THE UNIVERSITY *of* EDINBURGH

Postgraduate Teaching Office
Social and Political Science
Dissertation Cover Sheet

Please enter your Examination Number and Dissertation Title into the following grid and insert this form on to the front of your dissertation before uploading your file.

Exam Number	**B265912**
Dissertation Title	**The Ctrl+Alt+Defeat Playbook: Platform-Intelligence Collaboration Against Russian Disinformation in the Russia-Ukraine War**
Word Count	**14,925**

Important information about how your dissertation may be retained and shared, please read.

The School may retain a copy of your dissertation and make it available to other students to assist them in completing their own dissertation. Where your work is made available, it will be anonymized add added to the subject area dissertation/ project library.

The freedom of information (Scotland) Act 2002 requires the University to make available to any enquirer any information held by the university, unless one of the legislation's narrowly defined exemptions apply.

If, with your consent, a copy of your dissertation is retained beyond the period of assessment, information contained within may be made available to any enquirer. If your dissertation or project contains confidential information, i.e data that has been provided to you by external bodies under a Non-Disclosure Agreement or Data Processing Agreement, where confidentiality and integrity must be preserved; or personal or sensitive data belonging to research participants (information relating to natural persons who can be identified directly or indirectly from the information contained within your dissertation), you should not consent to your dissertation/project being retained.

If you do not consent to your work being retained we will destroy the work as per Taught Assessment Regulation 49.

Please read the full details of our retention of course assessments policy on our webpage and edit the below statement as applicable.

I consent for the School to retain a copy of my dissertation or project and make it available to other students to assist them in completing their own dissertation: *Yes*

If in future you wish to withdraw consent, please email pgtaught.sps@ed.ac.uk

THE UNIVERSITY *of* EDINBURGH
School of Social
& Political Science

The Ctrl+Alt+Defeat Playbook: Platform-Intelligence Collaboration Against Russian Disinformation In The Russia-Ukraine War

By: Malek Shedid

A Dissertation
Submitted to the School of Social and Political Science
University of Edinburgh
In Fulfillment of the Requirements
For the Degree of Masters of Science in International Relations
August 2025

Abstract

On February 24, 2022, Russia's full-scale invasion of Ukraine triggered an unprecedented surge in state-aligned disinformation campaigns targeting Euro-Atlantic audiences. This dissertation investigates how NATO-affiliated intelligence agencies and the social media platforms Meta and Twitter collaborated to counter these operations during the 2022-2024 escalation. A central and original contribution is the development of a five-level collaboration typology, ranging from symbolic alignment to integrated response, to systematically assess the depth and institutionalisation of public-private security partnerships. Through comparative case studies of Twitter's handling of the Ghostwriter campaign and Meta's disruption of Operation Doppelgänger, this research applies securitization theory and networked governance to process-traced, open-source evidence, including transparency reports, NATO documentation, and investigative journalism. Findings show Meta achieved Level 4 (Integrated Response), characterized by synchronized takedowns, co-produced attribution, and embedded institutional linkages, while Twitter's Level 3 (Strategic Partnership) reflected informal, reactive coordination amid governance instability. The study advances understanding of digital governance in hybrid warfare, offering a transferable framework for evaluating intelligence-technology collaboration and interrogating the democratic tensions that emerge when counter-disinformation efforts blur the boundaries between state and corporate authority.

Keywords: NATO, disinformation, hybrid warfare, social media, platform governance, securitization, networked governance, public-private partnerships.

Acknowledgements

I would like to firstly say Alhamdulillah. I want to thank Allah SWT for allowing my life to shift in a way I wouldn't have imagined possible. It has been a privilege to further my education and experience all the beauty that this life and the amazing nation of Scotland have to offer.

To my mother, Elham. I hope I have made you proud. I will never be able to repay you for the number of times you had to leave work for my disciplinary issues as a child. I could not have asked for a better parent to raise me. Sometimes I felt defeated being abroad alone for this long, but I remind myself that you did it with three children and barely knew the language. While I can only offer this moment, I promise I will show you the amazing places I have been so lucky to witness. I love you.

My oldest sister, Eman. The only person who knows exactly what is going on in my head, regardless of how well I think my poker face is. Thank you for pushing me to pursue this opportunity and for answering every phone call when I felt isolated here. Also for listening to my post-training rants when football wasn't going the way I would have liked. If the career market ever allows me to finally find employment, I will finally help with buying Daisy some litter. I love you as well.

My second-oldest sister, Omnia. The only person who knows exactly what button to push to irritate me. Thank you for challenging me every time we argue. I wouldn't have even known what to study when I started this journey if it weren't for you. I will never admit this, but I do look up to you, and I appreciate that you helped me polish my professional skills. I will forever annoy you at home, but I will always have your back in public. I guess I love you too.

Dr. Adam Chalmers, thank you from the bottom of my heart for being there and guiding me through my first dissertation process. It has been a pleasure and my honor working with you, and I wish you the best in all your future endeavors.

Dr. T.C. Lloyd, thank you for being one of the most impactful people throughout my entire university process. When I transferred into UMD, you told me to make sure I took a big place and make it small. I have carried that mindset since, and it has allowed me to connect with people from every corner of the globe.

Eban, Marci, and both of their families. I appreciate you two for being the closest thing to a brother for me. Eban, I will never forget the endless help you and your family provided me when I was younger, a place to sleep, dinners, and rides to school. Marci, I think the amount of travel stories we have could equate to almost 5 dissertations. Thank you both from the bottom of my heart.

Lastly, all my friends and teammates have made this city as magical as J.K. Rowling described it. Thank you all.

Table of Contents

Table of Contents

Introduction 6

Research Objective 7

Framework 8

Significance of This Study 8

Outline of the Paper 9

Chapter One: Literature Review and Theoretical Framework 11

Literature Review 11

Disinformation as Hybrid Warfare 12

Platform Governance and Crisis Response 13

Theoretical Framework: Networked Governance and Securitization Theory 17

Networked Governance 17

Securitization Theory 19

Complementarity of the Frameworks 20

Application to Case Studies and Evaluating Degrees of Collaboration 21

Degrees of Collaboration 22

Factors Influencing Collaboration 23

Analytical Contribution 23

Chapter Two: Methodology 25

Research Design 25

Case Studies 26

Data Sources 28

Limitations 29

Chapter Three: Twitter Takedown 30

Introduction 30

Ghostwriter 31

Platform Response 33

Intelligence-Platform Coordination 35

Thematic Analysis of Collaboration 37

Collaboration Typology Assessment 39

Factors Influencing Collaboration 40

Summary 42
Chapter Four: Operation Doppelgänger 44
Introduction 44
The Doppelgänger Campaign 45
Platform Response 47
Intelligence-Platform Coordination 50
Thematic Analysis of Collaboration 52
Collaboration Typology Assessment 54
Factors Influencing Collaboration 56
Summary 57
Chapter Five: Comparative Analysis and Discussion 59
Comparative Evaluation of Collaboration Effectiveness 60
Tensions and Trade-Offs 62
Theoretical Insights Compared 64
Key Findings 65
Broader Implications 66
Recommendations and Future Research 68
Summary 68
Conclusion 70
Bibliography 74

List of Acronyms

ASEAN	Association of Southeast Asian Nations
CCDCOE	Cooperative Cyber Defense Centre of Excellence
CIB	Coordinated Inauthentic Behavior
DSA	Digital Services Act
EEAS	European External Action Service
EU	European Union
GCHQ	Government Communications Headquarters (United Kingdom)
GEC	Global Engagement Center (United States of America)
GRU	Main Directorate of the General Staff of the Armed Forces of the Russian Federation
NATO	North Atlantic Treaty Organization
OSINT	Open-Source Intelligence
PPP	Public-Private Partnerships
STRATCOM COE	Strategic Communications Centre of Excellence
UK	United Kingdom
US	United States (of America)

Introduction

"Never attempt to win by force what can be won by deception."

-Niccolò Machiavelli, The Prince

Machiavelli's insight remains timely: deception often succeeds where force cannot. In modern warfare, truth is both a casualty and a target. From ancient empires to digital regimes, propaganda has shaped perceptions and weakened resistance. Today, algorithmic bias and decentralized networks enable disinformation to spread faster and further than ever before, outpacing traditional methods such as leaflet drops and state-run radio. This is especially evident in the Russia-Ukraine conflict, which began in 2014 and escalated sharply in 2022. The war is waged not only with missiles but also through comment sections, encrypted chats, and viral falsehoods, a battlefield where deception is as strategic as firepower.

The Ukraine conflict exemplifies hybrid warfare, in which information manipulation rivals conventional military operations in strategic importance. Hoffman (2007) defines hybrid warfare as combining military force, cyber operations, economic pressure, and strategic communication to achieve geopolitical objectives without triggering full-scale war. Russia deployed multiple information tools, state media, troll networks, fabricated personas, and synthetic media to manipulate narratives and undermine Ukrainian sovereignty and Western resolve (Giles, 2016; Pomerantsev & Weiss, 2014; Nocetti, 2015). These strategies aim to weaken Western political unity, splinter democratic consensus, and skew public perception in addition to obscuring local truths.

Consequently, NATO intelligence agencies have expanded counter-disinformation operations. Key actors include the Cooperative Cyber Defense Centre of Excellence (CCDCOE), the UK's

Government Communications Headquarters (GCHQ), and the now-defunct U.S. Global Engagement Center (GEC). To detect, disrupt, and publicly condemn Russian disinformation networks, these agencies have started collaborating with major social media companies, particularly Meta and Twitter[1] (Bradshaw et al., 2011; Douek, 2020). Data exchange, this partnership includes coordinated content filtering, cooperative threat assessments, and the attribution of state-sponsored activity campaigns. Nevertheless, little is known about the workings, efficacy, and unexpected repercussions of these initiatives.

Research Objective

This research examines how Meta and Twitter collaborated with NATO intelligence agencies to counter Russian disinformation during the 2022-2024 Ukraine conflict escalation. This timeframe encompasses Russia's intensified 2022 invasion, which marked unprecedented hybrid warfare scales and triggered heightened NATO responses (NATO StratCom COE, 2023). Against this backdrop, the central question guiding this is:

"What factors drove collaboration between NATO-affiliated intelligence agencies and social media platforms (Meta and Twitter) in countering Russian disinformation during the 2022-2024 Russia-Ukraine escalation, and what unintended consequences emerged from these public-private partnerships?"

[1] This dissertation uses 'Twitter' to refer to the social media platform, regardless of its post-2023 rebranding to 'X' under Elon Musk's ownership. The choice reflects scholarly convention and avoids conflating branding shifts with substantive changes in platform architecture or historical analysis.

The dissertation also examines the geopolitical and technological context shaping these collaborations, with special focus on how well or poorly they correspond with national and intern ational legal frameworks.

Framework

This study employs two complementary theoretical lenses to analyze intelligence-platform collaboration. Securitization theory (Buzan et al., 1998) explains how disinformation becomes framed as an existential threat, justifying extraordinary measures that transform platforms from content moderators into quasi-security actors. Networked governance theory (Ansell & Gash, 2008) illuminates the operational mechanics of collaboration, how intelligence agencies and platforms coordinate through informal, trust-based networks rather than hierarchical command structures.

Together, these frameworks analyze both why collaboration emerges (securitization creates urgency) and how it operates (networked governance enables coordination), while revealing tensions between security effectiveness and democratic accountability that define modern digital governance.

Significance of This Study

This study examines how social media platforms and NATO-affiliated intelligence services worked together to combat Russian disinformation during the 2022-2024 crisis in Ukraine, through digital governance, international security, and platform regulation. Although previous

studies have extensively documented Russia's disinformation strategies (Giles, 2016; Pomerantsev & Weiss, 2014) and looked at platforms as independent regulators (Gillespie, 2018; Klonick, 2018), this dissertation takes a novel approach by considering intelligence-platform collaboration as a unique governance model that blurs the lines between public and private authority by combining security requirements with corporate responsibility. The study offers the first thorough empirical examination of this hybrid strategy, exposing its underlying conflicts (due process issues, transparency deficiencies) as well as its working mechanisms (data-sharing, cooperative attribution procedures).

In addition to providing practical insights for policymakers tackling pressing issues, such as how the EU's Digital Services Act might regulate crisis responses, how NATO's 2024 Digital Strategy could formalize public-private coordination, and ultimately how democracies can defend against information warfare without compromising their core values, these findings advance theoretical discussions about networked governance (Ansell & Gash, 2008) and securitization (Buzan et al., 1998). Thus, the study is equally relevant to academics studying conflict studies, digital policy, and international relations as well as security professionals negotiating the democratic-authoritarian gap in the information realm.

Outline of the Paper

Having established the strategic importance of information warfare and the study's focus on public-private collaboration, this dissertation now turns to the existing literature on disinformation, platform governance, and intelligence cooperation, particularly NATO-affiliated efforts and the evolution of Russian information warfare. Chapter One reviews key debates and

identifies gaps this study addresses, while also providing historical context on the Russia-Ukraine conflict (2014-2024). It concludes by outlining the theoretical frameworks of securitization and networked governance. Chapter Two details the research methodology, including the qualitative case study design, data sources (e.g., transparency reports, OSINT), and analytical tools such as thematic coding and process tracing. Chapter Three presents the case study of Twitter's collaborative engagement with NATO-aligned intelligence agencies during the Ghostwriter campaign. Chapter Four examines the case of Meta and their response to Operation Doppelgänger and their collaboration with NATO-aligned intelligence agencies. Chapter Five provides a comparative analysis and theoretical discussion, evaluating how differences in platform governance, regulatory context, and securitization shaped the depth and effectiveness of these public-private partnerships. The final chapter consolidates the findings, evaluates the theoretical and normative implications of public-private security collaboration, and proposes directions for future research on its global relevance and sustainability.

Chapter One: Literature Review and Theoretical Framework

Literature Review

The 2022-2024 Russia-Ukraine conflict demonstrates how digital platforms function as battlegrounds for information dominance, establishing disinformation as a primary weapon in contemporary hybrid warfare (Bradshaw & Howard, 2018, p. 24). This section examines how counter-disinformation efforts increasingly depend on partnerships between social media companies and intelligence organizations, emphasizing Meta and Twitter's role in disrupting Russian influence operations. This dissertation conceptualizes collaboration through three components: (1) informational interdependence; bilateral or multilateral exchange of threat intelligence, attribution assessments, and operational insights; (2) operational synchronization; coordination of enforcement actions, public messaging, and response timing; and (3) institutional alignment; convergence of threat narratives, policy frameworks, and strategic objectives (Emerson et al., 2012; Ansell & Gash, 2008; Sørensen & Torfing, 2005). Absence and Integrated Response are the two extremes of the collaboration spectrum, with higher levels denoting further institutionalization, formalization, and reciprocal adaptation of organizational processes. The literature review's three main goals are to: (1) summarize scholarly research on disinformation as a hybrid warfare tool (Giles, 2016); (2) examine changing platform governance models (Douek, 2020); and (3) develop the theoretical framework of networked governance (Ansell & Gash, 2008) and securitization theory (Buzan et al., 1998) that directs this investigation.

