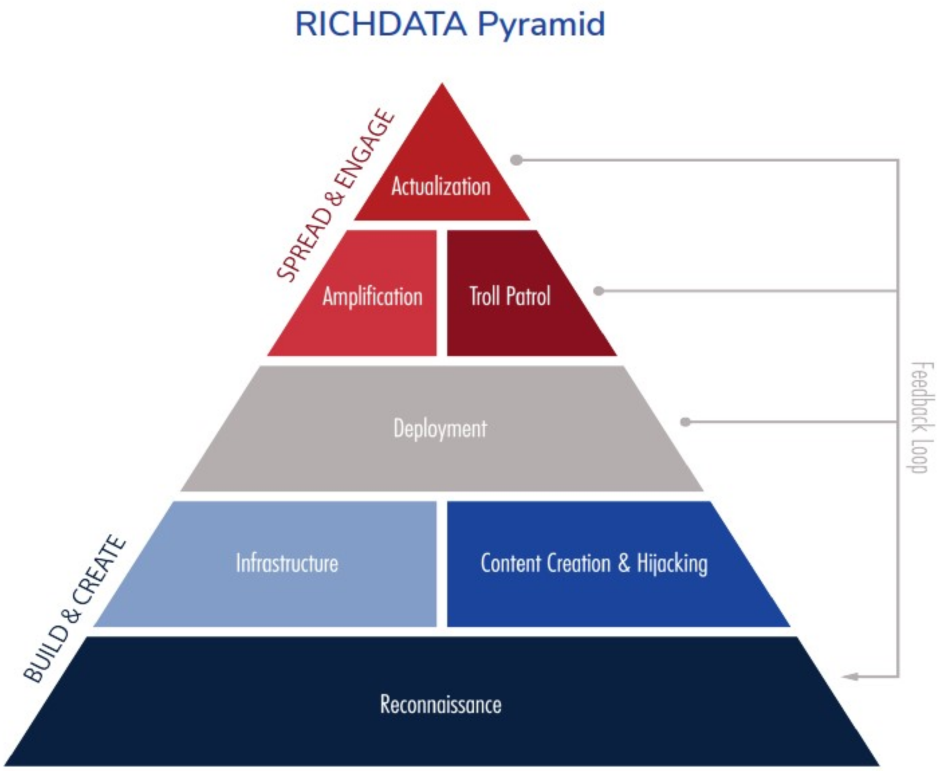
The age of information has brought with it the age of disinformation. Powered by the speed and data volume of the internet, disinformation has emerged as an insidious instrument of geopolitical power competition and domestic political warfare. It is used by both state and non-state actors to shape global public opinion, sow chaos, and chip away at trust. Artificial intelligence (AI), specifically machine learning (ML), is poised to amplify disinformation campaigns—influence operations that involve covert efforts to intentionally spread false or misleading information.[1]

In this series, we examine how these technologies could be used to spread disinformation. Part 1 considers disinformation campaigns and the set of stages or building blocks used by human operators. In many ways they resemble a digital marketing campaign, one with malicious intent to disrupt and deceive. We offer a framework, RICHDATA, to describe the stages of disinformation campaigns and commonly used techniques. Part 2 of the series examines how AI/ML technologies may shape future disinformation campaigns.

We break disinformation campaigns into multiple stages. Through **reconnaissance,** operators surveil the environment and understand the audience that they are trying to manipulate. They require **infrastructure**—messengers, believable personas, social media accounts, and groups—to carry their narratives. A ceaseless flow of **content**, from posts and long-reads to photos, memes, and videos, is a must to ensure their messages seed, root, and grow. Once **deployed** into the stream of the internet, these units of disinformation are **amplified** by bots, platform algorithms, and social-engineering techniques to spread the campaign's narratives. But blasting disinformation is not always enough: broad impact comes from sustained engagement with unwitting users through **trolling**—the disinformation equivalent of hand-to-hand combat. In its final stage, a disinformation operation is **actualized** by changing the minds of unwitting targets or even mobilizing them to action to sow chaos. Regardless of origin, disinformation campaigns that grow an organic following can become endemic to a society and indistinguishable from its authentic discourse. They can undermine

a society's ability to discern fact from fiction creating a lasting trust deficit.

Figure 1. RICHDATA Pyramid



Source: CSET.

This report provides case studies that illustrate these techniques and touches upon the systemic challenges that exacerbate several trends: the blurring lines between foreign and domestic disinformation operations; the outsourcing of these operations to private companies that provide influence as a service; the dual-use nature of platform features and applications built on them; and conflict over where to draw the line between harmful disinformation and protected speech. In our second report in the series, we address these trends, discuss how AI/ML technologies may exacerbate them, and offer recommendations for how to mitigate them.

Table of Contents

Introduction

> "Falsehood flies, and the Truth comes limping after it; so that when Men come to be undeceiv'd, it is too late; the Jest is over, and the Tale has had its Effect . . ."[2]
>
> - Jonathan Swift

Disinformation is a potent tool of geopolitical power competition and domestic political warfare, and one that has been gathering force. Sustained and well-funded disinformation operations are weaponizing the fractured information environment and creating real-world effects. Their use has generated public angst and media attention, yet their impact remains difficult to measure.[3] The word "disinformation" reentered U.S. political discourse in 2016, when Russian operators executed a campaign to influence the U.S. presidential election.[4] Dusting off "active measures" and *dezinformatsiya*—the Soviet-era terms for operations meant to meddle in the internal politics of rival nations—this campaign marked an escalation of the long-standing Russian efforts to discredit democratic institutions, sow discord, and undermine public trust.[5]

Russia reinvented the twentieth century "active measures" and repurposed common digital marketing techniques to usher in a new cyber-enabled era of asymmetric conflict in the grey zone between war and peace. Since the late 2000s, the Kremlin has waged increasingly brazen disinformation campaigns—operations meant to intentionally spread false or misleading information. These operations were initially honed against domestic opposition targets, then in the Baltics and Georgia, before reaching deadly potency in Ukraine.[6] While democratic governments and social media platforms were still learning about its new tactics, the Kremlin continued experiments to sow discord throughout Europe before setting its sights on influencing U.S. voters.

As targets changed, malicious actors innovated. Russia's early operations at home and in its near abroad used a model, sometimes termed as the "firehose of falsehood," to produce repetitive, high-volume messaging aimed at audiences on the political extremes to exacerbate tensions.[7] Bots and teams of human trolls amplified the messaging across social media platforms with support from pro-regime broadcast media.[8] The Russian campaigns launched against U.S. and European elections in 2016–2017 were informed by research into the political segmentation of target societies. They relied upon tactics such as hack-and-leak operations and forgeries to discredit political candidates unfavorable to the Kremlin. Proxy organizations made up of professional trolls worked to manipulate real people into real-world actions to exploit societal fissures, driving political polarization.[9] In recent years, these methods covertly cultivated unwitting users and amplified disinformation from homegrown conspiracy theory communities, including some anti-vaccination groups, QAnon, and domestic extremist organizations.[10]

While Russia was the pioneer, other states have also adopted these tactics. Chinese and Iranian disinformation operations, developed against their own internal opposition, are now used to sway international opinions in support of their geopolitical goals.[11] Both nations have stepped up malicious disinformation campaigns, including targeting—or contemplating to target—the 2020 U.S. presidential election.[12] Echoing well-known Russian tactics, China deployed a high-volume network of inauthentic accounts—across multiple social media platforms in multiple languages—to promote COVID-19 disinformation and conspiracy theories.[13] Today, some 81 countries use social media to spread propaganda and disinformation, targeting foreign and domestic audiences.[14]

Disinformation campaigns in the age of social media are notable for their scope and scale, and are difficult to detect and to counter. The tactics used in these campaigns capitalize on central features of today's social media platforms—maximizing user engagement and connection. They exploit the biological dynamics of the human brain: cognitive shortcuts, confirmation bias, heightened emotion, and information overload that overwhelms cognitive resources under stress.[15] They seek to deepen the natural fissures within open societies, erode trust, and chip away at the sense of a common foundation of political discourse, which is critical to a functioning democracy.

Recent advances in artificial intelligence offer the potential to exacerbate the volume, velocity, variety, and virality of disinformation, automating the process of content creation and the conduct of disinformation campaigns. A crucial question is the degree to which applications of AI will change the largely manual operations of the past. There are many unknowns. Will AI only marginally impact the scale and reach of the current tactics, or will it change the game? Can the labor-intensive operations of Russia's 2016 campaign be partially or fully automated in the future? Which techniques are likely to benefit from the rapidly evolving AI research, and how? Commentary and research have focused on synthetic video and audio known as "deepfakes." However, there are other AI capabilities, such as powerful generative language models, conversational AI chatbots, and audience segmentation techniques that may be more impactful.

In this series, we examine how advances in AI and its subfield of machine learning (ML) are likely to enhance malicious operations to spread disinformation. In this report, we analyze the building blocks of disinformation campaigns from the perspective of those who build them, outlining common techniques threat actors have used in recent years. We offer case studies that illustrate these techniques, hinting at systemic challenges that perpetuate these operations and suggest how they may evolve.

In the companion paper in this series, we identify AI technologies, such as natural language processing (NLP), Generative Adversarial Networks (GAN), and Conversational AI, that may amplify future campaigns. We then offer recommendations for how governments, technology platforms, media, and AI researchers can prepare, thwart, and respond to the malicious use of AI/ML in disinformation campaigns.

The Disinformation Frameworks

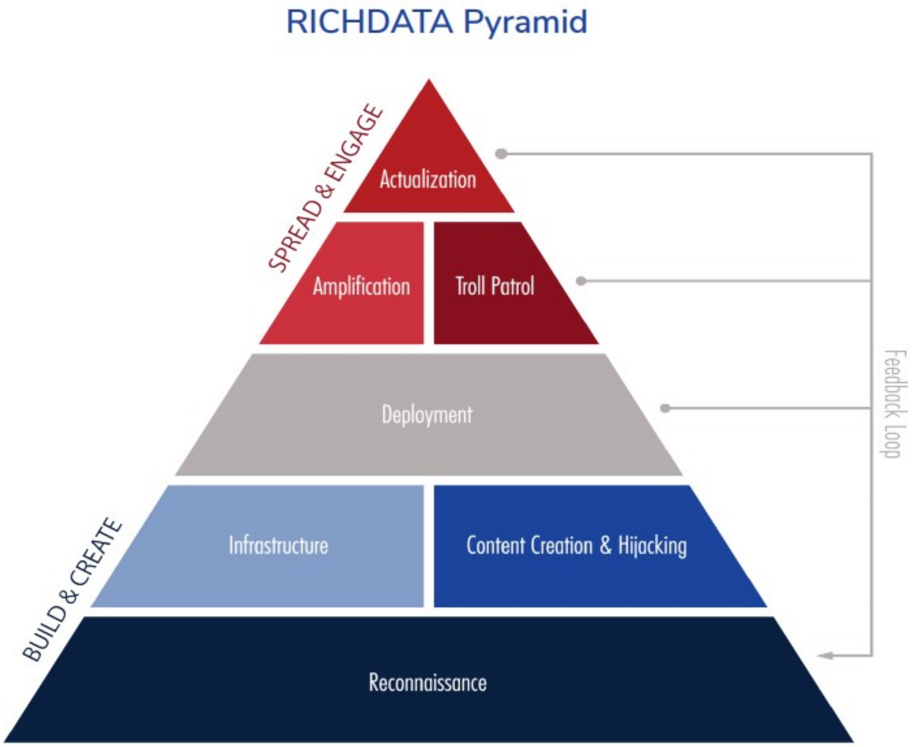
Though no two disinformation campaigns are alike, they share common elements and follow patterns that exploit the underlying features of social media platforms. A number of researchers and organizations have built frameworks to conceptualize various aspects of these campaigns, including phases, objectives, types of messages, and tactics.[16] One conceptualization used within cybersecurity is that of the "kill chain," which describes the various phases and associated tactics and techniques a hacker may use when conducting a cyber operation.[17] Clint Watt's "Social Media Kill Chain," the U.S. Department of Homeland Security's "Disinformation Kill Chain," and data scientist Sara Jayne-Terp's Adversarial Misinformation and Influence Tactics and Techniques framework have all applied similar models to disinformation operations. [18]

It is challenging to capture the full spectrum of disinformation operations in a single framework explaining the various ecosystems of actors, tactics, and platforms. There is an additional difficulty of how to account for the role of traditional media in the ecosystem, whether as the target, the platform, or the witting or unwitting agent of influence. Given this complexity it is perhaps not surprising that the community researching, disrupting, and countering disinformation among the government, private sector, and civil society stakeholders has not coalesced on the single model akin to the cyber kill-chain. In this paper, we draw on the above frameworks and other research efforts to build a threat-model of disinformation campaigns.