The section transitions thematically from broad notions of warfare to specific instances of public-private collaboration. At the same time, highlighting operational examples such as Twitter's 2022 takedowns (McSweeney, 2022) and Operation Doppelgänger, which entailed a concerted effort to spoof reputable European media sites with phony domains. Kremlin-aligned

narratives were disseminated through this on Meta platforms. The NATO StratCom COE report emphasized the hybrid nature of this campaign and detailed how joint attribution was enabled by interagency coordination, especially between German and Lithuanian cyber units alongside open-source intelligence. This case demonstrates how NATO-affiliated organizations are increasingly playing an operational role in influencing threat detection pipelines that platforms like Meta employ (NATO StratCom COE, 2023). The report also highlights critical gaps in the research on real-time counter-disinformation during active conflicts. As content moderation increasingly overlaps with national security imperatives, concerns arise over legitimacy, due process, and jurisdictional boundaries (Klonick, 2018, p. 1623; Kaye, 2019).

By addressing the current lack of research on operational partnerships in live conflict zones, this review positions the dissertation to contribute fresh empirical insights into NATO-platform collaborations during the most intense phase of the Ukraine war (2022-2024). The integration of networked governance and securitization theory offers a novel lens for evaluating the structure and effectiveness of these emerging public-private security systems.

Disinformation as Hybrid Warfare

Disinformation has emerged as a key component of hybrid warfare, especially in Russia's military and geopolitical strategy. A concerted strategy combining cyber, kinetic, and cognitive instruments to cause uncertainty, destabilize enemies, and weaken Western unity was first used with the annexation of Crimea (Giles, 2016; Pomerantsev & Weiss, 2014). This is known as "information confrontation" in Russian doctrine, and it includes both offensive and defensive psychological operations directed at both internal and foreign audiences (Thomas, 2014).

The operational mechanisms of Russia's digital intervention have been outlined in studies by Nocetti (2015) and Polyakova et al. (2018). These studies frequently combine state-sponsored narratives with fake amplification strategies on social media. Early in the 2022 full-scale invasion of Ukraine, false narratives about Ukrainian 'Nazism' and manufactured massacres spread rapidly on Facebook and Twitter, demonstrating the scope and adaptability of these operations (NATO StratCom COE, 2023).

Much of this literature emphasizes disinformation's dual function: it acts as both a tool of persuasion and a mechanism that blurs the boundary between war and peace. These evaluations, however, hardly ever examine the planning and execution of countermeasures, especially those that are collaborative. Furthermore, although the dangers posed by Russian influence operations have been well described, insufficient attention has been paid to the governance frameworks that have been put in place to deal with them. Russian disinformation tactics have militarized digital space, which calls for a reassessment of the ways in which digital platforms regulate speech, particularly during times of geopolitical upheaval.

Platform Governance and Crisis Response

The second important body of research focuses on platform governance, which includes the procedures and frameworks used by social media corporations to filter, delete, or algorithmically devalue material. Platforms are active intermediaries that shape public discourse through design decisions and policy enforcement, rather than being "neutral" bearers of material, as Gillespie (2018) contends.

This role becomes especially pronounced during geopolitical crises. Klonick (2018) and Douek (2020) argue that platform governance has shifted from rule-based moderation to quasi-sovereign adjudication, wherein platforms define and enforce boundaries of legitimacy during international emergencies. While the Russia-Ukraine war intensified these dynamics, earlier turning points included the COVID-19 pandemic and the 2020 U.S. election.

NATO member states have increasingly pressured Twitter and Meta to respond swiftly to coordinated disinformation attacks. However, unlike domestic law enforcement agencies, platforms operate transnationally and typically support state agendas only when collaboration serves their operational, regulatory, or reputational interests (Kaye, 2019).

This literature has not sufficiently examined the micro-politics of platforms' interactions with intelligence services. Few researchers examine how such coordination plays out in high-intensity conflict zones or in situations where attribution is questioned, despite some pointing out the rise of public-private threat-sharing organizations (Bradshaw et al., 2011). In the current study, the nature of collaboration, whether it be directed, negotiated, or symbolic, remains primarily anecdotal.

While the platform governance literature documents how tech companies shape online discourse, it pays insufficient attention to the role of state and intelligence actors within these regimes. The next section situates these relationships within the broader literature on public-private collaboration in security and intelligence.

Public-Private Collaboration in Security and Intelligence Contexts

The cooperation between social media companies and NATO-linked intelligence agencies should be situated within the broader history of public-private partnerships (PPPs) in security governance. While PPPs have traditionally been associated with public services and infrastructure, they have expanded into domains like cybersecurity and counterterrorism, areas that, like disinformation, blur the lines between military and civilian responsibilities.

Scholars such as Cavelty and Egloff (2019) highlight that mutual dependency often underpins cybersecurity PPPs, with both sides leveraging each other's unique capabilities. Businesses rely on state intelligence for threat mitigation guidance, while states depend on access to private digital infrastructures. This interdependence has led to what Carr (2016) terms "co-produced security," where roles are shared, and authority is flexible. However, these collaborations vary widely: some are formalized through memoranda of understanding or multistakeholder forums (e.g., the EU Internet Forum), while others operate informally, especially during periods of geopolitical volatility.

U.S. and U.K. intelligence services, for example, have long partnered with tech companies to identify and remove jihadist content (Awan, 2017; Conway, 2017). Unlike traditional policing, these collaborations often occur without judicial oversight or formal accountability, a pattern that also applies to counter-disinformation partnerships. The GEC, before its disbandment, was a key liaison between government and tech firms in countering foreign propaganda (Global Engagement Center, 2024). NATO's StratCom COE has similarly worked with private threat intelligence firms to trace coordinated inauthentic behavior (NATO StratCom COE, 2023).

These examples reveal that collaboration frequently arises in response to urgency, even in the absence of formal institutional frameworks. However, the informality of such arrangements raises accountability concerns, particularly in liberal democracies where oversight is a normative expectation. This tension underscores the need not only to examine the existence of collaboration but to evaluate its normative implications.

In their idea of weaponized interdependence, Farrell and Newman (2019) observe that nations are increasingly taking advantage of the asymmetric leverage brought about by control over international communication networks. On the other hand, platforms may cooperate to avoid regulation, preserve user confidence, or prevent reputational harm rather than because they share values.

Despite the growing body of research on digital security partnerships, few studies systematically classify or evaluate cooperation in the context of active information operations. Key questions remain: What distinguishes episodic coordination from sustained collaboration? What institutional conditions shape the form and frequency of engagement? These issues remain under-theorized and insufficiently mapped in the current literature. To analyze the means and implications of these public-private partnerships, the next section will look at the dual-theoretical framework.

Theoretical Framework: Networked Governance and Securitization Theory

This dissertation draws on two complementary theoretical frameworks to explain the emergence, dynamics, and variation of collaboration between NATO-affiliated intelligence agencies and social media platforms: networked governance and securitization theory. Each provides distinct analytical insights. Securitization theory helps explain the conditions under which such collaboration becomes politically and normatively acceptable, while networked governance clarifies how these partnerships are structured and operationalized. Together, they illuminate both the discursive processes and institutional mechanisms that underpin these collaborations. This dual-theoretical lens captures the hybrid nature of digital security governance and moves beyond binary categories like collaboration versus coercion or public versus private. The following subsections outline each framework and its relevance to the empirical case studies.

Networked Governance

Networked governance involves horizontal, decentralized relationships between public and private actors (Ansell & Gash, 2008). In contrast to market-based contracts or conventional bureaucratic structures, networked governance is predicated on reciprocal resource interdependence, shared aims, and iterative negotiation. It is frequently used to explain how non-state players, such as businesses and civil society groups, take on the role of co-governors in the formulation and execution of policies, particularly in areas where the state lacks the authority or capability to act alone.

In digital security, networked governance arises when state actors rely on the voluntary cooperation of social media platforms, which they do not directly control. This is particularly

evident in counter-disinformation efforts: governments possess strategic intelligence, but they rely on platforms to implement enforcement, such as content takedowns, demotions, or labeling. Conversely, platforms benefit from intelligence support to validate attribution claims or detect subtle forms of inauthentic behavior. This mutual reliance fosters informal yet operationally meaningful cooperation.

Networked governance is particularly useful for analyzing crisis-driven collaboration, where legal mechanisms may be too slow or rigid. In such contexts, cooperation may take the form of ad hoc working groups, private threat-sharing channels, or joint attributions involving NATO-linked analysts. When facing state-sponsored disinformation, platforms like Meta have operated as security actors. Meta's response to Operation Doppelgänger included synchronized takedowns, attribution of inauthentic behavior to Russian sources, and an unusual public partnership with European intelligence services (Meta, 2023; Miguel et al., 2024). This informal, adaptive structure aligns with what Sørensen and Torfing (2005) call "interactive political coordination," where authority is distributed and legitimacy emerges through repeated, multi-actor engagement. Though spontaneous, these networks are often effective due to their speed and scale.

Networked governance isn't necessarily fair or cooperative, though. Such structures frequently suffer from unequal power relations, a lack of legitimacy, and blurred lines of accountability, as Klijn and Skelcher (2007) contend. When national security logics are introduced into public-private networks, these characteristics can be particularly troublesome as they may evade democratic scrutiny. Concerns regarding transparency and the degradation of civil liberties are raised by the possibility that intelligence agencies and platforms may collaborate outside the

purview of judicial review or parliamentary supervision. There is a need for an integrated paradigm that incorporates securitization theory because, whereas networked governance sheds light on the organizational architecture of collaboration, it does not adequately account for the political processes that either permit or restrict these partnerships.

Securitization Theory

The Copenhagen School's securitization theory explains how certain issues are framed as existential threats to justify extraordinary responses (Buzan et al., 1998; Waever, 1995). Securitization is a performative act in which a securitizing actor, such as a state or political leader, identifies a referent object (e.g., national sovereignty or democratic institutions) as being under threat. They then persuade an audience to accept emergency measures that bypass normal political processes. Crucially, successful securitization depends not only on the actor's framing but also on the audience's acceptance.

This framework helps explain how disinformation has been elevated from a content moderation issue to a national security concern. In response to Russia's 2022 invasion of Ukraine, Western governments, NATO entities, and intelligence services employed securitizing rhetoric that framed digital deception as a form of cognitive warfare rather than mere propaganda (Miguel et al., 2024). This discursive shift legitimized the move away from conventional moderation practices toward real-time, security-driven interventions, enabling deeper collaboration between platforms and intelligence agencies.

Platforms have positioned themselves as digital first responders in the information battle and have adopted rhetoric that is increasingly infused with security. For instance, Meta's "Threat Report" series now uses references to hostile state actors, influence campaigns, and adversarial behavior, mimicking the format and tone of intelligence briefings (Meta, 2023). According to this alignment, securitization not only makes collaboration possible but also determines its nature by establishing the framework for the formation and normalization of extraordinary partnerships, which permits platforms to function as de facto players in national security governance.

However, normative trade-offs are also highlighted by securitization theory. There is less room for due process, openness, and democratic discussion when disinformation is viewed as a national security issue. Ordinary legal and ethical restrictions do not apply to exceptional actions, such as cross-border intelligence cooperation or bulk takedowns (Krebs & Jackson, 2007). As a result, securitization contributes to the understanding of how intelligence-platform interactions are formed and their disputed legality. It highlights the contradiction of protecting democracy via means that could jeopardize its core values. The next section examines how these two theories complement one another.

Complementarity of the Frameworks

Securitization theory describes the discursive and political circumstances that lead to the institutionalization and acceptance of collaboration, whereas networked governance analyzes the structure and operational logic of collaboration. When combined, the two frameworks allow for a multi-level examination of the following topics: the reasons behind collaboration (securitization), its structure and operationalization (networked governance), and the democratic tensions that

develop (both theories).

The secondary goal of the dissertation, which is to evaluate the level of collaboration in each instance and determine the variables influencing variance, is also supported by this dual-framework method. This theoretical toolbox is used in two incidents, as explained in the next section: Twitter's elimination of Ghostwriter accounts connected to the GRU and Meta's removal of Operation Doppelgänger.

Application to Case Studies and Evaluating Degrees of Collaboration

This dissertation applies the integrated frameworks of securitization and networked governance to two case studies: Twitter's 2022 takedown of GRU-linked "Ghostwriter" accounts and Meta's response to the Russian-affiliated "Operation Doppelgänger" network. These cases were selected based on three criteria: (1) availability of public data, including transparency reports, independent investigations, and OSINT repositories; (2) clear involvement of NATO-affiliated or Western intelligence actors; and (3) explicit attribution to Russian state or state-linked entities. This selection allows for systematic comparison across platform responses within a shared geopolitical and temporal context, while ensuring empirical feasibility.

Finding the level and type of cooperation between platforms and intelligence agencies in each instance is the main goal of the analysis. The study reconstructs the initiation, operationalization, and public framing of these collaborations by process tracing. Key indicators that assist in placing each case along the collaboration spectrum are identified using thematic coding. These indicators include synchronicity in takedown actions, narrative alignment, and information-

sharing modalities. Both cross-case evaluation of collaborative dynamics and within-case causal process analysis are possible with this dual-level methodology.

This section lays the groundwork for evaluating the motivations, structures, and implications of intelligence-platform partnerships by firmly establishing the empirical inquiry within a theoretical framework that is easily understood. The following section introduces a typology to gauge levels of collaboration, and the institutional and political elements that influence variation between cases are evaluated.