The RICHDATA Architecture of Disinformation

Understanding the key stages of disinformation campaigns offers insight into how AI may enhance them. While strategies and intent vary by the actor and target—from sowing chaos to shaping public opinion—they share common tactics. To aid in understanding, this paper provides a framework, which we term RICHDATA, based upon the various stages often seen in disinformation campaigns, as depicted in Figure 1. A continuous feedback loop runs through these concurrent stages to assess efficacy, monitor engagement, and refine the tactics. The pyramid shape reflects the greater importance of preparatory stages that are often overlooked in discussion of disinformation yet can be the most valuable and time-consuming for operators. The depicted stages can occur concurrently and are dynamic, evolving based on the feedback received throughout the campaign. The RICHDATA acronym reflects the critical role data and digital footprints play in both disinformation operations and machine learning.

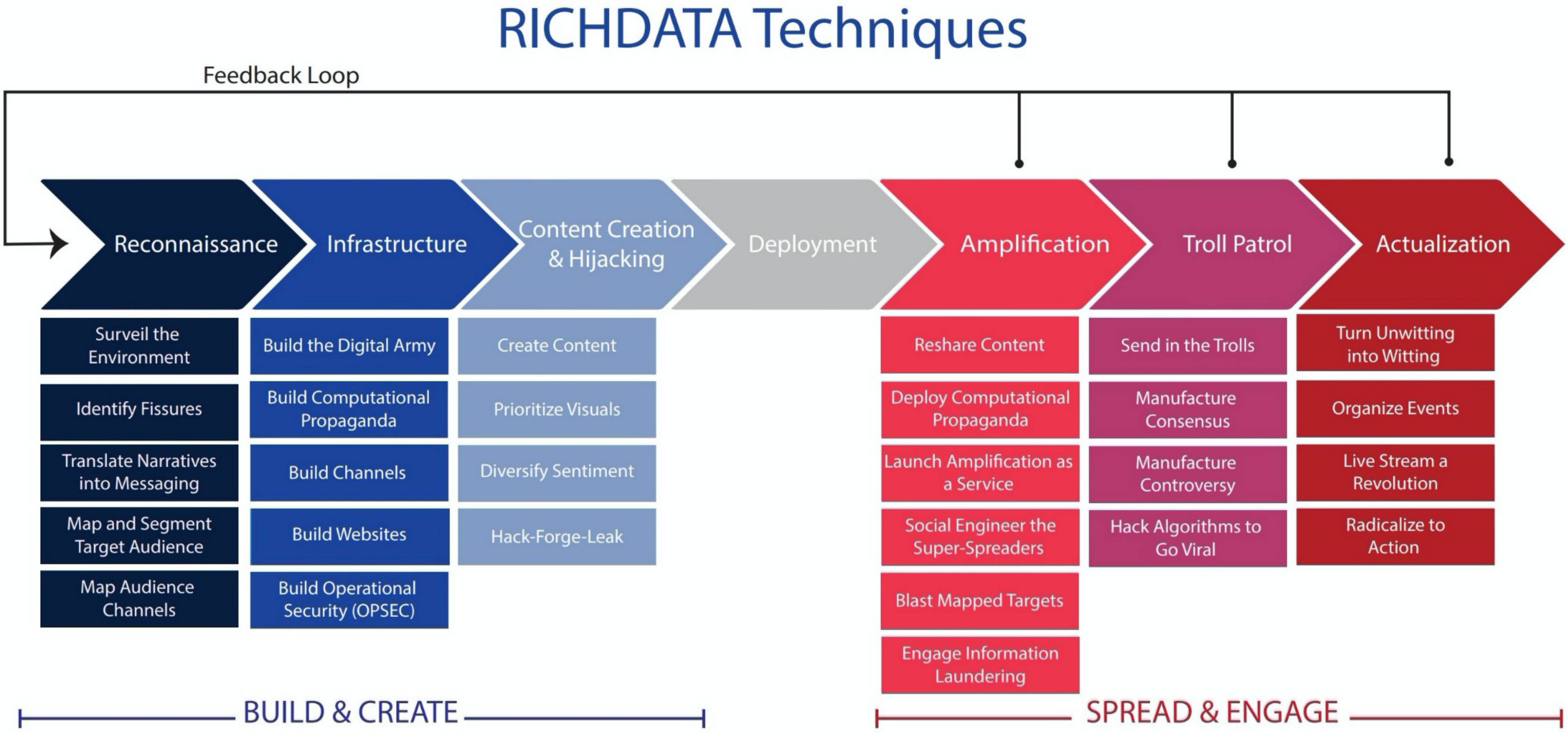
Figure 1. RICHDATA Pyramid



Source: CSET.

In popular discourse, the term “disinformation” is often associated with bots and trolls on social media.[19] However, they are just a part of the picture. In this section, we describe various techniques disinformation operators use at each stage of campaigns, though this is by no means an exhaustive list. Defining the manual and automated parts of the job allows us to frame the spaces where AI/ML may enhance traditional automation, and those where operations are more likely to be run by humans.

Figure 2. RICHDATA Techniques



Source: CSET.

(R) Reconnaissance: Understanding the Audience

Disinformation campaigns often begin with an examination of the target society and its information environment in order to shape and target their messages. Disinformation operators read and watch news and entertainment to understand cultural references, tracking content on traditional and social media to identify the themes and influencers driving the conversation. To perform this labor-intensive process, they build teams with the necessary language and data analytics skills, as the Internet Research Agency (IRA), the infamous Russian professional "troll farm," did in their campaign to influence the 2016 U.S. election.[20] Similarly, throughout its 2014–2015 military campaign in Ukraine, Russian operatives closely studied Ukraine's media and political landscape, producing daily reports including topics covered in the local media, and the number of citations for the pro-Russian term "Novorossiya." [21]

Operatives seek to understand the historic grievances and fault lines of the target society in order to shape broad narratives into their specific messages and talking points. In 2014, Russian operators depicted Euromaidan and the Revolution of Dignity in Ukraine, mass protests against the corruption of the Yanukovych regime, as a "fascist coup." They tapped into a sensitive fault-line—the collaboration of some Ukrainian nationalists with Nazi Germany during World War II against the Soviets in return for the unfulfilled promise of Ukraine's independence. Contemporary Russian narratives continue to project the actions of a few Ukrainians onto the entire nation.[22] They also ignore the contribution of Ukrainians to the underground resistance and to the allies' war effort on the eastern front.[23] Historical narratives are a potent tool, and malicious operators can use them to develop specific themes that contain messages and keywords to include in their social media posts.[24] Messages are more likely to resonate when they align with a target's pre-existing worldview, connect to a grievance, or contain a kernel of truth.[25]

Operators exploit societal fissures to segment audiences by their positions on divisive issues. They identify core influencers in each segment, as well as potential "super-spreaders" with large social media followings: public figures, officials, activists, and journalists. Manipulating super-spreaders into sharing a disinformation campaign message can pay off with broad resharing and even traditional media coverage.[26] The fact that the super-spreaders are unwitting perpetrators also helps to evade countermeasures by social media platforms.

Disinformation operators also exploit evolving crises and the ambiguity inherent in incomplete information with messaging to target audiences across the political spectrum. China's and Russia's efforts to target audiences in the United States and around the world with disinformation narratives about COVID-19 origins, treatments, and the efficacy of U.S. response present a compelling demonstration.[27] They also underscore the importance of developing narratives that tap into deeply rooted values and existential fear.

Exploiting Ambiguity Case Study: COVID-19 Disinformation

Incomplete and evolving information in a crisis, such as a global health pandemic, presents opportunities for disinformation actors to fill voids with alternative narratives that feed panic and undermine trust in government institutions and expertise. Aiming to sow fear and chaos, early in the pandemic, a China-linked source spread a message via social media and text messaging apps that warned Americans of the impending lockdown by the U.S. federal government.[28] Capitalizing on incomplete information about the origins of COVID-19 and U.S. conspiracy theorist's claims linking U.S. scientists to the "gain of function" research, Chinese and Russian actors amplified the narratives that COVID-19 was engineered by the United States, exploiting the anti-government sentiment on the U.S. political extremes.[29] While this message did not resonate with U.S. audiences, others, such as those that promoted doubts about vaccines, spread more widely. The efforts to fuel vaccine hesitancy and discredit vaccines made by European and U.S. firms are ongoing. In August 2021 Facebook banned an inauthentic network originating in Russia that claimed AstraZeneca vaccine would turn people into chimpanzees and Pfizer vaccine caused higher casualties than other vaccines. [30] The network was linked to Fazze, a UK-registered marketing firm operating in Russia, which also recruited YouTube influencers to claim that Pfizer vaccines were deadly.[31] The campaign supported its claims with a document allegedly hacked and leaked from AstraZeneca, which in reality was forged from different sources and taken out of context.[32] This activity targeted countries where Russia wanted to sell its own Sputnik-V vaccine and coincided with the amplification of the same narratives on Russian state-sponsored media and false news outlets operated by Russian intelligence.[33] Ironically, Russia's promotion of vaccine hesitancy around the world is backfiring at home: as of June 2021, only 11 percent of Russian population had been fully vaccinated.[34]

Having segmented their audience, operators identify the best channels to reach them. In their reconnaissance efforts targeting the U.S. 2016 elections, Russian operators researched groups on social media dedicated to political and social issues, tracking metrics such as size, frequency of content, and the level of audience engagement.[35] To aid in their audience segmentation efforts, operators may leverage the automated advertising features of social media platforms. As of 2019, digital advertisers could target their audiences through hundreds of thousands different attributes on mainstream platforms.[36] Anyone with a personal Facebook account can test the advertising mechanisms, targeting audiences based upon demographic data (e.g., age, gender, location) and the more personal information users share with the platform: interests, hobbies, politics, and life events.[37] Twitter provides potential advertisers with information on what a user searches for or views, their location, and apps installed on their device, as well as targeting based on interests, attended events, and an option to target a specific account's followers.[38] Microtargeting through advertising remains relatively easy, despite recent restrictions and changes to advertising policies to limit political ads on major platforms.[39]

(I) Infrastructure

Key elements of campaign infrastructure include a digital army of fake personas, computational propaganda tools (e.g., automated accounts), channels for distributing disinformation (e.g., social media groups and pages), and websites for hosting false content. In addition, threat actors need tools to cover their tracks, and financial infrastructure to purchase ads, such as PayPal accounts or credit cards under stolen identities.

Inauthentic personas, or "sock puppets," are the digital soldiers of an online disinformation campaign. Threat actors can create these personas in-house, acquire them on the dark web, or rent them from the growing coterie of firms that offer influence as a service. They can also rent authentic accounts of witting or unwitting humans or pay them to propagate campaign messages. As social media platforms get better at identifying inauthentic activity, threat actors increasingly resort to burner accounts with a short life span that blast disinformation widely and then disappear. These accounts may be effective for seeding a conspiracy theory narrative or injecting forgeries into the information stream, but seldom reach the audience engagement levels of curated inauthentic personas.[40]

Threat actors place a premium on messengers that appear authentic. These personas can be highly curated, including a profile photo, biographic data, presence across multiple platforms, and a history of relevant posts. To develop realistic personas, threat actors scrape photos from public social media profiles, stock photos, or images of lesser-known celebrities. Operators increasingly use new ML techniques that can generate realistic human faces, providing a pipeline of untraceable avatars. These are becoming more important because threat hunters can use a reverse image search to identify stolen photos. A degree of credibility is necessary to set up new groups or hijack existing ones, and to evade automated detection systems, which look for accounts with a lack of history of posting locally relevant content, fewer connections, and recent creation date.[41]

Building Personas Case Study: The Russian IRA Impersonates the American Political Spectrum

The IRA's posts from a small group of highly curated accounts—created with a focus on credibility—were the most successful in getting traction online. Some of the most active inauthentic accounts in the IRA's disinformation campaign against the 2016 U.S. presidential election were personas imitating Black Lives Matter activists and the Tennessee Republican Party.[42] Out of 1,500 fake accounts the IRA deployed in the English-language operation, this group of 300 personalized accounts garnered 85 percent of all retweets in English (18.5 million retweets of 1 million tweets).[43] These accounts were "consistent, persistent, political personas that reflected caricatures of U.S. political participants" and evolved to "behave like real people."[44] These personas allowed the IRA to infiltrate existing grassroots activist communities on both sides of the political spectrum, in some cases coordinating with real activists to organize protests.

Threat actors in a hurry can outsource persona creation and rent a digital army. The growing "influence as a service" industry ranges from seemingly legitimate marketing and PR firms to outfits selling "armies" of inauthentic accounts, likes, and comments via the dark web.[45] This practice, however, carries risks. These inauthentic accounts may have been used in other campaigns and can be easier to detect. Nevertheless, they can be useful to amplify or inflate follower numbers for a narrowly targeted campaign.

As platforms get better at detecting fake accounts, operators may resort to purchasing or renting authentic user accounts. The technique of "franchising," paying witting or unwitting individuals to establish online personas and spread disinformation, has been documented several times, including during the 2019 elections in Ukraine and in the 2020 IRA-linked operation in Ghana and Nigeria targeting the United States.[46] This technique makes operations look as if they are perpetrated by authentic users. Due to heavy human resources, such operations may be more expensive to scale and sustain overtime.

Threat actors with limited resources, or those looking for a quick ramp up, may use automation to build accounts en masse. This tactic is part of what some researchers call “computational propaganda,” the use of automated accounts (bots) and scripts to publish posts.[47] Social bots are user accounts automated to interact with other user accounts.[48] This involves writing computer code that will talk to the platform application programming interface (API), automatically generate bot accounts, connect them into networks of bots, aka 'botnets', and automate posting. While the platforms have enhanced countermeasures to detect botnets, their presence persists. In the 2020 U.S. election, bots accounted for 13 percent of users engaging with conspiracy theory content and reshared hyper-partisan content “in a systemic effort to distort political narratives and propagate disinformation.”[49] In recent years, the use of bots has grown exponentially, fueled in part by availability of open-source code to build them at more than four thousand GitHub sites and over forty thousand other public bot code repositories on the internet and the dark web underground market.[50]

Accounts created by automation often betray their inauthenticity through metadata or behavior. Telltale signs include missing avatars, mismatched male/female names, names that resemble computer code, or accounts created within seconds of each other. These accounts may also publish posts at the same intervals and retweet every few seconds, faster than the average human could. These technical markers can be detected by the social media platform algorithms.[51] Russian IRA operators built a human operation in part because their bots were getting shut down, a lesson that China-linked actors also learned in their 2020 “Spamouflage Dragon” campaign.[52]

Having built personas to propagate their messages, disinformation operators may create groups and pages to begin building an audience. These need time to build credibility. Often groups start with pre-populated inauthentic personas to give the appearance of legitimacy before they can attract authentic members. One of the

IRA's top performing Facebook pages, "Blacktivist," cultivated a large community with messages of empowerment aimed at African Americans and garnered 11.2 million engagements from authentic users.[53]

Operators may also create general interest pages on innocuous subjects, such as food or tourism, to reach a broader audience of authentic and inauthentic users and then pivot to disinformation content later.

Building Channels Case Study: A Road Trip to Falsehood

A covert Russian network of accounts and pages linked to Rossiya Segodnya and Sputnik, Kremlin-affiliated media outlets, encouraged followers to travel around the Baltic nations, featuring food and health tips,[54] and occasionally interjecting Sputnik stories with pro-Russia and anti-NATO messaging.[55] Through the use of innocuous content in its 266 pages, the network amassed an audience of 850,000 followers, ready to ingest a coordinated disinformation campaign. Come for the pictures of beautiful Riga, stay for the falsehoods.

Disinformation operatives use audience-building techniques such as infiltrating authentic groups. Social media platforms allow and encourage the creation of interest-based subpopulations, giving threat actors easy targets for infiltration and ready-made audience segmentation. In response to the privacy concerns following the 2018 Cambridge Analytica scandal, Facebook's change to promote private interest-based groups more prominently on users' feeds had unintended consequences.[56] Promotion of these private groups, some with tens of thousands of members, opened an avenue for malicious actors to identify politically interested communities, infiltrate them, and build trust before deploying deceptive content.[57] While Facebook scaled back on promotion of political interest groups on its platform and now requires stringent content moderation from group administrators, the avenue of hijacking innocuous Facebook groups to push disinformation remains open.[58] Telegram users can create channels of two hundred thousand members and protect messages with end-to-end encryption and self-destruct capability.[59] Features designed to build trusted spaces can give operators another way to target and build an audience, presenting an ongoing challenge for social media platform engineers, policy managers, and threat hunters.[60]

Tools of operational security (OPSEC) are a key component of building or acquiring infrastructure. As disinformation operators create personas, activate channels and, in some cases, build a web presence, they use tools to hide their operations. In a well-run operation, threat actors cover their tracks as they register domains and create inauthentic accounts. This includes identity-protection measures like proxy servers and virtual private networks (VPN) to mask their IP addresses as they register domains and create accounts to appear to originate from a geographically appropriate location. The reuse of email addresses, domain registrars, or hosting services can expose the operator.

Technical markers identify disinformation operations, so advanced actors have to be hypervigilant. In one case, a small slip in operational security—the failure to use a VPN while accessing a social media profile of "Guccifer 2.0"—allowed U.S. intelligence analysts to trace this inauthentic persona to an officer at the Moscow headquarters of the Russian military intelligence service (GRU). This proved that Guccifer 2.0 was a Russian operative and not a Romanian hacker claiming to have hacked and leaked emails from the Democratic National Committee.[61] That said, some operators want to be discovered. Creating a perception that they have greater impact than it is in reality, known as "perception hacking," helps malicious actors weaponize distrust, play into the societal expectations of foreign interference, and sow doubt in the integrity of institutions.[62]

(CH) Content Creation and Hijacking

Content is the fuel of disinformation campaigns. After identifying narratives and target audiences, operators seek to create a steady stream of engaging content to shape the saturated information environment, marked by a 24-hour news cycle and competition for attention.[63] The process of content development is iterative and labor intensive, demanding responsiveness to user engagement, refinement, and a ceaseless stream of content.

During the Russian IRA operations, six hundred operators worked in its 40-room St. Petersburg headquarters, behind doors with labels such as “graphics,” “blogging,” “video production,” and “social media specialists.”[64] Managers placed a premium on creating compelling content that would go viral organically.[65] The human content farm churned out thousands of posts for the required weekly “technical tasks” reiterating its themes through talking points and “must have” keywords.[66] At the height of the 2016 campaign, one hundred operatives from the IRA’s American Department created one thousand unique pieces of content per week, reaching 30 million users in the month of September, and 126 million over the course of the election, on Facebook alone.[67] This scale requires a dedicated team, skilled in creating specific types of content, with nuanced command of the language and culture of the target society.

Operators create content to elicit a range of emotional responses: posts that are uplifting and positive, angry and negative, or neutral and unrelated to the campaign. Different types of content are useful for different campaigns. Neutral content helps build audience and funnel traffic towards pages operated by the threat actors, while positive posts can help operators infiltrate and gain credibility with authentic groups. They can also garner more engagement. Counterintuitively, the Russian IRA’s posts with most likes and reshares on Facebook leading up to the 2016 U.S. election were benign and positive, rather than polarizing, in emotional profile.[68] That said, threat actors continue to deploy negative messaging in campaigns when it serves their purposes. For example, COVID-19 disinformation efforts targeting Indo-Pacific countries have predominantly used negative emotional hooks, such as anger, in their messaging.[69]

The types of content range from short messages and medium-length articles to visual media. Operators can hijack posts from authentic users, screenshot and repost them with a slant. Screenshotting helps hide the operator’s origins, minimizes the effort entailed in creating content from scratch, and retains the authentic voice of local users. These techniques can help avoid common grammatical errors or misused idioms that can be red flags to a vigilant target audience or a social media threat investigator.

Visual content and memes, in particular, are a potent tactic of disinformation. The interplay between image and text increases the virality of memes and challenges platforms' detection algorithms.[70] Memes capitalize on both the human ability to quickly process imagery and their own quality to “shed the context of their creation.”[71] Once amplified, the best memes lose ties to their creators, are reinterpreted, and integrate into the consciousness of the internet, becoming collective cultural property.

Prioritize Visuals Case Study: #DraftOurDaughters Memes

The “#DraftOurDaughters” campaign, created by anonymous users on the 4chan fringe network in the final week of the 2016 U.S. presidential campaign, spread falsehoods to mainstream platforms about candidate Hillary Clinton.[72] Disinformation operators pushed memes masquerading as Clinton campaign ads, alleging the candidate’s intention of instituting a mandatory draft of women into the military.[73] Private individuals found their wedding photos scraped and turned into memes with the "#DraftOurDaughters" and "#DraftMyWife" hashtags.[74]

From their innocent origins, memes offer a potent tool[75] that operators have weaponized not only against political candidates but also journalists and ethnic minorities.[76] In 2016–2017, the Myanmar military pushed anti-Muslim Facebook posts against the Rohingya population before the platform shut them down. Offensive imagery thinly veiled in a cartoon form contributed to the spread of violence through their ease of interpretation and crude attempts at humor, unleashing real-world horror.[77]

The leaking of embarrassing or compromising content, sometimes laced with forgeries, is a key tactic used by threat actors.[78] Threat actors can develop forgeries and leak them, claiming they are authentic hacked information. They can hack targets' network to acquire compromising content or purchase hacked content on the underground market, then enhance this illegally obtained material with forgeries.