Degrees of Collaboration

This dissertation views collaboration as a spectrum rather than a binary variable that is either present or absent. Drawing on Sørensen and Torfing (2005), Ansell and Gash (2008), and recent cybersecurity collaboration models (Cavelty & Egloff, 2019), the following typology is used to assess the degree of collaboration in each case, ranging from no interaction (Level 0) to fully integrated, real-time cooperation (Level 4):

Degree	Form	Indicators
0	Absence	No evidence of intelligence- platform interaction
1	Symbolic Alignment	Public statements or mutual concern, no operational cooperation
2	Informal Coordination	Ad hoc exchanges (e.g., email alerts, private briefings)
3	Strategic Partnership	Regularized collaboration (e.g., shared threat dashboards, joint reports)
4	Integrated Response	Co-produced attribution, synchronized removals, real-time cooperation

The relevant platform's collaboration with intelligence actors along this spectrum is assessed in each case study chapter (Chapters 3 and 4) using process tracing and thematic classification of

primary and secondary source material. In addition to analyzing the political context, for example, whether collaboration was presented as a reaction to an existential threat, therefore triggering securitizing logic, this method enables the analysis to separate the operational processes of collaboration.

Factors Influencing Collaboration

Building off the previous section, the theoretical framework enables the identification of key elements that influence the stability and depth of intelligence-platform cooperation. These consist of:

Level of securitization: To what extent was the threat presented as a wartime or existential peril?

Model of platform governance: Did the platform follow crisis protocols, such as Twitter's post-acquisition moderation uncertainty versus Meta's CIB policy?

Leverage of the state: Did the intelligence actors work from a country that had normative or regulatory control over the platform?

Networks of trust: Was information sharing facilitated by pre-existing trust ties (such as prior cyber collaboration or NATO working groups)?

Regimes of accountability: Which legal or normative structures (such as parliamentary oversight or transparency standards) limited or validated the partnership?

Analytical Contribution

This chapter lays the theoretical groundwork for analyzing when and how intelligence-platform collaboration happens. By integrating insights from networked governance and securitization

theory, it explores an area that remains largely unclear and underdeveloped empirically. The next chapter outlines the qualitative, comparative methodology used to investigate NATO-platform collaboration, detailing case selection, data sources, and analytical approach.

Chapter Two: Methodology

Research Design

This dissertation adopts a qualitative, comparative case study design to investigate collaboration. A qualitative approach is well-suited to capturing the complexity, fluidity, and opacity of public-private security partnerships. It enables the unpacking of decision-making processes, institutional coordination, and strategic framing.

The primary objective is explanatory: to understand the circumstances in which collaboration arises, the forms it takes, and the outcomes it yields. To achieve this, the study uses a cross-case thematic comparison in conjunction with a within-case process tracing technique, which enables the identification of both patterns of variation and causative mechanisms. The objective is to develop empirically supported knowledge of how intelligence-platform cooperation is formulated, operationalized, and justified in high-intensity geopolitical crises rather than proving a predetermined premise.

The five-point typology introduced in Chapter One is used to evaluate collaboration, allowing cooperative behavior to be systematically categorized based on observable indicators such as shared narratives, synchronized attribution, and platform response timing. This typology supports both within-case and cross-case comparisons by standardizing how collaborative activity is interpreted.

The selection of a qualitative design is further supported by the nature of the available data. The empirical data, which includes academic analyses, policy documents, OSINT investigations, and transparency reports, is detailed but difficult to measure. An interpretive, process-oriented

analysis that prioritizes institutional interactions, governance structures, and discursive shifts over aggregate results is the most appropriate use case for these sources.

The case study design also aligns with the theoretical framework outlined in Chapter One. Both networked governance and securitization theory require contextual interpretation of how collaboration occurs across institutional boundaries and how threats are constructed. A qualitative approach allows for tracing securitizing speech acts, identifying coordination nodes, and mapping the formal and informal pathways through which public-private partnerships emerge.

Lastly, the two chosen cases represent important events that provide opposing real-world examples of cooperation. Although there are differences in platform governance models, regulatory frameworks, and operational transparency levels, both entail Russian state-affiliated disinformation, intelligence involvement, and social media response. The findings' explanatory power is strengthened, and the typology created in Chapter One's generalizability is improved through a comparative study of these cases. The following section will take a closer look at the case studies, outlining the rationale behind the choices.

Case Studies

Through a deliberate sample method based on theoretical relevance, empirical accessibility, and strategic variety, the two cases considered for this study were chosen. These cases are perfect for comparative research since they both provide examples of public-private coordination during the Russia-Ukraine crisis (2022-2024), but they also differ in terms of institutional background and

platform reaction.

First, both cases are attributed to Russian state-linked or affiliated actors, as confirmed by platform threat research teams, independent intelligence firms (e.g., Graphika, Mandiant), and NATO-affiliated bodies such as the StratCom COE. This attribution is crucial for applying securitization theory, as it frames the disinformation as a hostile, foreign-led threat.

Second, there is evidence indicating the platforms have interacted with Western or NATO-affiliated intelligence services, either through coordinated messaging, shared assessments, or direct attribution. Both provide enough open-source signals, including shared declarations and parallel disclosures, to support a process-tracing methodology, even though not all collaboration is publicly acknowledged.

Third, the instances provide information that is easily comprehensible and adequately documented. NATO StratCom assessments, quarterly Threat Reports, and independent research by EU DisinfoLab have all documented Meta's removal of Operation Doppelgänger. In a similar vein, OSINT communities and independent investigations have validated Twitter's takedowns, which have also been examined in Twitter Transparency Reports. The creation of trustworthy event timelines and the thematic classification of significant cooperative moments are made possible by the availability of rich, triangulable data sources.

Finally, the cases differ strategically in terms of political environment, organizational capacity, and governance structure. Meta maintains an internal threat intelligence team and operates under a formal CIB policy, which targets networks that mislead users about their identity or purpose (Murero, 2023). Twitter, by contrast, experienced layoffs, policy shifts, and governance

disruption following its 2022 acquisition, factors likely affecting its capacity or willingness to collaborate with intelligence actors. This divergence allows for analysis of how internal platform dynamics shape external cooperation.

In summary, the chosen cases are conceptually interesting and analytically rich, but they are not typical of all intelligence-platform interactions. They are essential case studies for investigating the operational aspects of digital security cooperation in times of war, as well as for testing and implementing the framework laid out in the previous chapter. These case studies establish the empirical foundation for evaluating collaboration. Building on the theoretical and case selection rationale, the following section outlines the data sources that enable the triangulation of platform-intelligence coordination patterns.

Data Sources

Due to the classified nature of direct intelligence-platform interactions, this study relies on open-source and publicly available materials to trace cooperation between social media companies and NATO-affiliated intelligence bodies. A triangulated dataset is constructed from five key sources: (1) Meta and Twitter's transparency reports and platform disclosures, which detail takedown data, threat attributions, and mentions of CIB, often referencing collaboration with entities like NATO StratCom COE; (2) NATO and Western government documents, such as reports from the GEC, EEAS, and NATO StratCom COE, which reflect threat-sharing dynamics and corroborate platform actions; (3) Investigative journalism and OSINT, which expose informal or undisclosed state-platform ties, as seen in Operation Doppelgänger. (4) Academic literature from scholars such as Douek (2020) and Farrell & Newman (2019), which provides theoretical grounding on

informal cooperation and asymmetric power in digital governance; and (5) Public interviews and statements from NATO officials and platform representatives, which offer discursive cues about strategic alignment and institutional narratives.

Together, these sources enable the reconstruction of cooperation patterns in a plausible, though indirect manner, despite the opacity of classified interactions. This is examined further in the following section, dealing with the research limitations.

Limitations

This research faces several limitations, primarily due to the classified and opaque nature of many intelligence-platform interactions. To compensate, it relies on triangulated open-source materials, including transparency reports, investigative journalism, and OSINT networks, which offer valuable but indirect evidence. These sources cannot uncover backchannel communications or internal strategic motivations. Likewise, the study's temporal scope (2022-2024) constrains its ability to assess the durability of observed collaboration patterns over time.

In essence, this methodological framework sets the stage for the upcoming case process tracing that takes place in the following chapters. Chapter Three will examine Twitter, while Chapter 4 will unpack Meta.

Chapter Three: Twitter Takedown

Introduction

This chapter observes Twitter's reaction to the Ghostwriter disinformation campaign during the 2022-2024 Russia-Ukraine conflict. Using process tracing and thematic analysis, it identifies informal coordination with NATO-affiliated intelligence actors through attribution patterns, semantic alignment, and synchronized enforcement. The chapter applies the five-level typology to classify this cooperation as a Level 3 Strategic Partnership, shaped by internal governance shifts, regulatory uncertainty, and trust-based networks.

The Ghostwriter operation exemplifies state-sponsored disinformation during the 2022-2024 escalation, directly linking Russian intelligence to attempts at influencing Western public opinion and undermining NATO political unity. Unauthorized access to digital accounts, the spread of false information, and coordinated influence operations on social media platforms to sway information ecosystems were the main components of the campaign (Cardiff University, 2023, p.3).

The operation is analytically significant for two key reasons. First, it provides a vivid illustration of hybrid warfare, in which digital platforms are used as battlefields for information control and geopolitical competition (Hakala & Melnychuk, 2021). Second, it provides a test case for assessing the dynamics and efficacy of cooperation between NATO-affiliated intelligence services and social media platforms, in this case, Twitter, in the fight against state-sponsored disinformation.

This case illuminates the cooperative measures deployed against recognized disinformation threats. Official transparency reports and independent analyses of Twitter's responses from early

2022 to mid-2023 offer a solid empirical foundation for evaluating public framing tactics, attribution procedures, and coordination levels (Clayton, 2022; McSweeney, 2022).

The main goals of this case study are to shed light on the operationalization of collaborative efforts, assess the degree to which public-private partnerships developed, and pinpoint the determinants that influenced the collaborative landscape, such as internal governance frameworks, regulatory environments, and crisis perceptions. The comprehensive analytical framework developed in previous chapters is used in the following sections to methodically unravel these themes. To understand how stakeholders collaborated against Ghostwriter, this analysis begins by examining the campaign's genesis and authoritative attributions before exploring Twitter's adaptive countermeasures.

Ghostwriter

The term Ghostwriter describes a widespread and ongoing disinformation campaign that first appeared in 2017 but became much more intense with the early 2022 escalation of the Russia-Ukraine conflict. The operation, which was carried out by actors claiming to be from the GRU, was designed to undermine NATO's eastern flank and destabilize Western democracies (Cardiff University, 2023). With a focus on Ukraine, Poland, Lithuania, Latvia, and Germany, nations strategically positioned within NATO's geopolitical framework, the campaign's main strategy involved hacking trustworthy websites and social media accounts to spread polarizing narratives and disinformation (Roslon et al., 2024).

To create and spread contentious political ideas, the Ghostwriter campaign employed strategies like posing as politicians, journalists, and well-known public figures. The main goals of the

material were to worsen political tensions, undermine public trust in democratic institutions, and erode support for Ukraine and NATO's strategic unity (Schuette, 2023). These techniques were deliberately used to take advantage of political tensions related to matters like migration, military assistance to Ukraine, and accusations of domestic political corruption (Golovchenko et al., 2018).

In coordinated statements released in the fall of 2021, NATO and European Union agencies formally and explicitly attributed this to the Russian State. The EU characterized Ghostwriter as a systematic, coordinated, and state-supported information operation and urged the Russian Federation to adhere to the norms of responsible state behavior in cyberspace (Council of the EU, 2021). Similarly, the NATO Strategic Communications Centre of Excellence specifically designated Ghostwriter as a "Russian-state backed cyber influence campaign", highlighting the dual-purpose operational techniques that integrated information warfare and cyber intrusion to accomplish strategic geopolitical goals (Cardiff University, 2023).

This unambiguous, cohesive attribution signaled a dramatic change in how disinformation was framed in Western security discourse. A significant step toward securitizing Western intelligence services and public diplomacy platforms was the designation of Ghostwriter as state-sponsored cyberwarfare rather than just a string of separate influence operations (Orenstein, 2025). It is important to keep in mind how this accurate portrayal of Ghostwriter affected Twitter's activities and its subsequent collaboration with NATO-affiliated organizations as this analysis progresses. Having established the strategic context and clear attribution of Ghostwriter to the Russian State, the next section explores the specific measures taken by Twitter in response to this threat.

Platform Response

Russia's 2022 invasion marked a turning point in Twitter's approach to state-sponsored disinformation. The platform implemented major policy changes targeting Russian disinformation and propaganda tactics. Twitter declared in February that it was expanding its labeling policy for "Russia state-affiliated media" to combat disinformation that favored the Russian government, just after Russia invaded Ukraine (Aguerri et al., 2024). Although Twitter never explicitly referenced Ghostwriter in its transparency reports, the timing, target geography, and thematic focus of enforcement actions in early 2022 strongly aligned with the characteristics of the campaign described by NATO StratCom COE and the European External Action Service. Therefore, this analysis infers Twitter's response to Ghostwriter based on correlated evidence rather than formal acknowledgment. In reaction to geopolitical events, this signified a change from reactive content control to proactive policy enforcement.

While Ghostwriter was not formally named in Twitter's reports, the platform's 2022 enforcement actions were particularly frequent. Takedowns targeting CIB closely tracked ongoing disinformation patterns identified by NATO and security researchers (Bradshaw & Howard, 2018). As Twitter attempted to strike a balance between countering foreign manipulation and protecting free expression, its evolving policy regime represented one of the platform's most comprehensive responses to state-linked influence operations.

In May 2022, Twitter expanded enforcement by algorithmically suppressing misleading posts about the Russian invasion. The platform stopped amplifying content that falsely accused individuals of war crimes or misrepresented conflict conditions (Press, 2022). Twitter took a more interventionist approach to information regulation by restricting the reach of content rather than merely labeling it. While this strategy enhances response speed and mitigates harm, it also

raises concerns about opaque decision-making and potential overreach in times of crisis, particularly without judicial or multistakeholder oversight. The action foreshadows more intricate forms of public-private crisis coordination and reflects Twitter's slow transition from a passive conduit to an active security actor.

Academic studies have documented the magnitude and significance of these platform responses. During the early stages of the invasion, pro-Russian messages received about 251,000 retweets and reached about 14.4 million users, according to an analysis of Twitter data (Geissler et al., 2023). However, there was evidence that bots disproportionately contributed to the spread of pro-Russian messages and increased their reach (Geissler et al., 2023). These results demonstrated the effectiveness of platform countermeasures as well as the scope of the disinformation problem.

When Elon Musk acquired Twitter in October 2022, the platform's governance regarding state-sponsored disinformation became more complex. Under Musk's control, the platform relaxed content regulations, reduced enforcement personnel, and retracted from the voluntary code of conduct for countering disinformation that was enthusiastically embraced in June 2022. EU authorities expressed concerns about the platform's ability to continue thwarting Russian propaganda attempts after the business also removed the state-affiliated media labels it had previously used (Menn, 2023).

Given Twitter's growing role as a major player in the dynamics of information warfare, these platforms' responses highlight the changing relationship between technology governance and global security issues (Bradshaw & Howard, 2018; Iuliia Alieva et al., 2022). Although the long-term efficacy and consistency of these measures are still being examined and debated. The documented policy changes and enforcement actions offer empirical evidence of how the

platform modified its operational framework in response to state-backed disinformation campaigns. Although Twitter's proactive enforcement measures addressed the Ghostwriter danger, a more thorough analysis of terminology, timing, and operational parallelisms is necessary to determine the degree of platform-intelligence cooperation. The next section applies the framework to evaluate better how these relations align.

Intelligence-Platform Coordination

Platform responses overview counter-disinformation tactics but assessing alignment with security actors reveals deeper insights into effectiveness. Although public records generally lack explicit disclosures of intelligence-platform coordination, the evidence that is currently available points to informal cooperation between NATO-affiliated intelligence agencies and Twitter in addressing threats such as the Ghostwriter campaign, a persistent Russian disinformation campaign that security researchers have been documenting since at least 2017 (Group, 2022; Mandiant, 2020)**.**

Twitter took several coordinated actions to combat Russian disinformation after the invasion in 2022. These included broader enforcement actions against CIB in March and April 2022 (Geissler et al., 2023) and expanded labeling policies for "Russia state-affiliated media" accounts announced on February 28, 2022 (Aguerri et al., 2024; Press, 2022).

These platform responses frequently coincided with threat intelligence cycles and assessments of the larger security community, indicating that platforms and security institutions had informal coordination mechanisms and a common understanding of new threats (Kausche & Weiss, 2024).

Twitter's usage of language specific to security indicates a stronger adherence to the threat frameworks of the EU and NATO. Currently, frequently used in Twitter's transparency reports, phrases like "state-backed influence operations" and "coordinated inauthentic behavior" are very similar to those found in NATO and EU documentation (Starbird et al., 2019). According to Lynskey (2017), this semantic overlap suggests that platform and state actors are responding to convergent threat perceptions and operating from comparable analytical baselines. Such convergence, however, can legitimize narratives driven by intelligence in platform governance, so excluding civil society perspectives and making it more difficult to guarantee accountability in content removal procedures. Accordingly, terminological alignment highlights the power imbalances present in informal security cooperation even as it improves tactical coordination (Iuliia Alieva et al., 2024).

Platform officials regularly reference collaboration with security partners and external stakeholders in public statements, indicating ongoing informal coordination networks (McSweeney, 2022). Such language emphasizes that Twitter's strategic response probably depended on intelligence inputs from reliable external entities, even in the absence of formal public agreements. This increased the effectiveness and speed of enforcement actions against threats like the Ghostwriter campaign and other Russian disinformation operations (Iuliia Alieva et al., 2024).

The information that is currently available points to what might be called informal coordination as opposed to formal partnership, using synchronized response timing, shared threat intelligence indicators, aligned analytical frameworks, and informal communication channels (Carr, 2016). The intricate task of combating state-sponsored disinformation in a setting where platforms preserve their independence while acknowledging the necessity of more extensive security

collaboration is reflected in this coordination (Gorwa, 2019). Although the lack of official disclosure requirements means that the full extent of coordination is still unknown and necessitates more research, this evaluation of coordination patterns offers crucial context for comprehending how social media platforms react to state-sponsored threats (Zia et al., 2023). This evidence forms the basis for a thematic analysis to determine precisely how and at what level Twitter collaborated with NATO-affiliated intelligence entities.

This section demonstrates how significant security collaboration can still occur even in the absence of formal agreements. Through the documentation of temporal synchronization, semantic convergence, and common danger perception patterns, this investigation contributes to a better understanding of the growing role of digital platforms as unofficial security partners. Semantic convergence refers to the alignment of language, terminology, and framing used by different actors when describing threats or security issues (Ameel et al., 2009). Additionally, it raises the possibility that public-private coordination is more systemic than is currently recognized in the literature. To better understand the nature of this alignment and its implications for collaboration, the next section applies a thematic analysis using the framework from Chapter One.

Thematic Analysis of Collaboration

The synchronization of enforcement activities is a crucial theme indicating of effective partnership. Although the precise number of accounts deleted due to Ghostwriter on Twitter is not made public, the platform's enforcement actions against Russian disinformation activities during this time coincided with security researchers and NATO institutions paying more

attention to the campaign (Gielewska, 2021; Twitter, 2021). According to cybersecurity company FireEye, the Ghostwriter effort has been going on since at least March 2017 and is being carried out by a state-sponsored cyber-espionage group. These timing patterns point to possible coordination since coordinated actions by many players could be a sign of cooperative operational planning and strategic intelligence sharing (Mandiant, 2020).

The narrative and vocabulary alignment between Twitter and NATO-affiliated groups demonstrated considerable consistency, further supporting the possibility of a strategic alliance. Often using specific terms like “state-backed influence operations” and “coordinated inauthentic behavior”, Twitter’s transparency reports reflected threat frameworks developed by the EU and NATO (Twitter, 2021; Gielewska, 2021). The adoption of common analytical frameworks or coordinated communication tactics, which are traits of informal coordination arrangements, may be indicated by such convergence in threat narratives (Starbird et al., 2019).

Informal channels for sharing knowledge are another important theme. Twitter mentioned using external threat intelligence for detection and response efforts, despite the absence of clear formal agreements (Twitter, 2021). Platform enforcement activities and indications recorded by security institutions were found to be aligned by independent security studies, indicating possible informal intelligence exchanges (Gorwa, 2019). This data points to the continuous, mutually trusting, semi-formalized intelligence sharing, which is typical of strategic collaborative settings (Ansell & Gash, 2008).

Public declarations from Twitter staff members and institutional involvement also pointed to cooperative dynamics. In public forums and industry talks, platform authorities mentioned continued collaboration with external security partners and stakeholders (McSweeney, 2022)**.** This type of language, frequently used in informal intelligence-sharing agreements, suggests that

regular communication occurred between the platform's security teams and security institutions, even in the absence of officially announced formal partnerships (Gorwa, 2019).

Instead of formal, institutionalized agreements, this case shows reliable collaboration that is primarily enabled by informal trust-based networks and shared intelligence frameworks. The importance of informal coordination in modern public-private security arrangements is confirmed by this thematic analysis, which offers a strong empirical basis for examining the underlying causes and wider ramifications of this cooperation in the sections that follow. Identifying the precise form of collaboration enables a more nuanced exploration of the factors driving and shaping these cooperative dynamics. Building on the thematic and coordination analysis, the next section will apply the typology to classify the nature of the collaboration.

Collaboration Typology Assessment

The evidence outlined in the preceding sections can now be mapped against the five-level collaboration typology introduced in Chapter One. The case of Twitter's engagement with NATO-affiliated intelligence services in response to the Ghostwriter campaign demonstrates key traits of Level 3: Strategic Partnership.

This level of collaboration is characterized by informal but action-oriented cooperation through shared threat recognition, semantic alignment, and synchronized enforcement cycles. Twitter's public use of terms that mirrored NATO and EU threat vocabularies (Twitter, 2021; Gielewska, 2021), suggesting discursive convergence. Similarly, enforcement actions against Ghostwriter-linked assets, though not formally attributed, coincided with public disclosures from NATO

StratCom COE and EU actors, indicating temporal alignment and shared threat assessments (Mandiant, 2020; McSweeney, 2022).

While no formal documents or legal frameworks were publicly disclosed, Twitter executives' frequent references to security partners and the platform's documented coordination with trusted intelligence and OSINT communities reveal the presence of stable, trust-based informal coordination networks (Gorwa, 2019; Carr, 2016). This is consistent with Level 3 dynamics, where cooperation is not institutionalized through treaties or permanent bodies but is nonetheless sustained through repeated interaction, convergent priorities, and reciprocal legitimacy.

This typological assessment underlines the nature of Twitter's engagement that reflects a strategic but less institutionalized form of public-private cooperation, shaped significantly by internal volatility and regulatory uncertainty under new ownership. To understand what enabled this level, the next section unpacks the institutional, political, and platform-specific conditions that facilitated this.

Factors Influencing Collaboration

The informal but significant strategic collaboration was shaped by several important elements. Securitization, platform governance, state leverage, trust networks, and accountability regimes are the five primary factors that account for the result, in accordance with the conceptual framework presented in Chapter One.

Collaboration dynamics were greatly impacted by internal governance structures. The framework for possible coordination efforts was supplied by Twitter's well-established content moderation guidelines and specialist teams tasked with countering state-sponsored disinformation (Twitter,

2021). However, major governance changes were brought about during Musk's takeover of Twitter in October 2022, including adjustments to policies and staff reductions in content moderation teams (Hickey et al., 2025). The platform's approach to external coordination and content control saw significant modifications because of these internal upheavals (Kahn, 2024).

The collaborative atmosphere was profoundly affected by NATO and EU organizations' securitization of threats. Since at least 2017, the Ghostwriter campaign has been known to target NATO partners, specifically Poland, Lithuania, and Latvia, to destabilize trust in NATO activities (Cardiff University, 2023). Increased securitization frequently validates more comprehensive information sharing and deeper collaboration between platforms and security organizations, according to academic research (Buzan et al., 1998; Cavelty & Egloff, 2019). By attributing the campaign to Russia and characterizing it as an attempt to undermine NATO, the EU has elevated the issue from information manipulation to one of regional security concerns.

Coordination dynamics with intelligence actors were considerably changed by regulatory uncertainties after Twitter's ownership shift. The platform now known as X changed its moderation guidelines and removed protections like verified labels, which led to uncertainty in enforcement priorities (Gillespie, 2018). Formal collaboration with security agencies was probably challenging in this dynamic climate, which increased the need for adaptable, ad hoc coordination. However, as judgments on speech suppression are made without explicit legal or procedural guidelines, this informality can erode transparency and democratic oversight. As a result, although regulatory ambiguity promoted operational flexibility, it also made governance gaps worse, which makes public-private partnerships in digital security less legitimate (Ansell & Gash, 2008).

Potential coordination was greatly aided by unofficial networks built on confidence. Operational coordination may have been streamlined by trust networks formed through previous professional interactions in cybersecurity communities. This then established working relationships between platform security teams and security professionals (Clark & Rid, 2021; Valeriano et al., 2018). Although attribution necessitates a comprehensive examination of the available material, security companies such as FireEye have established the Ghostwriter campaign's congruence with Russian security aims. Coordinated responses to state-sponsored disinformation efforts may benefit from regular, if unofficial, exchanges of threat intelligence (Ischebeck-Baum, 2015; Aldrich, 2009).

When taken as a whole, these elements produced circumstances that could help with informal cooperation dynamics when combating state-sponsored disinformation. Understanding these factors provides a framework for examining how platforms and security institutions navigate the challenge of countering information operations while maintaining operational flexibility and adapting to evolving threats. (Ansell & Gash, 2008). By demonstrating that ad hoc, trust-based networks can produce strategic coordinating outcomes even in the absence of formal legal structures, these findings add to broader discussions on platform governance and hybrid threat mitigation.

Summary

To combat the Ghostwriter campaign, this chapter analyzed the collaboration between Twitter and intelligence organizations associated with NATO. The relationship was categorized as a Level 3 Strategic Partnership using the established analytical framework, which was defined by

coordinated operational responses, consistent narrative alignment, and substantial informal information exchange. Despite the lack of official, publicly announced agreements, these components were evident.

To combat geopolitical risks, key insights show that platforms such as Twitter have become increasingly integrated into national and international security systems. A significant change in public-private security partnerships is reflected in informal but reliable coordination that is powered by flexible frameworks and trust-based networks. These unofficial processes bring up important issues about accountability, transparency, and the moral implications of contemporary security governance.

A closer examination of platform governance, securitization processes, and regulatory ambiguity under Musk's ownership reveals how each shaped Twitter's fluctuating cooperation with NATO-affiliated intelligence actors. While coordination did occur largely through semantic alignment and temporal synchronization, it lacked formalization and was constrained by internal instability. These findings offer a strong empirical foundation for the next chapter, which turns to Meta's more institutionalized and proactive response to Russian disinformation during Operation Doppelgänger. Chapter Five will then bring both cases into direct comparison to assess how different platform governance models influence the depth, durability, and risks of public-private security collaboration.

Chapter Four: Operation Doppelgänger

Introduction

This chapter investigates Meta's handling of Operation Doppelgänger during the 2022-2024 escalation of the Russia-Ukraine war. Drawing on process tracing and thematic analysis, it reveals high levels of collaboration with NATO-affiliated intelligence agencies, reflected in coordinated attribution, aligned narratives, and joint enforcement rhythms. Using the five-tier typology developed earlier, the case is categorized as a Level 4 Integrated Response, shaped by robust platform governance, regulatory engagement, and institutionalized trust networks.

The Russia-Ukraine conflict has served as a live testbed for advanced information operations. Among the most complex of these was Operation Doppelgänger, a pro-Russian disinformation campaign that sought to influence European public opinion by mimicking legitimate news outlets and disseminating divisive narratives, particularly via Meta's Facebook platform.

Discovered in mid-2022 and active through 2024, Doppelgänger employed paid ads, fabricated websites, and AI-generated personas to portray NATO as an aggressor and erode Western support for Ukraine. Its analytical significance lies in both its operational sophistication and the coordinated response it triggered across NATO-affiliated intelligence bodies and Meta's internal threat units. As Sessa and Miguel (2024) note, Doppelgänger exemplifies cognitive warfare, where digital infrastructures are contested spaces and public trust becomes the primary target. This case offers a unique opportunity to examine real-time collaboration between a major tech platform and Western security institutions. Whereas Chapter Three analyzed Twitter's engagement amid institutional instability, this chapter focuses on Meta's comparatively

structured and proactive coordination, particularly as it intersected with emerging governance frameworks such as the EU DSA and NATO's 2024 Digital Strategy.

Applying the dual theoretical framework developed earlier in the dissertation, this chapter uses securitization theory to explore how actors justified extraordinary measures by framing Doppelgänger as an existential threat, and networked governance theory to examine how collaboration was negotiated through synchronized actions, shared narratives, and institutional linkages. Mirroring Chapter Three, this chapter begins with an analysis of the operation, followed by an examination of Meta's response.