(Hack)-Forge-Leak Case Study: Ghostwriter and Secondary Infektion

Hacking and digital forgeries are a tactic of choice for Russian state-affiliated actors. Operation Ghostwriter distributed forged military documents alleging that NATO troops were responsible for the spread of COVID-19 in Europe. It also published anti-NATO op-eds authored by fake personas impersonating journalists and invoking forged “leaked” emails between officials at NATO and national governments of Estonia, Latvia, Lithuania, and Poland.[79] Leading up to the 2021 German elections, actors connected to this group escalated their efforts to hack members of the German Parliament to obtain compromising information for use in pre-election disinformation campaigns. This prompted German and European Union authorities to call on Russia “to adhere to norms of responsible behavior in cyberspace.” Despite this call, security researchers at Mandiant recently linked this campaign to Belarusian government, suggesting that Russia’s techniques are diffusing. [80] Similarly, Operation Secondary Infektion targeted the United States and its allies with forgeries aiming to stoke diplomatic tensions between NATO allies and partners.[81] In six years, over 2,500 pieces of forged content appeared in seven languages on three hundred platforms.[82] Only a small portion of activity surfaced on Facebook, with most of the campaign percolating through small platforms that lacked threat hunters.[83] Higher quality digital forgeries would make this tactic more effective, particularly if mixed with authentic leaked documents. Secondary Infektion showed the difficulty of detecting operations that are rolled out across multiple platforms, an increasing trend.

(D) Deployment

After setting up infrastructure and crafting content, operators put their preparation into action by interacting with authentic users. In the deployment stage, operators identify and activate the persona(s) that will deliver the initial payload. They prepare the payload in the chosen form, such as a blog, a video, a social media post, or a meme. They then drop the payload into targeted channels, such as a conspiracy theory forum, a social media page, a false news website managed by the operators, an authentic pre-infiltrated group, or a persona's social media feed.

Disinformation actors often begin their campaigns away from mainstream social media platforms, which are moderated. They may instead seed disinformation content to an obscure website or channel and reference it repeatedly on mainstream platforms. In recent years, several state-sponsored actors have deployed content in the form of false news stories and research by establishing fake news sites and think tanks. In 2018, an Iranian disinformation operation used six public-facing websites to push content aligned with Iranian national security interests.[84] The sites linked into a web and hosted both original and hijacked content. Likewise, Russian actors operate multiple proxy sites, such as Strategic Culture Foundation, SouthFront, and Geopolitica.ru, and publish disinformation disguised as analysis in support of the Russian government interests.[85]

(A) Amplification: Pushing the Message

With content created, operators turn their efforts to amplification—exposing their message to the maximum number of eyes and ears, getting it picked up by trending topics or hashtags and spread organically. If authentic voices engage and unintentionally spread malicious content, content moderation is more challenging for the platforms.

Operators push the payload across multiple social media platforms to maximize exposure. Research on cognition and disinformation stresses that the first engagement with false information leaves a lasting imprint, particularly if it is seen via trusted networks of personal connections.[86] For this reason, efforts to debunk disinformation often fall short of the scale and reach of the initial interaction and fail to fully dissuade the previously targeted audience.[87] A diversity and variety of channels all pushing the same deceptive message enables threat actors to mask coordination, appear authentic, and ensures repetitive, high-volume messaging.

Resharing Content Case Study: China's "The Spamouflage Dragon"

In 2020, China launched a disinformation operation in an effort to shape positive perceptions of its handling of the COVID-19 pandemic. The campaign, conducted on Twitter, relied on 23,750 core posting accounts and 150,000 additional amplifier accounts designed to "artificially inflate impressions metrics and engage with the core accounts."[88] While the majority of these were in Chinese, around 9.4% were in English and 1.8% - in Russian.[89] The English-language tweets focused on COVID-19, pushing narratives from prominent Chinese state media accounts and Chinese officials, arguing that China was both transparent and efficient in their response to the outbreak.[90] Technical indicators linked this operation to the same actor spinning a pro-PRC portrayal of Hong Kong protests and responsible for a network of two hundred thousand accounts previously suspended by Twitter in August 2019. The network "Spamouflage Dragon" also posted video content extensively on YouTube and amplified the same content across Twitter and Facebook.[91] Recent findings suggest that the same network operated far beyond the mainstream platforms: posting thousands of identical messages across 30 social media platforms and 40 other online forums in additional languages, including Russian, German, Spanish, Korean, and Japanese.[92]

Operators may also use the tools of computational propaganda, launching bots to barrage the followers with queued-up posts linked to payload content. Bots can automate tasks such as rapidly posting links to media outlets or categorizing tweets by high-volume influencers. They can also automate interactions with other user accounts and hijack conversations.[93] In the context of disinformation campaigns, bots act maliciously by impersonating real users. Networks of bots or botnets can act in concert to distort conversations online by driving up “likes” and retweets, hijacking trending hashtags and topics by spamming them with unrelated content or sharing links to websites that publish disinformation.[94]

For example, a legitimate nonprofit group created the hashtag #SaveTheChildren to spread awareness about child trafficking and fundraise to combat it. Followers of the QAnon conspiracy theory hijacked the hashtag, expanding their reach through misuse of the platform features.[95] These techniques draw attention of unwitting users to the topic or a hashtag and expose them to the disinformation payload.

Threat actors may also deploy custom or third-party applications that leverage platform features to manage multiple bots and botnets. Many platforms expose hooks into their platform for programmers to encourage the development of engaging apps and advertising tools within their platforms. While users experience the social media platforms through the front end of a website or an application, bots access information and functionality of the platforms by “talking” to the API through code.

Through this access, digital marketing applications can automate the propagation of content for campaigns across multiple social media networks. This allows bots to collect the data from the social media site more efficiently than from the front-end-user’s view of the platform, and then complete any action that a regular user can, including posting and interacting with content. Through the Twitter API, researchers and businesses can access conversations in real time, analyze past related posts, measure tweet engagement, monitor for news events as they break, or explore a specific account and its history in depth.[96] Services like MasterFollow, Botize, and UberSocial allow users to upload large amounts of content, manage delivery schedules, connect multiple automated accounts, and integrate bots across multiple platforms.[97] In an effort to combat spammy behavior, Twitter increased restrictions in access to data through the API, particularly by governments, and prohibited users and applications from simultaneously posting "duplicative or substantially similar" tweets across multiple accounts.[98] However, its policy still permits automation of multiple accounts for "related but non-duplicative use cases."[99] This enables coordination as long as accounts appear adequately authentic and the messages are sufficiently varied.

These programming hooks have many legitimate uses but can be misused by threat actors. The “good” bots that push notifications from the traditional news outlets to Twitter feeds are built on the same infrastructure as the “bad” bots that push disinformation payloads. The same APIs can enable malicious actors to post high volumes of disinformation content and ensure a continuous and repetitive stream of content across platforms. This dual-use nature of API access requires the platforms to perform nuanced detection and vetting to distinguish legitimate use and misuse, particularly as disinformation actors can masque their actions through third-party applications and firms.

It is not just bots that may spread the message. In 2020, private firms in 48 countries have offered automated and human-powered inauthentic amplification as a service and deployed it on behalf of a political actor to shape public conversation, domestically or internationally.[100] Since 2009, these firms have spent almost $60 million on computational propaganda. These firms offer a variety of services including the creation of inauthentic personas (sock-puppet accounts), micro-targeting, and message amplification with fake likes and fake followers, allowing threat actors to outsource disinformation operations and complicate attribution.[101]

While the marketing claims of some of these “amplification as a service” firms can be dubious or overblown, threat actors are turning to them. A large takedown by Facebook in July 2020 implicated a Canadian political strategy consulting firm Estraterra in a campaign of coordinated inauthentic behavior. The campaign targeted Ecuador, El Salvador, Argentina, and other countries in Latin America around elections, aiming to manipulate political debate.[102] This technique has cropped up around the world, including the case of an Indian public relations firm creating inauthentic bot accounts to spread an anti-Saudi and pro-Qatari narratives in the Gulf.[103] The services can be relatively inexpensive. In Taiwan, “cyber armies” cost as little as $330 per month.[104]

In addition to relying on bots or outsourcing, operators can also turn to key individuals in the social media ecosystem, super-spreaders, the “verified” influencer accounts belonging to public figures, celebrities, politicians, media personalities, and journalists, whose audience reach helps increase distribution. When journalists discover and report on public figures who have been duped into spreading disinformation, the additional media coverage further exposes and amplifies the campaign’s messages.

Hack and Leak Case Study: UK Trade Leaks

Secondary Infektion threat actors combined hack and leak techniques with superspreading to powerful effect in the 2019 UK parliamentary election. The threat actors known for their six-year campaign that used forgeries of U.S. and European government documents deviated from their tactics by injecting leaked authentic documents containing details of the US-UK trade agreement negotiations.[105] Oblivious to their origins, Jeremy Corbyn, the leader of the Labour Party, re-amplified the leaked documents at a campaign event. The media picked up the story, assuring wall to wall coverage for days leading up to the election.[106]

The implications of covering hacked material by the press are far reaching. Together with domestic political actors with vested interest in spreading such material to undermine their opponents, the media can play an important, if sometimes unwitting, role as a vehicle of amplification.[107] In 2016, the disinformation operation by the GRU precisely targeted the media into amplifying leaks of hacked and slightly forged information from the Democratic National Committee, the Democratic Congressional Campaign Committee, and the Clinton campaign.[108] The hackers integrated the stolen information into the broader campaign to manipulate the media into promoting the material. The coverage played into the narrative of Secretary Clinton and “lost emails.” Mainstream media is in a difficult position when deciding whether to publish hacked and leaked material, and must determine whether the public's right to know outweighs the potential harm caused by unwittingly amplifying a disinformation campaign. As long as there is lack of consensus on disclosure, operators will continue to rely on this tactic.

Distribution of malicious content is not limited to super-spreaders or curated groups, as operators may spread bespoke payloads into authentic online communities representing different sides of a divisive issue. By infiltrating authentic communities, threat actors can blend in, exploit the preexisting biases, and ensure the disinformation seeds can root and grow to spread organically to other adjacent sympathetic audiences.

Homegrown conspiracy theory communities provide fertile ground for "information laundering" by foreign and domestic actors.[109] This technique draws narratives from fringe sources, cultivates them, and amplifies them into the mainstream to gain legitimacy. Researchers attribute the rapid growth of QAnon conspiracy communities in part to the technique of cross-posting QAnon content into loosely related groups on mainstream platforms, quickly drawing them into a single anti-establishment tent.[110] Threat actors have leveraged these communities as a pipeline of ready-made disinformation that fits their goals—to undermine trust and sow discord. Russian disinformation actors historically cultivated and exploited conspiracy theories, and in recent years have engaged and amplified anti-vaccination groups and the QAnon networks. [111] Chinese disinformation operators likewise resorted to this technique to fan COVID-19 origin conspiracies as we described in the COVID-19 disinformation case study earlier in this paper.

Operators can use the recommender algorithms of social media platforms to their advantage, both to identify adjacent interest-based groups and communities to cross-pollinate and to draw more users into the conversation. One technique is to manufacture debates to generate activity, artificially inflate engagement, and make the online dialog appear lively for any authentic observer. Because many platform algorithms prioritize high user engagement, this inflated activity causes social media recommendation systems to push content with higher levels of engagement across more networks, thereby exposing more eyes to the disinformation content. We discuss recommendation algorithms in the second installment of this paper series, which focuses on AI/ML technologies.

(T) Troll Patrol: Controlling the Message

Disinformation thrives on dynamic and nuanced engagement. While blasting deceptive narratives across the social media ecosystem in a one-to-many amplification campaign can get the message before the maximum number of people, organic debates help disinformation to gain a foothold and thrive. Legitimate users are forced to expend cognitive and emotional resources defending their positions, and the platform’s algorithms are likely to prioritize lengthy debates and push them into additional feeds. This is a theatrical and labor-intensive exercise. Whether the goal is to cement a specific message or to sow discord, inauthentic personas engage authentic users—and each other—in elaborate back-and-forth arguments. Malicious actors measure their success—and the size of their paycheck—in the volume and the intensity of exchanges that drive up controversy and draw in authentic viewers.

Controlling the Narrative Case Study: The Downing of MH17

In the immediate aftermath of the MH17 crash over eastern Ukraine in the summer of 2014, trolls swarmed a variety of posts about Russia's lack of culpability, peddling many shades of falsehood.[112] The IRA's Ukraine operation monitored related conversations and steered the arguments towards several familiar themes in an effort to control the narrative. Some suggested that the flight was full of passengers who were already dead. Others pointed the blame at Ukraine with narratives such as "the Ukrainian air force shot down the plane" or "the missile came from the Ukrainian government-controlled territory." These claims were debunked by satellite imagery, intercepted communications between Russian proxy forces in the Donbas, and open source intelligence tracing the journey of the Russian missile system across the Ukrainian border.[113] These efforts infused falsehoods into the comment sections and social media pages of reputable international media outlets, leaving audiences confused and misinformed.[114]

Manipulation of the target audience is often dynamic, iterative, and agile. Human trolls with "hands on the keyboard" switch methods based on the desired effect.[115] They may use provocation techniques to start arguments within posts, private groups, and on media publications by adding comments that raise the temperature of debate. In other cases, trolls may subvert authentic conversations by sharing information that advances the troll's narrative. Through social engineering, they may incite users to join a group, observe a boycott, or vote for a candidate. They use ad hominem attacks to discredit authentic users, delegitimizing specific positions. Operators use diversion techniques and nuisance attacks to subvert a thread, derailing discussion and irritating other users. Swarming techniques "flood the zone" with a counter narrative.

If the goal of a campaign is to drive a particular message, trolling techniques must control the narrative. Trolls swarm threads and debate legitimate users in order to route the conversation in the desired direction. Some trolls coordinate entirely fake debates.[116] In an interview with *The Washington Post*, a former IRA troll described operatives creating plays in three acts. One would post a negative opinion about an issue. The others would comment in disagreement with supporting links. Eventually the first naysayer would succumb to the "evidence" and declare themself convinced, providing an illusion of consensus to unwitting audiences.[117]

If the goal of a disinformation campaign is to disrupt, divide, and antagonize, trolling techniques focus on provoking the target audience to react emotionally. The more dialog and greater degree of manipulative responses in comments, the higher the volume of engagement from authentic users. Accusing a threat actor of being a troll or spreading disinformation tends to lead to denial, counteraccusation, or "what-about-ism"—all aiming to push the target into a heated argument.[118]

As with amplification, the task of the trolls is to drive up engagement. Professional trolls are paid by the number of comments they post and reactions they get, while real people may be drowned out, whipped into a fury, or even driven to disconnect.

(A) *Actualization: Mobilizing Unwitting Participants*

To sustain itself, and to avoid detection, a campaign needs to attract authentic messengers. Ideally, disinformation operators can reduce their direct involvement, as the target audience, comprised of authentic users, begins to create organic content to support the injected narratives. At this stage, the line between foreign and domestic disinformation blurs, as does the line between intentional disinformation and unintentional misinformation. The campaign may move into the physical world, culminating in the mobilization of unwitting participants through protests and rallies. The pinnacle of a disinformation campaign is the creation of a state of contradiction: a readiness to mobilize and a disempowerment to act. At this stage, a threat actors' approach is targeted and more closely resembles the activation of an intelligence asset rather than widespread opportunistic and exploratory messaging fanned through the social media ecosystem by bots. If a campaign is to have a sustained effect and continue rooting and growing in the target society, it needs authentic messengers and an organic following. Threat actors may precisely identify susceptible individuals, cultivate them into an operational role, and incite them into political action. Radicalized susceptible individuals can step into the roles of disciples and carry on the campaign.

Actualization and Mobilization Case Study: From Hot Dogs to High Noon

Disinformation campaigns may culminate in real-world confrontation. Early in its U.S. operations, the IRA tested its capacity to have an impact on the ground half a world away. In 2015, operatives posted an event on Facebook offering free hot dogs at a certain time and place in New York City. They watched through a live cam as New Yorkers showed up looking for the promised hot dogs. The test exceeded expectations, opening operatives' eyes to new possibilities: using the "event" feature on Facebook.[119] Having tapped into a cleavage—racial, ethnic, and religious tensions—in American politics, the IRA chose its target. One May afternoon in 2016, followers of the Facebook page "Heart of Texas" marched in front of the Islamic Da'wah Center in Houston, Texas, brandishing confederate flags and wearing shirts that read "White Power." The protesters had joined the vitriolic Facebook page and viewed ads promoting a rally to "Stop the Islamization of Texas." Waiting at the center were other Texans manipulated by the IRA: members of another Russian page, "United Muslims for America." Like a scene from a western, the two groups came face to face on Travis Street, waving banners and hurling insults at one another.[120] They were unaware that their confrontation had been arranged for roughly $200-worth of advertising space.[121] Though protests were small, the IRA had successfully exploited one of democracy's most sacred tools—the freedom to peacefully protest.

At this stage, the narratives of the disinformation campaign gain a foothold in the authentic population, blending and reinforcing their own sentiments. Attribution to threat actors becomes more challenging as more authentic voices amplify the campaign.

Conclusion

The RICHDATA framework provides an overview of the core aspects of modern disinformation campaigns. Threat actors use a variety of techniques across many stages to achieve their objectives. After surveying their target audience and building the necessary infrastructure to support a campaign, these operators begin to seed the information environment with tailored content often hijacking legitimate online forums. As they disseminate their message, they apply a variety of amplification techniques to broaden their reach and impact. If authentic users stand in their way, they may apply a variety of trolling techniques to marginalize dissenting voices while leveraging sometimes unwitting super-spreaders to propagate their message. At the pinnacle of a disinformation campaign, threat actors may operationalize their target audience and mobilize them to action.

The case studies highlight systemic challenges in the current information environment. The lines between foreign and domestic disinformation operations are blurring, particularly with the outsourcing of these operations to companies that provide influence as a service. The dual-use nature of platform features and applications built upon them makes vetting and due diligence critical. Lack of consensus over where to draw the line between harmful disinformation and protected speech requires platforms to make decisions that are politically unpopular.

Many of the tactics discussed in this report could become even more powerful with the application of ML techniques, the subject of our second paper in this series. In that paper, we discuss how advancements in AI/ML may augment the operators in each phase of a disinformation campaign and exacerbate these systemic challenges.

Authors

Katerina Sedova is a research fellow with the CyberAI Project at CSET.

Christine McNeill, Aurora Johnson, Aditi Joshi, and Ido Wulkan are former CSET student research analysts.

Acknowledgments

For their feedback and insights, we thank our CSET colleagues: John Bansemer, Ben Buchanan, Margarita Konaev, Andrew Imbrie, Micah Musser, Dakota Cary, and Andrew Lohn. We are grateful for the thoughtful reviews provided by Miriam Matthews of RAND Corporation and Alicia Wanless of Partnership for Countering Influence Operations at Carnegie Endowment for International Peace. Finally, we thank Hannah Stone, Melissa Deng, and Shelton Fitch for their editorial support. Any errors that remain are the fault of the authors alone.

Document Identifier: doi: 10.51593/2021CA005

Endnotes

[1] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia's Use of Social Media with Additional Views* (Washington, DC: U.S. Senate, 2019), https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

[2] Jonathan Swift, *The Examiner*, Number 15, November 9, 1710, https://books.google.com/books?id=KigTAAAAQAAJ&q=%22Truth+comes%22#v=snippet&q=%22Truth%20comes%22&f=false.

[3] Jon Bateman et al., "Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research" (Carnegie Endowment for International Peace, June 28, 2021), https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824.

[4] Office of the Director of National Intelligence, *Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution* (Washington, DC: Office of the DNI, January 6, 2017), https://www.dni.gov/files/documents/ICA_2017_01.pdf.

[5] "Dezinformatsiya" is a Russian word, defined in the Great Soviet Encyclopedia as the "dissemination of misleading or false information, used as a method of political propaganda aimed to mislead public opinion." Great Soviet Encyclopedia Online, https://bse.slovaronline.com/10240-DEZINFORMATSIYA. See also U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*.

[6] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*.

[7] Christopher Paul and Miriam Matthews, "The Russian 'Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It" (RAND Corporation, 2016), https://www.rand.org/pubs/perspectives/PE198.html.

[8] Paul and Matthews, "The Russian 'Firehose of Falsehood" Propaganda Model."

[9] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*. See also, Naima Green-Riley and Camille Stewart, "A Clapback to Russian Trolls," *The Root*, February 28, 2020, https://www.theroot.com/a-clapback-to-russian-trolls-1841932843.

[10] National Intelligence Council, *Foreign Threats to the 2020 US Federal Elections* (Washington, DC: Office of the Director of National Intelligence, March

10, 2021), https://www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf. See also: Center for an Informed Public, Digital Forensic Research Lab, Graphika, and Stanford Internet Observatory, “The Long Fuse: Misinformation and the 2020 Election,” *Stanford Digital Repository: Election Integrity Partnership*, v1.3.0, https://purl.stanford.edu/tr171zs0069.

[11] For China’s foray into international influence campaigns, see: Twitter Safety, “Disclosing Networks of State-linked Information Operations We’ve Removed,” *Twitter*, June 12, 2020, https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html; Jacob Wallis et al., “Retweeting Through the Great Firewall” (Australian Strategic Policy Institute, June 12, 2020), https://www.aspi.org.au/report/retweeting-through-great-firewall; Carly Miller et al., “Sockpuppets Spin COVID Yarns: An Analysis of PRC-Attributed June 2020 Twitter Takedown,” *Stanford Internet Observatory*, June 17, 2020, https://stanford.app.box.com/v/sio-twitter-prc-june-2020. For Iran’s evolution: Ben Nimmo et al., “Iran’s Broadcaster: Inauthentic Behavior” (Graphika, May 2020), https://public-assets.graphika.com/reports/graphika_report_irib_takedown.pdf; See also: Alice Revelli and Lee Foster, “‘Distinguished Impersonator’ Information Operation That Previously Impersonated U.S. Politicians and Journalists on Social Media Leverages Fabricated U.S. Liberal Personas to Promote Iranian Interests,” *FireEye*, February 12, 2020, https://www.fireeye.com/blog/threat-research/2020/02/information-operations-fabricated-personas-to-promote-iranian-interests.html. See also: Mona Elswah, Philip N. Howard, and Vidya Narayanan, “Iranian Digital Interference in the Arab World” (Oxford Internet Institute, April 3, 2019), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/04/Iran-Memo.pdf. See also: FireEye Intelligence, “Suspected Iranian Influence Operation Leverages Network of Inauthentic News Sites & Social Media Targeting Audience in the U.S., UK, Latin America, Middle East,” Mandiant, August 21, 2018, https://www.fireeye.com/blog/threat-research/2018/08/suspected-iranian-influence-operation.html.