The Doppelgänger Campaign

Doppelgänger represented one of the most sophisticated pro-Russian influence campaigns during the 2022-2024 Russia-Ukraine conflict. Following the exposure, it was revealed that the operation deployed spoofed media domains, AI-generated avatars, and algorithmically amplified propaganda across Meta. With an emphasis on targeting NATO-aligned audiences in Germany, France, Poland, and Italy (Nimmo, Gleicher, et al., 2023; Alaphilippe et al., 2022). At its core, the campaign created forged news sites that closely mimicked the branding and tone of legitimate outlets. These domains circulated Kremlin-aligned disinformation, which was amplified by networks of inauthentic social media accounts crafted to simulate genuine civic engagement. The narratives centered on NATO expansionism, Ukrainian corruption, and economic fatigue were engineered to erode European solidarity with Ukraine and undercut NATO cohesion (Frühwirth & Nazari, 2024).

By early 2023, attribution to Russian military-intelligence networks had gained traction. Although Meta did not explicitly name the GRU, its reports pointed to infrastructure and behavioral patterns previously linked to Russian entities such as the Social Design Agency and Structura National Technologies (Ronzaud et al., 2023; Insikt Group, 2023). Independent assessments reinforced this attribution. NATO StratCom's 2023 digital threat report identified linguistic and infrastructural parallels with earlier GRU-linked operations, including "Secondary Infektion" (Ronzaud et al., 2023). Similarly, the EEAS described Doppelgänger as "a continuation of Russian state-aligned disinformation practices, amplified through platform vulnerabilities and jurisdictional gaps" (EU External Action, 2024).

Although the campaign relied on private contractors and marketing intermediaries, the Swedish Psychological Defence Agency's 2025 forensic review concluded that its scale, timing, and infrastructure indicated centralized planning and state direction (Pamment & Tsurtsumia, 2025). The same report noted that many of the spoofed domains were hosted on servers previously associated with the Social Design Agency, which the UK government sanctioned in October 2024 for its role in malign foreign influence operations (Office, 2024). Parallel to these technical findings, state and intergovernmental actors contributed to shaping the narrative around Doppelgänger. NATO and EU institutions increasingly framed the operation as part of a broader shift in Russian hybrid warfare doctrine where disinformation no longer functioned merely as reputational disruption, but as an existential challenge to the information sovereignty of the Euro-Atlantic community.

NATO formally characterized Doppelgänger as part of Russia's broader use of "hybrid means, including disinformation, against NATO Allies and partners," highlighting the integration of conventional, cyber, and informational tactics to undermine regional stability (NATO, 2024).

The European Commission echoed this stance, accusing Doppelgänger of violating the EU Code of Practice on Disinformation and invoking the transparency provisions of the DSA to pressure platforms such as Meta to disclose their response mechanisms (European Commission, 2024).

This discursive shift reframing disinformation from a policy violation to a national security concern enabled deeper institutional convergence between Meta and NATO-aligned actors. It established a common attribution framework and paved the way for unprecedented cooperation, including coordinated enforcement actions and large-scale ad bans. In effect, Meta began operating not only as a content platform but as an informal actor within European security governance. With the strategic intent and state alignment of the campaign now established, the next section examines how Meta operationalized its response through enforcement infrastructure, threat disclosures, and public attribution.

Platform Response

Meta's response to Doppelgänger represents one of its most structured and high-profile interventions in adversarial threat disruption. Beyond removing assets tied to the campaign, Meta employed its internal CIB policy and threat intelligence infrastructure to issue public disclosures that shaped discourse on platform security, digital sovereignty, and foreign influence (Gleicher, 2019).

In its Quarterly Adversarial Threat Report for Q4 2022, Meta publicly acknowledged the Doppelgänger campaign, describing it as involving "elaborate website spoofing, advertising spend, and cross-platform amplification" (Nimmo, Franklin, et al., 2023). The report reiterated tactics already detailed to spread false narratives across the European information ecosystem. In

December 2022, Meta formally attributed the campaign to two Russian entities. As previously mentioned, the Social Design Agency and Structura National Technologies, both sanctioned earlier that year by the EU for conducting a "digital information manipulation campaign" (Nimmo & Torrey, 2022). According to Meta, these actors had run a network of websites impersonating news outlets and government agencies to post fake articles promoting pro-Russia narratives since at least May 2022 (Meta, 2023).

The campaign evolved through 2023 and 2024, adapting its narratives and targeting strategies. In the lead-up to the European Parliament elections. Doppelgänger intensified its activity using seven inauthentic websites impersonating legitimate European media entities to influence public opinion in Germany, France, Italy, and Poland. According to the EEAS Technical Report (June 2024), the fake domains were registered between January and May 2024 (EU External Action, 2024), reflecting both operational adaptability and strategic foresight typical of state-aligned disinformation. Meta reported that the campaign spent approximately $105,000 on ads, paid primarily in euros and US dollars (Bouchaud & L, 2024; Nimmo & Torrey, 2022). These funds were used to amplify disinformation targeting NATO, Ukraine, and Western sanctions (Goujard, 2024).

Meta's enforcement did not rely on a single takedown event but followed a recurring pattern of detection, disruption, and attribution. As Doppelgänger resurfaced with new domains and tactics, Meta updated its findings and released new data to researchers and the public. These actions often aligned with signals from OSINT groups, threat intelligence firms, and EU cyber defense assessments, underscoring Meta's growing role as both an enforcement mechanism and a public-facing security actor in Europe's response ecosystem. From the outset, Meta's CIB team framed

Doppelgänger as a persistent threat actor rather than a mere policy violation. Terms such as coordinated influence operations, foreign adversarial campaigns, and multi-language cognitive disruption revealed a convergence between platform and security-sector lexicons (Meta, 2023; Ronzaud et al., 2023)

This rhetorical alignment was matched by a shift in format. Meta's Threat Reports increasingly resembled intelligence briefings, featuring digital infrastructure maps, persona breakdowns, and targeting strategies. The 2022 report's Threat Indicators section outlined technical markers such as shared hosting metadata and tracking codes (Nimmo & Torrey, 2022), aimed at both public and policymaking audiences. Meta's platform governance model was central to its responsiveness. Its multilingual Threat Intelligence Team, coupled with a structured CIB framework and partnerships with actors like Graphika, Recorded Future, and DFRLab, enabled rapid threat classification, attribution, and enforcement (DFRLab, 2022).

The Ad Transparency Portal and public ad library allowed Meta to trace and document the campaign's sponsored content. It also used domain blacklisting to block redirects from fake news websites, frequently coordinating these interventions with EU-based registrars and hosting providers (Sekoia TDR, 2024; EU External Action, 2024). This combination of technical capacity, consistent policy, and external coordination enabled Meta to respond effectively to Doppelgänger. However, as later sections will explore, this response also raised transparency and accountability concerns. While Meta's enforcement was robust, its full implications can only be understood concerning its coordination with NATO and EU actors. The next section evaluates the extent and nature of that collaboration.

Intelligence-Platform Coordination

Although much of Meta's coordination with NATO-affiliated intelligence entities during Operation Doppelgänger occurred behind closed doors, open-source evidence points strongly to structured collaboration. This includes mirrored attribution timelines, coordinated disclosures, consistent language in public reports, and convergence in how both state and platform actors framed the threat.

The clearest indicator is the alignment of threat announcements. Meta's 2023 Quarterly Threat Report detailed Doppelgänger-linked assets with tactical breakdowns (Nimmo, Gleicher, et al., 2023). Around the same time, NATO StratCom COE published similar findings, citing overlapping technical indicators such as IP clusters and DNS data. In June 2024, the EEAS released a technical report that echoed Meta's earlier conclusions, describing Doppelgänger as a malign influence operation rooted in Russian interests (EU External Action, 2024). That three separate actors released nearly identical analyses within a short time frame of one another suggests coordinated assessment, if not direct information exchange.

This alignment extended beyond timing. Meta described Doppelgänger as coordinated inauthentic activity targeting geopolitical narratives across EU member states, mirroring NATO StratCom's description of transnational cognitive operations (Sessa and Miguel, 2024). Both identified the same core themes: economic insecurity, war fatigue, NATO skepticism, and anti-refugee sentiment, framing them as strategic efforts to weaken the democratic structure. Each also emphasized that the campaign was cross-platform, including Twitter, Telegram, and smaller networks, implying intelligence sharing that went beyond Meta's internal capabilities (Ronzaud et al., 2023).

Further coordination is evidenced in operational actions. According to EU DisinfoLab and Recorded Future, multiple Doppelgänger domains were suspended by EU-based registrars within 24 to 48 hours of Meta's takedowns (Insikt Group, 2023). Cybersecurity firms Qurium and Sekoia reported that enforcement timelines regularly coincided with alerts from European governments (Qurium, 2022; Sekoia TDR, 2024). These patterns suggest a coordinated response loop between platform moderation and state-level cyber operations.

Pre-existing institutional ties between Meta and NATO-linked actors facilitated this responsiveness. Since 2019, Meta has participated in forums involving law enforcement, intelligence services, and other platforms, co-authoring research and engaging in StratCom policy discussions (Nimmo, Franklin, et al., 2023; EEAS, 2019). These forums function as informal governance hubs, enabling fast information exchange and coordinated threat mitigation. Meta retained both the institutional memory and external trust to serve as a consistent security actor.

Platforms like Meta have also internalized securitized threat framings, treating disinformation as an existential risk, in line with NATO and EU policy logic (Bernhard, 2023; Viginum, 2023). Meta's use of terms like hybrid warfare and cross-border influence in its threat bulletins reflects this alignment (Nimmo, Gleicher, et al., 2023; EEAS, 2025). While Meta avoided explicitly naming NATO or intelligence partners, this strategic ambiguity functions as a form of signaling within intelligence-sharing regimes constrained by legal and diplomatic sensitivities.

Operation Doppelgänger reflects more than reactive content moderation. It demonstrates an evolving model of collaborative security governance in which attribution, narrative framing, and operational response are increasingly synchronized between platforms and state-linked intelligence actors (Justice Gov, 2024; USCYBERCOM Public Affairs, 2024). To move beyond

descriptive alignment and assess the actual depth of cooperation, the next section applies the thematic analysis that was formed in Chapter One.

Thematic Analysis of Collaboration

This section assesses the type, degree, and institutional framework of cooperation between Meta and NATO-affiliated intelligence actors during the Doppelgänger campaign. Through thematic coding of government statements, public reports, and third-party investigations, this analysis finds that Meta's response reflects high levels of strategic alignment and operational integration, indicative of Level 4 collaboration on the typology.

Attribution synchronicity, or the timing and consistency with which state actors and platforms publicly attribute campaigns, provides strong evidence of intelligence sharing. Meta's 2023 report identified Doppelgänger as a campaign originating in Russia, involving cloned websites and fabricated personas (Nimmo, Gleicher, et al., 2023). The EEAS and EU DisinfoLab published concurrent assessments affirming Russian state alignment, referencing the same technical markers: overlapping IP addresses, domain registries, and ad spending patterns (Alaphilippe et al., 2022; EU External Action, 2024). Ansell and Gash characterize such co-produced threat assessments as joint problem definition, a form of reciprocal consultation common in networked governance (Ansell and Gash, 2008).

Both Meta and NATO-linked actors framed Doppelgänger not merely as a platform violation, but as a systemic threat to democratic resilience. NATO referred to the campaign as "cognitive warfare" and "hybrid aggression," while Meta adopted similar terminology, describing it as an "adversarial foreign influence operation" involving "cross-border coordinated inauthentic

behavior" (NATO StratCom COE, 2023; Nimmo and Torrey, 2022). This convergence illustrates successful securitization, in which actors construct a shared threat narrative to legitimize exceptional actions (Buzan et al., 1998). By invoking national security rhetoric, Meta positioned itself as a legitimate actor in the democratic defense space without bearing the full weight of state accountability.

Operational coordination further supports these conclusions. Following Meta's platform restrictions, several Doppelgänger-linked domains were deregistered within 24 to 48 hours. Hosting companies in Germany and France acted quickly, likely prompted by upstream alerts from national cyber defense authorities (Sekoia TDR, 2024). The United Kingdom's public designations of sanctioned Russian accounts were closely aligned with Meta's ad bans (Office, 2024). These near-simultaneous interventions reflect synchronized operational rhythms, one of the strongest indicators of integrated response capacity.

Institutional continuity also shaped collaboration. Meta's sustained participation in NATO StratCom workshops and the EU Internet Forum fostered shared procedural knowledge. Notably, the Swedish Psychological Defence Agency's 2025 report on Russian information warfare drew directly on Meta's proprietary data to track Doppelgänger's evolution (Pamment and Tsurtsumia, 2025). Despite the operation's adaptive tactics, Meta's CIB team maintained a coherent and proactive reporting strategy, offering threat maps, actor archetypes, and multi-quarter updates (Nimmo and Torrey, 2022). These patterns reflect the institutionalization of platform-intelligence partnerships across transatlantic security architectures. As Sørensen and Torfing argue, interactive political coordination is shaped by repeated engagement rather than formal hierarchies (Sørensen and Torfing, 2005). These informal governance channels likely enabled

Meta to sustain a strategic tempo, align its language with NATO framing, and contribute to the integrated disruption effort.

Taken together, Operation Doppelgänger illustrates the emergence of collaborative governance in hybrid threat environments. The convergence of attribution, framing, operational timing, and institutional embedding meets the criteria for Level 4 collaboration: a structured, cross-sectoral security response involving shared decision-making and institutionalized cooperation between Meta and NATO-affiliated actors. The upcoming section will dive deeper into what makes this a Level 4 case.

Collaboration Typology Assessment

This section will follow the same structure as that of Chapter Three. It will use the typology created in Chapter One to evaluate Meta's engagement with intelligence actors during Operation Doppelgänger. Which uses indicators such as information sharing, operational coordination, public framing, and institutional embedding. Drawing on OSINT sources, NATO and EEAS documentation, investigative reports, and Meta's threat disclosures, this case aligns with Level 4: Integrated Response.

The case demonstrates how collaborative mechanisms became systematically embedded. As mentioned before, Meta's threat attributions closely mirrored those of EEAS and NATO StratCom COE, indicating co-produced assessments and alignment before public disclosure (EU External Action, 2024; NATO StratCom COE, 2023). Takedowns of CIB on Meta's platforms often occurred within 24 to 48 hours of domain-level actions across EU member states, suggesting synchronized operational timelines (Sekoia TDR, 2024). Consistent framing across

Meta and NATO materials using terms like “foreign adversarial behavior” and “threats to democratic infrastructure” underscores narrative convergence (Sessa and Miguel, 2024). Meta’s actions extended beyond content moderation to include ad bans, domain blocking, cross-platform tracking, and alignment with public sector communications.