[12] National Intelligence Council, *Foreign Threats to the 2020 US Federal Elections*. For attempts to influence the 2020 U.S. Presidential elections by Iran and China, see: John Ratcliffe, “Remarks at the Press Conference on Election Security,” Office of the Director of National Intelligence, October 21, 2020, https://www.dni.gov/index.php/newsroom/press-releases/press-releases-2020/item/2162-dni-john-ratcliffe-s-remarks-at-press-conference-on-election-security; William Evanina, “100 Days Until Election 2020,” Office of the Director of National Intelligence, July 24, 2020, https://www.dni.gov/index.php/newsroom/press-releases/item/2135-statement-by-ncsc-director-william-evanina-100-days-until-election-2020.

[13] Miriam Matthews, Katya Migacheva, and Ryan Andrew Brown, “Superspreaders of Malign and Subversive Information on COVID-19: Russian

and Chinese Efforts Targeting the United States" (RAND Corporation, 2021), https://www.rand.org/content/dam/rand/pubs/research_reports/RRA100/RRA112-11/RAND_RRA112-11.pdf. Ben Nimmo et al., "Return of the (Spamouflage) Dragon: Pro-Chinese Spam Network Tries Again" (Graphika, April 2020), https://public-assets.graphika.com/reports/Graphika_Report_Spamouflage_Returns.pdf. Ryan Serabian and Lee Foster, "Pro-PRC Influence Campaign Expands to Dozens of Social Media Platforms, Websites, and Forums in at Least Seven Languages, Attempted to Physically Mobilize Protesters in the U.S," *FireEye Threat Research*, September 8, 2021, https://www.fireeye.com/blog/threat-research/2021/09/pro-prc-influence-campaign-social-media-websites-forums.html.

[14] Samantha Bradshaw, Hannah Bailey, and Philip N. Howard, "Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation" (Computational Propaganda Project, Oxford Internet Institute, University of Oxford, January 2021), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/CyberTroop-Report-2020-v.2.pdf. Diego A. Martin, Jacob N. Shapiro, and Julia Ilhardt, "Trends in Online Influence Efforts" (Empirical Studies of Conflict Project, Princeton University, 2020), https://esoc.princeton.edu/publications/trends-online-influence-efforts.

[15] Daniel Kahneman, *Thinking, Fast and Slow* (New York, NY: Farrar, Straus and Giroux, 2011).

[16] For frameworks focused on threat actor motivations, see: Allan Kelly and Christopher Paul, "Decoding Crimea. Pinpointing the Influence Strategies of Modern Information Warfare" (NATO Strategic Communications Centre of Excellence, January 31, 2020), https://stratcomcoe.org/publications/decoding-crimea-pinpointing-the-influence-strategies-of-modern-information-warfare/64; Ben Nimmo, "Anatomy of an Information War: How Russia's Propaganda Machine Works and How to Counter It," *StopFake*, May 19, 2015, https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it; Claire Wardle, "Fake News: It's Complicated," *First Draft*, February 16, 2017, https://firstdraftnews.org/articles/fake-news-complicated/. For frameworks considering the role of media, see: Joan Donovan, "The Life Cycle of Media Manipulation," in "Verification Handbook: For Disinformation and Media Manipulation," https://datajournalism.com/read/handbook/verification-3/investigating-disinformation-and-media-manipulation/the-lifecycle-of-media-manipulation. For a conceptualization of Hack-Forge-Leak campaigns, see Thomas Rid, *Active Measures: The Secret History of Disinformation and Political Warfare* (New York, NY: Farrar, Straus and Giroux, April 2020).

For platform-focused defensive approaches, see: Camille François, "Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral

Deception to Guide Industry & Regulatory Responses" (Graphika and Berkman Klein Center for Internet & Society, Harvard University, September 20, 2019), https://www.ivir.nl/publicaties/download/ABC_Framework_2019_Sept_2019.pdf.

[17] "The Cyber Kill Chain," Lockheed Martin, https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html.

[18] For frameworks borrowed from cybersecurity and based on cyber Kill Chain and MITRE ATTACK, see Clint Watts, "Advanced Persistent Manipulators, Part Three: Social Media Kill Chain," *Alliance for Securing Democracy*, German Marshall Fund, July 22, 2019, https://securingdemocracy.gmfus.org/advanced-persistent-manipulators-part-three-social-media-kill-chain/; Peter M. et al., "Combatting Targeted Disinformation Campaigns: A Whole of Society Issue" (2019 Public-Private Analytic Exchange Program, October 2019), https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf. See also Bruce Schneier, "Toward an Information Operations Kill Chain," *Lawfare*, April 24, 2019, https://www.lawfareblog.com/toward-information-operations-kill-chain. For discussion of AMITT framework, see Sara-Jayne Terp, "Adversarial Misinformation and Influence Tactics and Techniques (AMITT)," GitHub, https://github.com/cogsec-collaborative/AMITT.

[19] Kate Starbird, "Disinformation's Spread: Bots, Trolls, and All of Us," *Nature*, July 24, 2019, https://www.nature.com/articles/d41586-019-02235-x.

[20] The U.S. Department of Justice indictment refers to this line of effort as the "translator project." See: *USA v. Internet Research Agency*: Indictment, Case 1:18-cr-00032-DLF, U.S. District of Columbia, February 16, 2018, https://www.justice.gov/file/1035477/download. See also: Rid, *Active Measures*.

[21] Alya Shandra, "What Surkov's Hacked Emails Tell about Russia's Hybrid War against Ukraine," *Euromaidan Press*, November 12, 2019, http://euromaidanpress.com/2019/11/12/what-surkovs-hacked-emails-tell-about-russias-hybrid-war-against-ukraine/, Alya Shandra, "A Guide to Russian Propaganda, Part 5: Reflexive Control," *Euromaidan Press*, March 26, 2020, http://euromaidanpress.com/2020/03/26/a-guide-to-russian-propaganda-part-5-reflexive-control/. Alya Shandra and Robert Seely, "The Surkov Leaks: The Inner Workings of Russia's Hybrid War in Ukraine" (Royal United Services Institute for Defence and Security Studies, July 2019), https://static.rusi.org/201907_op_surkov_leaks_web_final.pdf.

[22] Serhii Plokhy, "Navigating the Geopolitical Battlefield of Ukrainian History," *Atlantic Council*, September 9, 2021, https://www.atlanticcouncil.org/blogs/ukrainealert/navigating-the-geopolitical-battlefield-of-ukrainian-history/. Serhii Plokhy, *The Frontline: Essays on*

Ukraine's Past and Present (Cambridge, MA: Harvard Series in Ukrainian Studies, September 2021).

[23] Note: Throughout Euromaidan protests, Russian-speaking Ukrainians in the east and south were bombarded with messages painting protestors supporting pro-European economic policy as "Banderovtsy," a Soviet term for Ukrainian nationalists that in a distorted narrative implied Nazi collaboration. Anne Applebaum and Peter Pomerantsev, et al., "From 'Memory Wars' to a Common Future: Overcoming Polarisation in Ukraine" (Arena Project, The London School of Economics and Political Science, July 2020), https://www.lse.ac.uk/iga/assets/documents/From-Memory-Wars.pdf.

[24] Aric Toler, "Inside the Kremlin Troll Army Machine: Templates, Guidelines, and Paid Posts," *Global Voices*, March 14, 2015, https://globalvoices.org/2015/03/14/russia-kremlin-troll-army-examples/.

[25] Paul and Matthews, "The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options to Counter It"; Starbird, "Disinformation's Spread."

[26] Joan Donovan and Brian Friedberg, "Source Hacking: Media Manipulation in Practice," *Data & Society*, September 4, 2019, https://datasociety.net/library/source-hacking-media-manipulation-in-practice/.

[27] Matthews et al., "Superspreaders of Malign and Subversive Information on COVID-19."

[28] Edward Wong, Matthew Rosenberg, and Julian E. Barnes, "Chinese Agents Helped Spread Messages That Sowed Virus Panic in US, Officials Say," *The New York Times*, April 22, 2020, https://www.nytimes.com/2020/04/22/us/politics/coronavirus-china-disinformation.html.

[29] Matthews et al., "Superspreaders of Malign and Subversive Information on COVID-19"; John Gregory and Kendrick McDonald, "Trail of Deceit: The Most Popular COVID-19 Myths and How They Emerged," *NewsGuard*, June 2020, https://www.newsguardtech.com/special-reports/covid-19-myths/; Nectar Gan and Steve George, "China doubles down on baseless 'US origins' Covid conspiracy as Delta outbreak worsens," *CNN*, August 6, 2021, https://www.cnn.com/2021/08/06/china/china-covid-origin-mic-intl-hnk/index.html.

[30] "July 2021 Coordinated Inauthentic Behavior Report," *Facebook*, July 2021, https://about.fb.com/news/2021/08/july-2021-coordinated-inauthentic-behavior-report/.

[31] Karissa Bell, “Facebook caught a marketing firm paying influencers to criticize COVID-19 vaccines,” *Engadget*, August 10, 2021, https://www.engadget.com/facebook-caught-a-marketing-firm-paying-influencers-to-criticize-covid-19-vaccines-181030579.htm; Liz Alderman, “Influencers Say They Were Urged to Criticize Pfizer Vaccine,” *The New York Times*, May 26, 2021, https://www.nytimes.com/2021/05/26/business/pfizer-vaccine-disinformation-influeners.html; John Gregory, “The Top COVID-19 Vaccine Myths Spreading Online,” *NewsGuard*, September 13, 2021, https://www.newsguardtech.com/special-reports/special-report-top-covid-19-vaccine-myths/.

[32] Charlie Haynes and Flora Carmichael, “The YouTubers Who Blew the Whistle on an Anti-Vax Plot,” *BBC News*, July 25, 2021, https://www.bbc.com/news/blogs-trending-57928647.

[33] Manveen Rana and Sean O’Neill, “Coronavirus: fake news factories churning out lies over ‘monkey’ vaccine,” *The Times*, October 16, 2020, https://www.thetimes.co.uk/article/fake-news-factories-churning-out-lies-over-monkey-vaccine-qhhmxt2g5; Michael R. Gordon and Dustin Volz, “Russian Disinformation Campaign Aims to Undermine Confidence in Pfizer, Other Covid-19 Vaccines, U.S. Officials Say,” *The Wall Street Journal*, March 7, 2021, https://www.wsj.com/articles/russian-disinformation-campaign-aims-to-undermine-confidence-in-pfizer-other-covid-19-vaccines-u-s-officials-say-11615129200.

[34] Alexandra Tyan, “Russia pushed vaccine conspiracies. Now it’s backfiring,” *Coda Story*, June 25, 2021, https://www.codastory.com/newsletters/infodemic-june-25/.

[35] *USA v. Internet Research Agency*: Indictment.

[36] Note: this figure is based on publicly available research and may have changed by the platforms since then, but the confirmation of any such change is not publicly observable. Athanasios Andreou et al., “Measuring the Facebook Advertising Ecosystem,” *NDSS 2019 - Proceedings of the Network and Distributed System Security Symposium* (February 2019): 1-15, https://hal.archives-ouvertes.fr/hal-01959145/document.

[37] Note: authors accessed the “Audience Insights” feature of Facebook for Business Suite with their personal accounts that were in the past administrators of a group, but not of a business page. They were shown a Potential Audience Size for their potential ad is 245,200,000 people on Facebook and Instagram. https://business.facebook.com/latest/insights/people?asset_id=370279763339235&nav_ref=audience_insights.