Meta’s structured response illustrates a form of hybrid threat management beyond what is observed in lower-tier collaborations. Real-time intelligence exchange, coordinated countermeasures, and narrative alignment go beyond the fragmented interventions of Level 2 or the unilateral state-led efforts of Level 3 (Ansell and Gash, 2008). Unlike Level 1 symbolic cooperation, the Doppelgänger case demonstrates tangible operational fusion of platform capabilities with intelligence-driven objectives.

This case highlights how high-threat perception and aligned interests can enable platforms to engage in near-institutionalized coordination with public security actors (Buzan et al., 1998). However, while effective, this depth of integration raises concerns about privatized security, transparency, and democratic accountability. As Sørensen and Torfing (2005) argue, such arrangements create “gray zones” of governance that blur lines of responsibility. The Doppelgänger response captures both the strengths and risks of advanced public-private collaboration. Having classified the response as a Level 4 collaboration, the next section turns to the institutional and political conditions that enabled this high degree of coordination.

Factors Influencing Collaboration

The institutional and strategic context previously described shaped the depth and structure of collaboration in this case. Drawing on the conceptual framework in Chapter One, five main factors explain the outcome: securitization, platform governance, state leverage, trust networks, and accountability regimes.

The most critical enabler was the high level of securitization surrounding Russian disinformation. To justify extraordinary actions, public-private data sharing, mass takedowns, and geopolitical attribution, NATO and EU institutions framed information manipulation as an existential threat to democratic stability (Buzan et al., 1998; Orenstein, 2025). Meta echoed this in its threat reporting (Nimmo, Gleicher, et al., 2023), internalizing the securitization logic and aligning platform enforcement with national security imperatives. Second, Meta's **i**nternal governance capacity was a key differentiator. Its well-resourced CIB regime, experienced in multilingual attribution and state-linked campaigns, enabled rapid escalation and disclosure, unlike Twitter, which faced rollbacks in trust and safety (Hickey et al., 2025; Pamment & Tsurtsumia, 2025). This institutional stability also facilitated seamless coordination with actors like NATO StratCom and the EU.

Third, state leverage**,** especially via the EU's DSA (European Commission, 2023), exerted regulatory pressure and aligned Meta with NATO's strategic goals. Regulatory embeddedness across EU jurisdictions (Kausche & Weiss, 2024) created both legal obligations and incentives for cooperation. Fourth, trust networks developed through multistakeholder forums (e.g., EU Internet Forum, NATO StratCom COE) since 2019 provided channels for real-time coordination

(EEAS, 2019; Sessa & Miguel, 2024). These informal infrastructures enabled cross-mandate collaboration toward shared security outcomes (Emerson et al., 2012).

Finally, accountability regimes such as Meta's Ad Library, transparency reports, and civil society engagement helped maintain legitimacy (Alaphilippe et al., 2022). By framing its actions as democratic corporate responsibility (Gillespie, 2018), Meta balanced cooperation with the state and platform autonomy. Together, these factors, securitization, governance capacity, regulatory alignment, trust infrastructure, and public accountability, explain the level 4 Integrated Response achieved in this case.

Summary

Bringing together the findings across background, platform actions, coordination, and theoretical assessment, this chapter has shown the most advanced form of intelligence-platform collaboration observed during the Russia-Ukraine conflict. Meta's efforts to detect, disrupt, and attribute the operation were closely aligned with NATO-affiliated actors, constituting a Level 4: Integrated Response, the highest tier on the typology outlined in Chapter One. Unlike routine content moderation, Meta's approach reflected strategic convergence: it provided enforcement infrastructure and narrative reach, while intelligence agencies contributed geopolitical insight and legitimacy. From a networked governance perspective, this collaboration was characterized by distributed authority, cooperative framing, and informal communication flows. Meta operated as a semi-autonomous security actor within a broader crisis governance system, rather than simply responding to state directives.

Securitization theory explains how this partnership became politically and institutionally viable. By framing Doppelgänger as an existential threat to democratic integrity, Meta and NATO actors justified mass takedowns, geopolitical attribution, and transnational coordination, actions otherwise difficult to defend under standard regulatory norms. In contrast, the Twitter-Ghostwriter case involved lower coordination, hindered by internal disruption and weakened trust networks. This divergence underscores how institutional capacity, regulatory context, and discursive alignment shape collaboration intensity. Having established Meta's structured, high-level engagement during Operation Doppelgänger, the next chapter compares both cases to assess how differences in governance, threat framing, and trust contributed to distinct outcomes. It draws broader conclusions about the conditions enabling or limiting public-private coordination in hybrid threat environments.

Chapter Five: Comparative Analysis and Discussion

This chapter addresses the dissertation's central research question: what factors drove collaboration between NATO-affiliated intelligence agencies and the social media platforms Meta and Twitter during the 2022-2024 escalation of the Russia-Ukraine conflict, and what unintended consequences emerged from these partnerships? Building on the empirical case studies presented in Chapters Three and Four, which analyzed Twitter's response to the Ghostwriter campaign and Meta's engagement during Operation Doppelgänger, respectively, this chapter provides a comparative analysis of the two platforms' coordination models.

The aim is threefold: first, to assess the relative effectiveness of each partnership using the five-level collaboration typology introduced in Chapter One; second, to interpret how these differences reflect broader patterns of securitization and networked governance; and third, to evaluate the implications of such cooperation for democratic accountability and digital security. While both platforms confronted similar disinformation threats, their internal governance capacity, regulatory exposure, and institutional ties differed sharply.

The chapter proceeds in six sections. It begins by evaluating the effectiveness of Meta and Twitter's collaborations. It then explores key trade-offs between security and civil liberties. Next, it analyzes the explanatory power of securitization and networked governance frameworks before situating the findings within the broader scholarly literature. The final two sections assess the normative and policy implications of hybrid collaboration and conclude with a synthesis of the core insights.

The empirical findings presented in Chapters 3 and 4 reveal differences in the quality and structure of collaboration between Meta and Twitter and NATO-affiliated intelligence agencies. Chapter Four showed that Meta's response to Operation Doppelgänger constituted a Level 4 Integrated Response, marked by synchronized takedowns, co-produced threat framing, and sustained institutional linkages. In contrast, Chapter Three demonstrated that Twitter's engagement during the Ghostwriter campaign corresponded to a Level 3 Strategic Partnership, characterized by informal, reactive coordination and eroded internal stability. The next section directly compares the two cases to assess how internal platform dynamics and external governance structures shaped the effectiveness of collaboration.

Comparative Evaluation of Collaboration Effectiveness

Looking at the five-level collaboration typology, Twitter's coordination during the Ghostwriter campaign corresponded to a Level 3 Strategic Partnership, characterized by informal yet sustained engagement. In contrast, Meta's response to Operation Doppelgänger demonstrated a Level 4 Integrated Response, involving formalized, institutional collaboration.

Meta's partnership featured real-time coordination, co-authored attribution reports, and consistent use of security-sector language terms. The platform actively coordinated with European intelligence units and NATO-linked institutions, publishing synchronized takedowns and intelligence-aligned bulletins. These activities reflected high institutional alignment, embedded trust networks, and a shared strategic framing of disinformation. Critically, Meta's internal architecture, especially the CIB unit, enabled rapid integration of external intelligence into enforcement operations, reinforcing the durability of the partnership.

In contrast, Twitter's response was reactive and episodic. While it undertook meaningful enforcement, such as account removals and labeling, its coordination with intelligence actors remained ad hoc, lacking formal structures or sustained procedural linkages. Despite some convergence in timing and language with EU/NATO sources, it fell short of joint attributions or institutionalized collaboration. After Musk's acquisition, instability, staff reductions, and diminished transparency eroded continuity in platform governance, weakening the reliability of external partnerships.

Collaboration effectiveness can be assessed across several dimensions. Both platforms successfully disrupted disinformation networks, but Meta displayed greater resilience, sustaining consistent coordination and policy implementation over time. Public legitimacy was also stronger in Meta's case, due to regular threat reports and visible disclosures, while Twitter's inconsistent communication weakened external trust.

These comparative findings suggest that collaboration effectiveness is shaped not only by external securitizing pressures but also by internal platform capacity, the strength of trust networks, and the institutionalization of governance systems. While Twitter demonstrated initial engagement, its lack of formal structures and internal unpredictability limited the depth and sustainability of its counter-disinformation strategy. For scholars, this reinforces calls to move beyond threat-centric models by incorporating organisational resilience into theories of public-private security cooperation. For policymakers, it signals that strengthening a platform's internal governance and trust-building mechanisms may be as crucial as external pressure in sustaining effective, long-term alliances against disinformation.

Dimension	Twitter / Ghostwriter	Meta / Doppelgänger
Collaboration Level	Level 3 - Strategic Partnership	Level 4 - Integrated Response
Attribution	GRU (informal, inferred)	Social Design Agency, etc. (publicly stated)
Securitization Intensity	Moderate	High
Platform Stability	Low (post-acquisition turbulence)	High (structured CIB unit)
Operational Coordination	Informal	Synchronized takedowns, shared reports
Regulatory Leverage	Limited	Strong (DSA)
Trust Networks	Partial	Embedded via EU/NATO forums

This table illustrates how Meta's embedded structures and strategic alignment produced a more durable and institutionally integrated response, whereas Twitter's approach constrained its collaboration depth and effectiveness.

While these operational differences offer valuable insights, they also expose deeper tensions between competing values, particularly when security objectives begin to clash with democratic principles. The next section explores these normative trade-offs.

Tensions and Trade-Offs

While Meta and Twitter both responded to disinformation, their approaches exposed deeper tensions between security imperatives and democratic principles. While Meta's alignment with NATO-aligned actors improved attribution speed and strategic coordination, it also intensified concerns over opaque data practices and the creeping normalization of private-sector surveillance under a security mandate. Such actions, though effective, can circumvent traditional oversight mechanisms and diminish user trust in democratic governance frameworks.

Enforcement actions targeting CIB frequently swept up gray-zone actors, satirists, anonymous dissidents, and fringe commentators. This blurred the line between hostile influence operations and protected expression. When platforms rely on intelligence-derived frameworks without transparent criteria or procedural safeguards, they risk reproducing state biases and constraining public discourse.

By contrast, Twitter's looser coordination avoided direct state alignment but came at the expense of consistency and reliability. The platform's moderation volatility post-2022 acquisition highlights the risks of governance instability. As Douek (2020) and Keller (2019) warn, platforms that act as security actors without institutional accountability may accelerate democratic erosion instead of preventing it.

These results suggest that security effectiveness alone cannot justify public-private partnerships. Their legitimacy depends equally on how well they preserve transparency, civil liberties, and pluralistic information ecosystems. These challenges underscore the need for clear procedural safeguards and multistakeholder oversight mechanisms. Without such checks, the rapid escalation of platform-led enforcement risks becoming a norm in crisis governance, one that circumvents democratic deliberation and entrenches asymmetries of power. The next section turns to the theoretical frameworks introduced in Chapter One, where it will take a closer look at how the frameworks compare.

Theoretical Insights Compared

As previously mentioned, securitization theory explains how disinformation became framed as an existential threat, justifying extraordinary platform interventions. Meta's adoption of terms like cognitive warfare reflected strategic alignment with NATO narratives, legitimizing unprecedented takedowns and data sharing. This securitizing move succeeded through mutual reinforcement between state and platform actors, which the Copenhagen School terms "audience acceptance" (Buzan et al., 1998).

However, securitization requires operational capacity to translate rhetoric into action. Twitter's failure to institutionalize its Ghostwriter response, despite similar NATO framing, demonstrates that discursive alignment alone is insufficient. Without robust internal governance, securitized narratives remain hollow. Networked governance helps explain why Meta could implement securitized goals effectively. Its embedded ties with EU and NATO actors enabled operational follow-through. By contrast, Twitter's internal disruption severed trust-based networks, revealing the fragility of informal governance under organizational instability. Collaboration depends on relational factors, institutional memory, mutual dependence, and embedded trust.

Together, these frameworks show that effective cooperation requires both narrative legitimacy and operational capability. Meta succeeded because securitization provided urgency while networked trust enabled execution. Twitter failed when problems within the platform undermined both elements, revealing that partnership sustainability depends on institutional coherence as much as shared threat perceptions. Having established how these theoretical lenses explain the observed collaboration patterns, the following section examines how these findings

contribute to existing scholarship on platform governance and public-private security partnerships.

Key Findings

This study offers four key contributions to the literature on platform governance and public-private security cooperation. First, it reinforces theories of semi-sovereign governance. Meta's coordination with NATO-aligned actors during Operation Doppelgänger illustrates how platforms can evolve into co-producers of security narratives rather than mere content moderators (Gorwa, 2019; Keller, 2019). By adopting similar language, Meta absorbed and operationalized state securitization discourses. This supports Ansell and Gash's (2008) model of networked governance, which argues that trust-based horizontal coordination can substitute for formal institutional hierarchies during crises.

Second, Twitter's case problematizes linear assumptions about platform institutional development. Scholars like Gillespie (2018) and Klonick (2018) suggest that platforms trend toward predictable, rule-based governance. However, Twitter's governance collapse, post-acquisition challenges, which include elite-driven decisions and weakened trust networks, rapidly eroded enforcement standards. These discoveries align with warnings by Douek (2020) and Keller (2019), who argue that collaboration depends not only on threat alignment but also on stable internal infrastructure.

Third, both cases expose enduring accountability gaps in informal intelligence-platform partnerships. Despite Meta's relative success in disrupting disinformation, its actions lacked

transparency regarding enforcement criteria, legal justification, and intelligence sourcing. This supports the critiques of Kaye (2019) and Bradshaw et al. (2011), who caution that such opaque arrangements risk circumventing judicial oversight and undermining civil liberties.

Conclusively, this dissertation offers a methodological innovation. The five-level collaboration typology provides a scalable framework for assessing public-private cooperation across different contexts. It allows researchers and practitioners to distinguish between rhetorical alignment and operational integration with greater analytical precision.

Together, these contributions advance theoretical, empirical, and methodological understanding of digital security governance in an increasingly hybrid and securitized information environment. Building on these insights, the following section explores the broader implications of public-private partnerships for democratic accountability, international norms, and the future architecture of digital governance.

Broader Implications

This comparative study revealed important implications for platform accountability, international norm-building, and democratic governance. Meta's response illustrates how platforms can adopt quasi-security functions when state institutions fall short. Its close alignment with NATO frameworks reflects a model of delegated governance, where enforcement authority and geopolitical signaling blur the boundaries between public and private actors. While this enabled effective action, it also entrenched opaque decision-making processes that evade democratic oversight.