[38] "How Twitter Ads Work," Ads Help Center, Twitter, https://business.twitter.com/en/help/troubleshooting/how-twitter-ads-work.html.

[39] "New Steps to Protect the US Elections," *Facebook Newsroom*, September 3, 2020, https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/.

[40] Ben Nimmo et al., "Secondary Infektion," *Graphika*, https://secondaryinfektion.org/report/secondary-infektion-at-a-glance/.

[41] Kate Starbird, "The Surprising Nuance Behind the Russian Troll Strategy," *Medium*, October 20, 2018, https://medium.com/s/story/the-trolls-within-how-russian-information-operations-infiltrated-online-communities-691fb969b9e4.

[42] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*," 38-54.

[43] Kate Starbird, Ahmer Arif, and Tom Wilson, "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations," *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (November 2019): 1-26, https://doi.org/10.1145/3359229.

[44] Starbird et al., "Disinformation as Collaborative Work."

[45] Ben Nimmo et al., "Operation Red Card: An Indian PR Firm Created Inauthentic Accounts and Used Coordinated Behavior to Post Anti-Saudi, Anti-Emirati, Pro-Qatar, and Football-related Content," *Graphika*, March 2020, https://public-assets.graphika.com/reports/graphika_report_operation_redcard.pdf; Singularex, "The Black Market for Social Media Manipulation" (NATO Strategic Communications Centre of Excellence, January 19, 2019), https://www.stratcomcoe.org/black-market-social-media-manipulation.

[46] Michael Schwirtz and Sheera Frenkel, "In Ukraine, Russia Tests a New Facebook Tactic in Election Tampering," *The New York Times*, March 28, 2019, https://www.nytimes.com/2019/03/29/world/europe/ukraine-russia-election-tampering-propaganda.html; Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior From Russia," *Facebook Newsroom*, March 12, 2020, https://about.fb.com/news/2020/03/removing-coordinated-inauthentic-behavior-from-russia/; Davey Alba, "How Russia's Troll Farm Is Changing Tactics Before the Fall Election," *The New York Times*, March 28, 2020, https://www.nytimes.com/2020/03/29/technology/russia-troll-farm-election.html.

[47] Samuel C. Woolley and Philip Howard, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (Oxford, UK:

Oxford University Press, 2018), as cited in Programme on Democracy & Technology, “What is computational propaganda?,” Oxford Internet Institute, https://navigator.oii.ox.ac.uk/what-is-comprop/.

[48] Philip N. Howard, Samuel Woolley, and Ryan Calo, “Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration,” *Journal of Information Technology & Politics* 15, no. 2 (2018): 81-93, https://doi.org/10.1080/19331681.2018.1448735.

[49] Emilio Ferrara et al., “Characterizing Social Media Manipulation in the 2020 U.S. Presidential Election,” *First Monday* 25, no. 11 (November 2020), https://doi.org/10.5210/fm.v25i11.11431.

[50] Stefano Cresci, “A decade of social bot detection,” *Communications of the ACM* 63, no. 10 (October 2020): 72-83, https://doi.org/10.1145/3409116.

[51] Donara Barojan (on behalf of DFRLab), “#TrollTracker: Bots, Botnets, and Trolls,” *Medium*, October 8, 2018, https://medium.com/dfrlab/trolltracker-bots-botnets-and-trolls-31d2bdbf4c13.

[52] Polina Rusyaeva and Andrei Zaharov, “How ‘Troll Factory’ worked the U.S. Elections,” *RBK Magazine*, October 17, 2017, https://web.archive.org/web/20210303095306/https://www.rbc.ru/magazine/2017/11/59e0c17d9a79470e05a9e6c1; See also Ben Nimmo et al., “Spamouflage Goes to America” (Graphika, August 2020), https://graphika.com/reports/spamouflage-dragon-goes-to-america/.

[53] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*, 38.

[54] Gleicher, “Removing Coordinated Inauthentic Behavior from Russia”; See also Nika Aleksejeva et al. (on behalf of DFRLab), “Facebook’s Sputnik Takedown — In Depth,” *Medium*, January 17, 2019, https://medium.com/dfrlab/facebooks-sputnik-takedown-in-depth-f417bed5b2f8.

[55] Nika Aleksejeva et al. (on behalf of DFRLab), “Facebook’s Sputnik Takedown – – Top Takeaways,” *Medium*, January 17, 2019, https://medium.com/dfrlab/facebooks-sputnik-takedown-top-takeaways-dbc22f7e9540.

[56] Mark Zuckerberg, “A Privacy Focused Vision for Social Networking,” *Facebook Newsroom*, March 6, 2019, https://about.fb.com/news/2019/03/vision-for-social-networking/.

[57] Nina Jankowicz and Cindy Otis, "Facebook Groups Are Destroying America," *WIRED*, June 17, 2020, https://www.wired.com/story/facebook-groups-are-destroying-america/.

[58] Tom Alison, "Changes to Keep Facebook Groups Safe," *Facebook Newsroom*, March 17, 2021, https://about.fb.com/news/2021/03/changes-to-keep-facebook-groups-safe/.

[59] Mia Bloom, "Telegram and Online Addiction to Terrorist Propaganda," *Minerva Research Initiative*, May 29, 2919, https://minerva.defense.gov/Owl-In-the-Olive-Tree/Owl_View/Article/1859857/telegram-and-online-addiction-to-terrorist-propaganda/. See also Michael Edison Hayden, "Far-Right Kyle Rittenhouse Propaganda 'Not Factually Based,' Says Kenosha Militia Participant," *Southern Poverty Law Center*, September 15, 2020, https://www.splcenter.org/hatewatch/2020/09/15/far-right-kyle-rittenhouse-propaganda-not-factually-based-says-kenosha-militia-participant.

[60] Note: One inauthentic network built audiences by living out manufactured dramas on Facebook. A longstanding sprawling network of 329 fake accounts took on a "telenovela"-style life of its own over a course of seven years. The fake account of "Alice Bergmann," a native of Chemnitz, Germany, whose sister was "brutally murdered in a Berlin park" by a Muslim immigrant, never existed. Nor did her concerned brother "David," or her caring cousin "Helena," who sympathizes with her anger and frustration toward Islam. The entire family of accounts, though realistically written into existence, were fakes. The complexity and believability of their backstory enable this type of a network to build audiences and stand ready to be repurposed by the network's operators to amplify campaign messages. For details, see: Maik Baumgärtner and Roman Höfner, "How To Fake Friends and Influence People," *Speigel International*, January 24, 2020, https://www.spiegel.de/international/germany/facebook-how-to-fake-friends-and-influence-people-a-4605cea1-6b49-4c26-b5b7-278caef29752. See also: Nika Aleksejeva and Zarine Kharazian (on behalf of DFRLab), "Top Takes: A Facebook Drama In Three Acts," *Medium*, January 24, 2020, https://medium.com/dfrlab/top-takes-a-facebook-drama-in-three-acts-a275e037c8be.

[61] Sean Gallagher, "DNC 'Lone Hacker' Guccifer 2.0 Pegged as Russian Spy After Opsec Fail," *ArsTechnica*, March 23, 2018, https://arstechnica.com/tech-policy/2018/03/dnc-lone-hacker-guccifer-2-0-pegged-as-russian-spy-after-opsec-fail/.

[62] Greg Myre, "A 'Perception Hack': When Public Reaction Exceeds The Actual Hack," *NPR*, November 1, 2020, https://www.npr.org/2020/11/01/929101685/a-perception-hack-when-public-reaction-exceeds-the-actual-hack; Gleicher, "Removing Coordinated Inauthentic Behavior," *Facebook Newsroom*, October 27, 2020,

https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-mexico-iran-myanmar/.

[63] Michael Ashley, "Sick of The Attention Economy? It's Time to Rebel," *Forbes*, November 24, 2019, https://www.forbes.com/sites/cognitiveworld/2019/11/24/sick-of-the-attention-economy-its-time-to-rebel/; See also Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (New York, NY: Knopf Doubleday Publishing Group, 2016).

[64] Anton Troianovski, "A Former Russian Troll Speaks: 'It was like being in Orwell's world,'" *The Washington Post*, February 17, 2018, https://www.washingtonpost.com/news/worldviews/wp/2018/02/17/aformer-russian-troll-speaks-it-was-like-being-in-orwells-world/; Adrian Chen, "The Agency," *The New York Times*, June 2, 2015, https://www.nytimes.com/2015/06/07/ magazine/the-agency.html; Andrei Soshnikov, "The Capital of Political Trolling," *Moi Rayon*, March 11, 2015, https://mr-7.ru/ articles/112478/; Max Seddon, "Documents Show How Russia's Troll Army Hit America," *BuzzFeed News*, June 2, 2014, https://www.buzzfeednews.com/article/ maxseddon/documents-show-how-russias-troll-army-hit-america.

[65] Nora Bitenice et al., "Digital Hydra: Security Implications of False Information Online" (NATO Strategic Communications Centre of Excellence, 2017), https://www.stratcomcoe.org/digital-hydra-security-implications-false-information-online; Julie Fedor and Rolf Fredheim, "'We need more clips about Putin, and lots of them': Russia's State-Commissioned Online Visual Culture," *Nationalities Papers* 45, no. 2 (2017): 161-181.

[66] Toler, "Inside the Kremlin Troll Army Machine."

[67] Rusyaeva and Zaharov, "How 'Troll Factory' worked the U.S. Elections"; U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*, 30.

[68] Rid, *Active Measures*, 407.

[69] Ryan Andrew Brown et al., "Rapid Analysis of Foreign Malign Information on COVID-19 in the Indo-Pacific: A Proof-of-Concept Study" (RAND Corporation, 2021), https://www.rand.org/pubs/research_reports/RRA1345-1.html.

[70] Douwe Kiela, Hamed Firooz, and Aravind Mohan, "Hateful Memes Challenge and dataset for research on harmful multimodal content," *Facebook AI*, May 12, 2020, https://ai.facebook.com/blog/hateful-memes-challenge-and-data-set/.

[71] Joan Donovan, "How Memes Got Weaponized: A Short History," *MIT Technology Review*, October 24, 2019,

https://www.technologyreview.com/2019/10/24/132228/political-war-memes-disinformation/.

[72] Abby Ohlheiser, "What Was Fake on the Internet This Election: #DraftOurDaughters, Trump's Tax Returns," *The Washington Post*, October 31, 2016, hhttps://www.washingtonpost.com/news/the-intersect/wp/2016/10/31/what-was-fake-on-the-internet-this-election-draftourdaughters-trumps-tax-returns/.

[73] Kim LaCapria, "Hillary Clinton and #DraftOutDaughters," *Snopes*, October 28, 2016, https://www.snopes.com/fact-check/hillary-clinton-and-draftourdaughters/.

[74] Donovan, "How Memes Got Weaponized."

[75] DARPA commissioned a study on how to integrate memetics into modern information warfare. While controversial, calls to integrate "memetic warfare" into the U.S. and NATO allied military information operations and strategic communications doctrine continue to percolate, in part due to the perception that the asymmetric effectiveness of these tools at the hands of terrorist propagandists and foreign influence operators must be countered in kind. For deeper examination, see: Brian J. Hancock, "Memetic Warfare: The Future of War," *Military Intelligence* 36, no. 2 (April-June 2010), https://fas.org/irp/agency/army/mipb/2010_02.pdf; Jeff Giease, "Hacking Hearts and Minds: How Memetic Warfare is Transforming Cyberwar," *OPEN Publications* 1, no. 6 (June 2017), https://www.act.nato.int/images/stories/media/doclibrary/open201706-memetic2.pdf.

[76] Jessikka Aro, a Finnish journalist who exposed Russian influence operations in Finland, found herself turned into a meme. Applying similar techniques, Johan Backman, a "Finnish propagandist" created memes from her pictures to attack her investigations into Russian disinformation. Following the publication of her 2014 investigations into Russian trolls active in Finland, Backman and an ensuing army of trolls harassed Aro with memes. An anonymous email accounts sent memes discrediting Aro's work to Finnish journalists, ministers, and even the president of Finland. Memes spread across Twitter, where they picked up traction and perpetuated the campaign against Aro. For details, see: Jessikka Aro, "How Pro-Russian Trolls Tried to Destroy Me," *BBC News*, October 6, 2017, https://www.bbc.com/news/blogs-trending-41499789.

[77] Alyssa Kann and Kanishk Karan (on behalf of DFRLab), "Inauthentic Anti-Rohingya Facebook Assets in Myanmar Removed," *Medium*, May 5, 2020, https://medium.com/dfrlab/inauthentic-anti-rohingya-facebook-assets-in-myanmar-removed-39eb7e069d9.

[78] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 5*; Jean-Baptiste Jeangène Vilmer, "The "Macron Leaks" Operation: A Post-Mortem," *Atlantic Council*, June 2019, https://www.atlanticcouncil.org/wp-content/uploads/2019/06/The_Macron_Leaks_Operation-A_Post-Mortem.pdf.

[79] Lee Foster et al., "'Ghostwriter' Influence Campaign: Unknown Actors Leverage Website Compromises and Fabricated Content to Push Narratives Aligned With Russian Security Interests" (Mandiant, July 29, 2020), https://www.fireeye.com/content/dam/fireeye-www/blog/pdfs/Ghostwriter-Influence-Campaign.pdf; Lee Foster et al., "Ghostwriter Update: Cyber Espionage Group UNC1151 Likely Conducts Ghostwriter Influence Activity," *Mandiant Threat Research*, April 28, 2021, https://www.fireeye.com/blog/threat-research/2021/04/espionage-group-unc1151-likely-conducts-ghostwriter-influence-activity.html.

[80] Tim Starks, "EU takes aim at Russia over 'Ghostwriter' hacking campaign against politicians, government officials," *CyberScoop*, September 24, 2021, https://www.cyberscoop.com/eu-high-representative-ghostwriter-russia-warning-german-elections/; High Representative, Council of the European Union, "Declaration by the High Representative on behalf of the European Union on respect for the EU's democratic processes," Council of the European Union, September 24, 2021, https://www.consilium.europa.eu/en/press/press-releases/2021/09/24/declaration-by-the-high-representative-on-behalf-of-the-european-union-on-respect-for-the-eu-s-democratic-processes/; Gabriella Roncone, Alden Wahlstrom, Alice Revelli, David Mainor, Sam Riddell, Ben Read, "UNC1151 Assessed with High Confidence to have Links to Belarus, Ghostwriter Campaign Aligned with Belarusian Government Interests," *Mandiant Threat Research*, November 16, 2021, https://www.mandiant.com/resources/unc1151-linked-to-belarus-government

[81] Nimmo et al., "Secondary Infektion."

[82] Nika Aleksejeva et al., "Operation 'Secondary Infektion': a Suspected Russian Intelligence Operation Targeting Europe and the United States," *DFRLab*, Atlantic Council, August 2019, https://www.atlanticcouncil.org/wp-content/uploads/2019/08/Operation-Secondary-Infektion_English.pdf.

[83] Nathaniel Gleicher, "Removing More Coordinated Inauthentic Behavior from Russia," *Facebook Newsroom*, May 6, 2019, https://about.fb.com/news/2019/05/more-cib-from-russia/.

[84] Kate Conger and Sheera Frenkel, "How FireEye Helped Facebook Spot a Disinformation Campaign," *The New York Times*, August 23, 2018, https://www.nytimes.com/2018/08/23/technology/fireeye-facebook-disinformation.html.

[85] Global Engagement Center, *GEC Special Report: Pillars of Russia's Disinformation and Propaganda Ecosystem* (Washington, DC: U.S. Department of State, August 2020), https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf.

[86] For discussion of effectiveness of repetition and first impression advantage on influence, see Paul and Matthews, "The Russian 'Firehose of Falsehood" Propaganda Model"; Gordon Pennycook et al., "Prior exposure increases perceived accuracy of fake news," *Journal of Experimental Psychology: General* 147, no. 12 (2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6279465/; Lynn Hasher et. al., "Frequency and the conference of referential validity," *Journal of Verbal Learning and Verbal Behavior* 16, no. 1 (1977), https://www.sciencedirect.com/science/article/abs/pii/S0022537177800121.

[87] Paul and Matthews, The Russian "'Firehose of Falsehood' Propaganda Model."

[88] Twitter Safety, "Information operations directed at Hong Kong," *Twitter*, August 19, 2019, https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html; Twitter Safety, "Disclosing networks of state-linked information operations we've removed," *Twitter*, June 12, 2020, https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html.

[89] Miller et al., "Sockpuppets Spin COVID Yarns."

[90] Miller et al., "Sockpuppets Spin COVID Yarns."

[91] Nimmo et al., "Return of the (Spamouflage) Dragon"; Nimmo et al., "Spamouflage Dragon Goes to America: Pro-Chinese Inauthentic Network Debuts English-Language Videos" (Graphika, August 2020), https://public-assets.graphika.com/reports/graphika_report_spamouflage_dragon_goes_to_america.pdf; Renee DiResta et al., "Telling China's Story: The Chinese Communist Party's Campaign to Shape Global Narratives," *Stanford Internet Observatory*, July 20, 2020, https://cyber.fsi.stanford.edu/io/news/new-whitepaper-telling-chinas-story.

[92] Serabian and Foster, "Pro-PRC Influence Campaign Expands to Dozens of Social Media Platforms, Websites, and Forums."

[93] Howard et al., "Algorithms, bots, and political communication in the US 2016 election."

[94] Howard et al., "Algorithms, bots, and political communication in the US 2016 election."

[95] Kevin Roose, “QAnon Followers Are Hijacking the #SaveTheChildren Movement,” *The New York Times*, August 12, 2020, https://www.nytimes.com/2020/08/12/technology/qanon-save-the-children-trafficking.html; Kevin Roose, “How ‘Save the Children’ Is Keeping QAnon Alive,” *The New York Times*, September 28, 2020, https://www.nytimes.com/2020/09/28/technology/save-the-children-qanon.html.

[96] Developer Platform, “Build for Business: Use Twitter’s Powerful APIs to Help Your Business Listen, Act, and Discover,” *Twitter*, https://developer.twitter.com/en/use-cases/listen-and-analyze.

[97] Howard et al., “Algorithms, bots, and political communication in the US 2016 election.”

[98] Developer Platform, “More about restricted uses of the Twitter APIs,” *Twitter*, https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases; Yoel Roth, “Automation and the use of multiple accounts,” *Twitter*, February 21, 2018, https://blog.twitter.com/developer/en_us/topics/tips/2018/automation-and-the-use-of-multiple-accounts; Developer Platform, “Developer Agreement and Policy,” *Twitter*, March 10, 2020, https://developer.twitter.com/en/developer-terms/agreement-and-policy.

[99] Help Center, “Twitter's automation development rules,” *Twitter*, November 3, 2017, https://help.twitter.com/en/rules-and-policies/twitter-automation.

[100] Bradshaw et al., “Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation.”

[101] Nathaniel Gleicher et al., “Threat Report: The State of Influence Operations 2017-2020,” *Facebook*, May 2021, https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf.

[102] Nathaniel Gleicher, “Removing Coordinated Inauthentic Behavior,” *Facebook*, July 8, 2020, https://about.fb.com/news/2020/07/removing-political-coordinated-inauthentic-behavior/.

[103] Nimmo et al, “Operation Red Card.”

[104] Nick Monaco, Melanie Smith, and Amy Studdart, “Detecting Digital Fingerprints: Tracing Chinese Disinformation in Taiwan” (DFRLab, Graphika, and The International Republican Institute, August 2020), https://graphika.com/reports/detecting-digital-fingerprints-tracing-chinese-disinformation-in-taiwan/.

[105] Ben Nimmo, “UK Trade Leaks and Secondary Infektion: New Findings and Insights from a Known Russian Operation” (*Graphika*, December 2019),

https://public-assets.graphika.com/reports/graphika_report_uk_trade_leaks_&_secondary_infektion.pdf

[106] Nimmo, "UK Trade Leaks and Secondary Infektion."

[107] Emma L. Briant and Alicia Wanless, "A digital ménage à trois: Strategic leaks, propaganda and journalism," in *Countering Online Propaganda and Extremism* (UK: Routledge, 2018), https://www.taylorfrancis.com/chapters/edit/10.4324/9781351264082-4/digital-m%C3%A9nage-trois-emma-briant-alicia-wanless.

[108] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 5*.

[109] Kirill Meleshevich and Bret Schafer, "Online Information Laundering: The Role of Social Media," *Alliance for Security Democracy*, German Marshall Fund, January 9, 2018, https://securingdemocracy.gmfus.org/online-information-laundering-the-role-of-social-media/.

[110] Melanie Smith, "Interpreting Social Qs: Implications of the Evolution of QAnon" (Graphika, August 2020), https://public-assets.graphika.com/reports/graphika_report_interpreting_social_qs.pdf.

[111] David A. Broniatowski et al., "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate," *American Journal of Public Health* 108, no. 10 (2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6137759/; Smith, "Interpreting Social Qs."

[112] Jan van Dijk and Robert van der Noordaa, "An Army of Trolls Pushing the Russian Narrative on MH17: Who Are They?," *Ukraine World International*, July 17, 2020, https://ukraineworld.org/articles/infowatch/MH17-trolls.

[113] Eliot Higgins, "MH17: The Open Source Evidence," *Bellingcat*, October 8, 2015, https://www.bellingcat.com/news/uk-and-europe/2015/10/08/mh17-the-open-source-evidence/.

[114] Seddon, "Documents Show How Russia's Troll Army Hit America"; Chris Elliott, "The Readers' Editor on… Pro-Russian Trolling Below the Line on Ukraine Stories," *The Guardian*, May 4, 2014, https://www.theguardian.com/commentisfree/2014/may/04/pro-russia-trolls-ukraine-guardian-online.

[115] Hal Berghel and Daniel Berleant, “The Online Trolling Ecosystem,” *Computer*, August 2018, http://www.berghel.net/col-edit/aftershock/aug-18/aftershock_8-18.pdf.

[116] Troianovski, “A Former Russian Troll Speaks.”

[117] Troianovski, “A Former Russian Troll Speaks.”

[118] “What about ‘whataboutism’? If everyone is guilty of something, is no one guilty of anything?,” Merriam Webster, https://www.merriam-webster.com/words-at-play/whataboutism-origin-meaning.

[119] Rid, *Active Measures*, 401; U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*, 46.

[120] Claire Allbright, “A Russian Facebook page organized a protest in Texas. A different Russian page launched the counterprotest,” *The Texas Tribune*, November 1, 2017, https://www.texastribune.org/2017/11/01/russian-facebook-page-organized-protest-texas-different-russian-page-l/.

[121] U.S. Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2*, 47.