From a global perspective, Meta's integration into Euro-Atlantic security structures reinforces Western information security norms while marginalizing alternative models, particularly in the Global South. As NATO and EU agendas shape enforcement standards, concerns around digital colonialism and elite-driven governance persist. Governments in the Global South are often excluded from shaping platform security protocols, leading to asymmetrical enforcement and the imposition of Western threat narratives that may not reflect local priorities, undermining efforts toward pluralistic and context-sensitive internet governance.

At the platform level, both cases raise doubts about whether security cooperation can coexist with transparency and due process. Meta offered more visibility than Twitter, but both lacked clear public standards on attribution, enforcement rationale, and data-sharing. Without institutional safeguards, deeper collaboration may erode public trust rather than strengthen it.

These conclusions should be understood considering several limitations. The study relied on open-source data, limiting access to internal decision-making and classified intelligence exchanges. Its focus on Western platforms restricts generalizability to non-NATO or authoritarian contexts. Additionally, while the five-level typology offered analytical clarity, real-world collaboration often defies fixed categories, particularly under crisis conditions.

Building on these constraints, the following section offers targeted recommendations to help policymakers, platforms, and researchers navigate the evolving challenges of digital security governance.

Recommendations and Future Research

The outcomes support several practical and scholarly recommendations. For policymakers and platforms, three priorities stand out. First, transparency protocols should be institutionalized: platforms must publish clear threat attribution logs and enforcement rationales when cooperating with security actors. Second, NATO and EU regulators should define legal and normative boundaries for such collaboration to ensure accountability without hindering crisis response. Third, independent oversight bodies should audit platform-intelligence partnerships, particularly in cases involving mass content takedowns, to safeguard civil liberties.

For researchers, three avenues merit further attention. Comparative studies should extend beyond NATO contexts to include alliances such as ASEAN (Association of Southeast Asian Nations) or the African Union, testing the typology's broader relevance. Longitudinal research could assess whether crisis collaborations endure or dissolve once threats subside. Finally, studies should prioritize interviews with platform officials, policymakers, and security practitioners to better understand internal coordination mechanisms and validate inferences drawn from public data.

Summary

This chapter has comparatively analyzed Meta's and Twitter's responses to Russian disinformation through empirical evidence and theoretical framing. Meta's high-level coordination reflected successful securitization and mature governance networks, while Twitter's fragmented engagement underscored the fragility of informal, unstructured collaboration.

The findings reaffirm that collaboration is not automatic but contingent, emerging through a confluence of platform stability, shared threat narratives, and institutional trust. Without these, even securitized environments can fail to produce effective coordination. The contrast between Meta and Twitter further reveals the normative and institutional stakes of hybrid security arrangements, particularly the tension between operational effectiveness and democratic accountability.

Ultimately, this chapter has demonstrated that the existence of shared infrastructures rather than shared adversaries shapes collaboration in the information domain. When institutional capacity and policy coherence coincide, securitization and networked governance can support each other, as demonstrated by Meta's integrated response. On the other hand, when discursive alignment is not supported by internal governance or strategic continuity, Twitter exposed the vulnerability of informal cooperation. In the future, the viability of these collaborations will rely on platforms' readiness to serve as long-term co-governors in the field of digital security, as well as who they collaborate with. With the empirical analysis complete, the final chapter now synthesizes these insights, reflecting on the theoretical contributions, practical implications, and future research directions stemming from this study.

Conclusion

This dissertation has investigated the collaboration between NATO-affiliated intelligence agencies and major social media platforms, demonstrating that it is no longer hypothetical; it is operational, strategic, and increasingly central to how democracies respond to disinformation threats. Through a dual-theoretical lens of securitization theory and networked governance, the study examined how Meta and Twitter engaged with NATO-aligned intelligence actors during the 2022-2024 escalation of the Russia-Ukraine war. The analysis focused on two case studies: Meta's response to Operation Doppelgänger and Twitter's involvement in the Ghostwriter campaign, utilizing a five-level collaboration typology to assess variations in depth, formality, and institutionalization of these partnerships.

Three core contributions emerged. First, a five-level collaboration typology has been introduced and operationalized, providing researchers and policymakers with a conceptual tool to evaluate platform-intelligence coordination along a spectrum, from integrated response to symbolic alignment, rather than as a binary. This typology improves empirical clarity and makes structured comparison across cases and contexts easier in a policy space characterized by informal arrangements and fragmented disclosures.

Second, the study reveals that internal platform governance is as crucial to collaboration outcomes as external threat intensity. Meta's structured internal capacity, predominantly the CIB unit, enabled high-level cooperation marked by co-produced attributions, real-time enforcement, and narrative alignment. By contrast, Twitter's internal instability and weakened trust and safety functions after its 2022 acquisition undermined its ability to sustain meaningful engagement, despite exposure to similar disinformation threats. This finding challenges the assumption that

securitized environments automatically produce deeper public-private coordination. Rather, collaboration is contingent on institutional readiness and mutual trust.

Third, the dissertation highlights the paradox of securitization. While framing disinformation as an existential threat enabled exceptional measures, mass takedowns, intelligence sharing, and policy alignment, it also raised significant risks. These include blurred lines between platform governance and state surveillance, the erosion of democratic oversight, and the suppression of marginal or dissenting voices under broad definitions of foreign adversarial behavior. The study thus positions securitization not only as a strategic enabler but as a normative dilemma.

These arguments were developed across five chapters. Chapter One situated the project within the literature on platform governance, hybrid warfare, and intelligence-policy coordination, identifying a gap in empirical understanding of real-time cooperation. Chapter Two outlined a qualitative, comparative methodology combining process tracing with thematic analysis of open-source materials. Chapter Three examined Twitter's response to Ghostwriter, identifying a Level 3 Strategic Partnership hindered by governance instability and ad hoc threat framing. Chapter Four assessed Meta's response to Operation Doppelgänger, finding a Level 4 Integrated Response characterized by synchronized enforcement, co-produced attributions, and embedded institutional linkages. Chapter Five shows that collaboration intensity and legitimacy hinge on platform capacity, regulatory leverage, and discursive convergence. Effective partnerships arise when operational resources, legal authority, and aligned framing reinforce one another, enabling sustained coordination.

The dissertation has limitations. Its reliance on open-source data prevented access to internal deliberations or classified communications, and its focus on the NATO context limits generalizability to non-Western or authoritarian regimes. However, these limitations suggest

valuable directions for future inquiry. Comparative research could apply the typology in regions where platform-state relationships follow different logics, such as India's evolving digital sovereignty model, ASEAN's cyber defense coordination, or South Africa's hybrid information control frameworks. These cases could test whether public-private cooperation is a uniquely Western strategy or a more global phenomenon with context-specific variations.

Longitudinal studies could explore the durability of such partnerships over time: do crisis-driven collaborations endure, or do they dissolve once threats recede? In addition, greater attention should be paid to the mechanisms of coordination, who authorizes cooperation, under what institutional conditions, and with what democratic safeguards? Interview-based research could help unpack the informal channels, trust networks, and political calculations that shape these dynamics behind the scenes. Future research should also investigate the normative consequences of security co-production: when platforms function as de facto intelligence actors, what happens to accountability, legitimacy, and pluralism?

Beyond the NATO-Russia context, this dual-theoretical framework offers a transferable lens for examining public-private security arrangements elsewhere. It could be applied to the examples listed earlier in this conclusion. While institutional logics and threat perceptions vary, these cases also involve forms of security co-production shaped by trust, discretion, and narrative alignment. Applying this framework comparatively can reveal whether NATO-style collaboration reflects a broader model or one context-specific pathway among many. In doing so, it would deepen understanding of how digital security governance adapts to local conditions while confronting shared challenges in the hybrid threat environment.

Theoretically, this study demonstrates the value of combining securitization theory with networked governance. Securitization explains how digital threats are constructed as exceptional and urgent, legitimizing extraordinary responses. Networked governance reveals how coordination unfolds horizontally, through informal, trust-based, iterative processes rather than hierarchical command. When applied together, these frameworks highlight both the discursive and institutional foundations of collaboration. This dual approach allows scholars to move beyond simplistic state-platform binaries and toward a more nuanced understanding of how security is co-produced in digitally mediated environments.

Ultimately, the findings suggest that intelligence-platform partnerships represent not just technical adaptations to new threat landscapes, but a broader transformation in democratic governance. As hybrid warfare increasingly targets the information domain, states are turning to private platforms not only for content enforcement but also for geopolitical signaling, attribution legitimacy, and public trust management. Platforms, in turn, are assuming quasi-sovereign roles, issuing threat assessments, enforcing norms, and co-constructing security narratives. This entanglement raises pressing questions about oversight, legitimacy, and the democratic costs of delegating core security functions to private actors.

Platforms will play a more strategic but contested role as hybrid warfare frontlines shift farther into the information sphere. Not only how they react, but also who establishes the conditions and for what purpose, is the question. The pressing challenge is no longer whether public-private collaboration should exist; it does, but how to institutionalize it in ways that are transparent, accountable, and democratically legitimate. Striking this balance will determine whether digital security governance safeguards democracy or inadvertently undermines it.

Bibliography

Aguerri, J. C., Santisteban, M., & Miró-Llinares, F. (2024). The fight against disinformation and its consequences: measuring the impact of "Russia state-affiliated media" on Twitter. *Crime Science*, *13*(1). https://doi.org/10.1186/s40163-024-00215-9

Alaphilippe, A., Machado, G., Miguel, R., & Poldi, F. (2022, September 27). *Doppelganger - Media clones serving Russian propaganda*. EU DisinfoLab. https://www.disinfo.eu/wp-content/uploads/2022/09/Doppelganger-1.pdf

Aldrich, R. J. (2009). US–European Intelligence Co-Operation on Counter-Terrorism: Low Politics and Compulsion. *The British Journal of Politics and International Relations*, *11*(1), 122–139. https://doi.org/10.1111/j.1467-856x.2008.00353.x

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, *60*(2), 270–290. https://doi.org/10.1016/j.jml.2008.10.001

Ansell, C., & Gash, A. (2008). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory*, *18*(4), 543–571. https://doi.org/10.1093/jopart/mum032

Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Society*, *54*(2), 138–149. https://doi.org/10.1007/s12115-017-0114-0

Bernhard, M. (2023, July 12). *Politicians of Germany's far-right AfD are benefitting from a pro-Russian propaganda campaign. Here is how.* Correctiv.org. https://correctiv.org/en/fact-checking-en/2023/07/12/politicians-of-germanys-far-right-afd-are-benefitting-from-a-pro-russian-propaganda-campaign-here-is-how/

Bouchaud, P., & L, A. (2024). *Pro-Russian Ads Campaigns Approved by Meta from May 1 to May 27, 2024in Italy, Germany, France& Poland*. https://aiforensics.org/uploads/Meta_Ads_Follow_up_docx_697d78b99f.pdf

Bradshaw, S., & Howard, P. N. (2018). THE GLOBAL ORGANIZATION OF SOCIAL MEDIA DISINFORMATION CAMPAIGNS. *Journal of International Affairs*, *71*(1.5), 23–32. https://www.jstor.org/stable/26508115

Bradshaw, S., Millard, C., & Walden, I. (2011). Contracts for clouds: comparison and analysis of the Terms and Conditions of cloud computing services*. *International Journal of Law and Information Technology*, *19*(3), 187–223. https://doi.org/10.1093/ijlit/ear005

Buzan, B., Wæver, O., & De Wilde, J. (1998). *Security: A New Framework for Analysis*. Lynne Rienner Pub.

Cardiff University. (2023). *Ghostwriter: Investigating a Russian cyber-enabled disinformation campaign*. Cardiff

University.
https://www.cardiff.ac.uk/__data/assets/pdf_file/0005/2699483/Ghostwriter-Report-Final.pdf

Carr, M. (2016). Public-private partnerships in national cyber-security strategies. *International Affairs*, *92*(1), 43–62. https://doi.org/10.1111/1468-2346.12504

Cavelty, M. D., & Egloff, F. J. (2019). The politics of cybersecurity: Balancing different roles of the state. St Antony's International Review, 15(1), 37-57.

Clark, J. R., & Rid, T. (2021). Review of Active Measures: The Secret History of Disinformation and Political Warfare. *American Intelligence Journal*, *38*(1), 185–187. https://www.jstor.org/stable/27087774

Clayton, J. (2022, March 19). How Kremlin accounts manipulate Twitter. *BBC News*. https://www.bbc.co.uk/news/technology-60790821

Conway, M. (2017). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies in Conflict & Terrorism*, *40*(1), 77–98. https://doi.org/10.1080/1057610x.2016.1157408

Council of the EU. (2021, September 24). *Declaration by the High Representative on behalf of the European Union on respect for the EU's democratic processes*. Consilium Europa. 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/

DFRLab. (2022, September 27). *Russia-based Facebook operation targeted Europe with anti-Ukraine messaging*. Medium; DFRLab. https://medium.com/dfrlab/russia-based-facebook-operation-targeted-europe-with-anti-ukraine-messaging-389e32324d4b

Douek, E. (2020). The Rise of Content Cartels. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3572309

EEAS. (2019). RAPID ALERT SYSTEMSTRENGTHENING COORDINATED AND JOINT RESPONSES TO DISINFORMATION. In *European External Action Service*. https://www.eeas.europa.eu/sites/default/files/ras_factsheet_march_2019_0.pdf

EEAS. (2025). *3rd EEAS Report on Foreign Information Manipulation and Interference Threats Exposing the architecture of FIMI operations*. https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf

Emerson, K., Nabatchi, T., & Balogh, S. (2012). An Integrative Framework for Collaborative Governance. *Journal of Public Administration Research and Theory*, *22*(1), 1–29. https://doi.org/10.1093/jopart/mur011

EU External Action. (2024). *Doppelganger strikes back: FIMI activities in the context of the EE24*. https://euvsdisinfo.eu/uploads/2024/06/EEAS-TechnicalReport-DoppelgangerEE24_June2024.pdf

European Commission. (2023, November 3). *The impact of the Digital Services Act on digital platforms*. Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms

European Commission. (2024, April 30). *Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act*. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

Farrell, H., & Newman, A. L. (2019). Weaponized Interdependence: How Global Economic Networks Shape State Coercion. *International Security*, *44*(1), 42–79. https://doi.org/10.1162/isec_a_00351

Frühwirth, L., & Nazari, S. (2024). FOOL ME ONCE: Russian Influence Operation Doppelganger Continues on X and Facebook. In *Alliance4Europe*. https://alliance4europe.eu/wp-content/uploads/2024/09/CDN-Report-%E2%80%93-Fool-Me-Once_-Russian-Influence-Operation-Doppelganger-Continues-on-X-and-Facebook-%E2%80%93-September-2024.pdf

Geissler, D., Bär, D., Pröllochs, N., & Feuerriegel, S. (2023). Russian Propaganda on Social Media during the 2022 Invasion of Ukraine. *EPJ Data Science*, *12*(1). https://doi.org/10.1140/epjds/s13688-023-00414-5

Gielewska, A. (2021). *The Ghostwriter playbook From cyber attacks to disinformation operations in Central Europe*. https://www.disinfo.eu/wp-content/uploads/2021/11/03_Disinfo-for-Hire_Anna.pdf

Giles, K. (2016). *Russia's "New" Tools for Confronting the West Continuity and Innovation in Moscow's Exercise of Power Russia's "New" Tools for Confronting the West*. https://www.chathamhouse.org/sites/default/files/publications/2016-03-russia-new-tools-giles.pdf

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press. https://doi.org/10.12987/9780300235029

Gleicher, N. (2019, October 21). *How We Respond to Inauthentic Behavior on Our Platforms: Policy Update*. About Facebook. https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/

Global Engagement Center. (2024, December 23). *About Us - Global Engagement Center - United States Department of State*. United States Department of State. https://2021-2025.state.gov/about-us-global-engagement-center-2/

Golovchenko, Y., Hartmann, M., & Adler-Nissen, R. (2018). State, Media and Civil Society in the Information Warfare over Ukraine: Citizen Curators of Digital Disinformation. *International Affairs*, *94*(5), 975–994. https://doi.org/10.1093/ia/iiy148

Gorwa, R. (2019). The Platform Governance triangle: Conceptualising the Informal Regulation of Online Content. *Internet Policy Review*, *8*(2), 1–22. https://doi.org/10.14763/2019.2.1407

Goujard, C. (2024, April 17). *Big, bold and unchecked: Russian influence operation thrives on Facebook*. Politico. https://www.politico.eu/article/russia-influence-hackers-social-media-facebok-operation-thriving/

Group, I. (2022, April 28). *Ghostwriter in the Shell: Expanding on Mandiant's Attribution of UNC1151 to Belarus*. Recordedfuture.com; Recorded Future. https://www.recordedfuture.com/research/ghostwriter-in-the-shell

Hakala, J., & Melnychuk, J. (2021, June 11). *StratCom | NATO Strategic Communications Centre of Excellence Riga, Latvia*. Stratcomcoe.org. https://stratcomcoe.org/pdfjs/?file=/publications/download/Nato-Cyber-Report_15-06-2021.pdf?zoom=page-fit

Hickey, D., Fessler, D. M. T., Lerman, K., & Burghardt, K. (2025). X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. *PLOS ONE*, *20*(2), e0313293. https://doi.org/10.1371/journal.pone.0313293

Hoffman, F. (2007). *Conflict in the 21st Century: The Rise of Hybrid Wars*. Potomac Institute for Policy Studies. https://www.potomacinstitute.org/images/stories/publications/potomac_hybridwar_0108.pdf

Insikt Group. (2023). *Obfuscation and AI Content in the Russian Influence Network "Doppelgänger" Signals Evolving Tactics*. https://go.recordedfuture.com/hubfs/reports/ta-2023-1205.pdf

Ischebeck-Baum, F. (2015). Intelligence cooperation and the war on terror: Anglo-American security relations after 9/11. *Defence Studies*, *15*(4), 376–377. https://doi.org/10.1080/14702436.2015.1079403

Iuliia Alieva, Hui, L., & Carley, K. M. (2022). Investigating the Spread of Russian Disinformation about Biolabs in Ukraine on Twitter Using Social Network Analysis.

2022 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata55660.2022.10020223

Iuliia Alieva, Kloo, I., & Carley, K. M. (2024). Analyzing Russia's propaganda tactics on Twitter using mixed methods network analysis and natural language processing: a case study of the 2022 invasion of Ukraine. *EPJ Data Science*, *13*(1). https://doi.org/10.1140/epjds/s13688-024-00479-w

Justice Gov. (2024). *IN THE UNITED STATES DISTRICT COURT FOR THE EASTERN DISTRICT OF PENNSYLVANIA UNITED STATES OF AMERICA v. CERTAIN DOMAINS*. https://www.justice.gov/d9/2024-09/doppelganger_affidavit_9.4.24.pdf

Kahn, G. (2024, December 3). *These reporters wrote a book on Musk's Twitter takeover. Here's what they think is next for journalism and X*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/news/these-reporters-wrote-book-musks-twitter-takeover-heres-what-they-think-next-journalism-and-x

Kausche, K., & Weiss, M. (2024). Platform power and regulatory capture in digital governance. *Business and Politics*, 1–25. https://doi.org/10.1017/bap.2024.33

Kaye, D. (2019). *Speech Police: The Global Struggle to Govern the Internet*. Columbia Global Reports. https://doi.org/10.2307/j.ctv1fx4h8v

Keller, D. (2019). *Who Do You Sue?* https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf

Klijn, E.-H., & Skelcher, C. (2007). DEMOCRACY AND GOVERNANCE NETWORKS: COMPATIBLE OR NOT? *Public Administration*, *85*(3), 587–608. https://doi.org/10.1111/j.1467-9299.2007.00662.x

Klonick, K. (2018). *THE NEW GOVERNORS: THE PEOPLE, RULES, AND PROCESSES GOVERNING ONLINE SPEECH* (pp. 1599–1669). https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf

Krebs, R. R., & Jackson, P. T. (2007). Twisting Tongues and Twisting Arms: The Power of Political Rhetoric. *European Journal of International Relations*, *13*(1), 35–66. https://doi.org/10.1177/1354066107074284

Lynskey, O. (2017). Regulating "Platform Power." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2921021

Mandiant. (2020, July 28). *Ghostwriter Influence Campaign Spreads Fabricated Content* (L. Foster, S. Riddell, D. Mainor, & G. Roncone, Eds.). Google Cloud Blog; Google Cloud. https://cloud.google.com/blog/topics/threat-intelligence/ghostwriter-influence-campaign

McSweeney, S. (2022). *Our ongoing approach to the war in Ukraine*. https://blog.x.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine

Menn, J. (2023, September 1). Musk's new Twitter policies helped spread Russian propaganda, E.U. says. *Washington Post*. https://www.washingtonpost.com/technology/2023/09/01/musk-twitter-x-russia-propaganda/

Meta. (2023). *DRAFT: Adversarial Threat Report*. https://www.politico.eu/wp-content/uploads/2023/08/29/NEAR-FINAL-DRAFT-Meta-Quarterly-Adversarial-Threat-Report-Q2-2023.pdf

Miguel, R., Sessa, M. G., & Alaphilippe, A. (2024, May 24). *Assessing cost-effectiveness: responses to the Doppelganger operation - EU DisinfoLab*. EU DisinfoLab. https://www.disinfo.eu/publications/assessing-cost-effectiveness-responses-to-the-doppelganger-operation/

Murero, M. (2023). Coordinated inauthentic behavior: An innovative manipulation tactic to amplify COVID-19 anti-vaccine communication outreach via social media. *Frontiers in Sociology*, *8*. https://doi.org/10.3389/fsoc.2023.1141416

NATO. (2024, October 24). *NATO-Russia: setting the record straight*. NATO. https://www.nato.int/cps/en/natohq/115204.htm

NATO StratCom COE (2023). Kremlin Communication Strategy for Russian Audiences Before and After the Full-Scale Invasion of Ukraine. NATO Strategic *Communications* Centre of Excellence. https://stratcomcoe.org/pdfjs/?file=/publications/download/Kremlin_Communication_Strategy_DIGITAL.pdf?zoom=page-fit

Nimmo, B., Franklin, M., Agranovich, D., Torrey, M., & Hundley, L. (2023). *DETAILED REPORT Quarterly Adversarial Threat Report*. https://about.fb.com/wp-content/uploads/2023/02/Meta-Quarterly-Adversarial-Threat-Report-Q4-2022.pdf#page=6.21

Nimmo, B., Gleicher, N., Franklin, M., Hundley, L., & Torrey, M. (2023). *Meta's Adversarial Threat Report, Third Quarter 2023*. Meta Transparency. https://scontent-lhr6-2.xx.fbcdn.net/v/t39.8562-6/406961197_3573768156197610_1503341237955279091_n.pdf?_nc_cat=105&ccb=1-7&_nc_sid=b8d81d&_nc_ohc=KC0wfQPMmMEQ7kNvwEwh1x1&_nc_oc=AdkGV2sh-IqYUXaanUjEGs9bhdghOdgXh9a9ESsRKKyvXOJriKyK1LFZWrraMUaiICI&_nc_zt=14&_nc_ht=scontent-lhr6-

2.xx&_nc_gid=bs0ixtNnfFGcvRdLpRQX4Q&oh=00_AfQnG29SgaZUgw1IlUrMZitjgN7Qx3DvqrevUIpw1L6dJA&oe=688FE592#page=6.06

Nimmo, B., & Torrey, M. (2022, September). *Taking down coordinated inauthenticbehavior from Russia and China*. Meta. https://about.fb.com/wp-content/uploads/2022/11/CIB-Report_-China-Russia-Sept-2022.pdf#page=12.19

Nocetti, J. (2015). Contest and conquest: Russia and global internet governance. *International Affairs*, *91*(1), 111–130. https://doi.org/10.1111/1468-2346.12189

Office, D. (2024, October 28). *UK sanctions Putin's interference actors*. GOV.UK. https://www.gov.uk/government/news/uk-sanctions-putins-interference-actors

Orenstein, M. A. (2025). Securitisation of EU policy: how Russia's invasion of Ukraine is changing Europe. *Journal of European Public Policy*, 1–22. https://doi.org/10.1080/13501763.2025.2497350

Pamment, J., & Tsurtsumia, D. (2025, May 12). *Beyond Operation Doppelgänger: A Capability Assessment of the Social Design Agency*. Https://Mpf.se/Psychological-Defence-Agency. https://mpf.se/psychological-defence-agency/publications/archive/2025-05-15-beyond-operation-doppelganger-a-capability-assessment-of-the-social-design-agency

Polyakova, A., Boyer, S., & Bosch, B.-R. (2018). *THE FUTURE OF POLITICAL WARFARE: RUSSIA, THE WEST, AND THE COMING AGE OF GLOBAL DIGITAL COMPETITION THE NEW GEOPOLITICS EUROPE*. https://www.brookings.edu/wp-content/uploads/2018/03/the-future-of-political-warfare.pdf

Pomerantsev, P., & Weiss, M. (2014). *The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money A Special Report presented by The Interpreter, a project of the Institute of Modern Russia*. https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev__The_Menace_of_Unreality.pdf

Press, T. A. (2022, May 19). Twitter aims to crack down on misinformation, including misleading posts about Ukraine. *NPR*. https://www.npr.org/2022/05/19/1100100329/twitter-misinformation-policy-ukraine

Qurium . (2022, September 27). *Under the hood of a Doppelgänger – Qurium Media Foundation*. Qurium.org. https://www.qurium.org/alerts/russia/under-the-hood-of-a-doppelganger/

Ronzaud, L., Carter, J., & Williams, T. (2023). Summit Old, Summit New Russia-Linked Actors Leverage New and Old Tactics in Influence Operations Targeting Online Conversations About NATO Summit. In *https://graphika.com/*. https://public-assets.graphika.com/reports/graphika_report_summit_old_summit_new.pdf

Roslon, D. T., Kruzhkova, E. M., & Syvulka, U. I. (2024). EU AND NATO STRATEGY TO COUNTER AND PREVENT RUSSIAN PROPAGANDA. *Scientific Journal "Regional Studies," 36*, 100–105. https://doi.org/10.32782/2663-6170/2024.36.16

Schuette, C. (2023). Russian Disinformation on NATO Expansion and the War in Ukraine. *Journal of Strategic Security*, *16*(4), 34–56. https://www.jstor.org/stable/48751614

Sekoia TDR. (2024, May 21). *Master of Puppets: Uncovering the DoppelGänger pro-Russian influence campaign* (C. Chavane, A. G., & K. S. Seznec, Eds.). Sekoia.io Blog. https://blog.sekoia.io/master-of-puppets-uncovering-the-doppelganger-pro-russian-influence-campaign/#h-facebook

Sessa , M. G., & Miguel, R. (2024, May 20). *The Doppelganger Case: Assessment of Platform Regulation on the EU Disinformation Environment*. Stratcomcoe.org. https://stratcomcoe.org/pdfjs/?file=/publications/download/The-Doppelganger-Case-DIGITAL.pdf?zoom=page-fit

Sorensen, E., & Torfing, J. (2005). The Democratic Anchorage of Governance Networks. *Scandinavian Political Studies*, *28*(3), 195–218. https://doi.org/10.1111/j.1467-9477.2005.00129.x

Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. Proceedings of the ACM on human-computer interaction, 3(CSCW), 1-26.

Thomas, T. (2014). Russia's Information Warfare Strategy: Can the Nation Cope in Future Conflicts? *The Journal of Slavic Military Studies*, *27*(1), 101–130. https://doi.org/10.1080/13518046.2014.874845

Twitter. (2021, December 2). *Disclosing state-linked information operations we've removed*. X.com. https://blog.x.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed

USCYBERCOM Public Affairs. (2024, September 3). *Russian Disinformation Campaign "DoppelGänger" Unmasked: A Web of Deception*. U.S. Cyber Command. https://www.cybercom.mil/Media/News/Article/3895345/russian-disinformation-campaign-doppelgnger-unmasked-a-web-of-deception/

Valeriano, B., Jensen, B., & Maness, R. C. (2018). Cyber Strategy. In *Oxford Scholarship Online*. Oxford University Press. https://doi.org/10.1093/oso/9780190618094.001.0001

Viginum. (2023, July 19). *RRN: A complex and persistent information manipulation campaign*. General Secretariat for Defence and National Security . https://www.sgdsn.gouv.fr/files/files/20230719_NP_VIGINUM_RAPPORT-CAMPAGNE-RRN_EN1.pdf

Wæver, O. (1995). Securitization and Desecuritization. In R. D. Lipschutz (Ed.), *On Security* (pp. 46-87). Columbia University Press.

Zia, H. B., Haq, E. U., Castro, I., Hui, P., & Tyson, G. (2023, June 20). *An Analysis of Twitter Discourse on the War Between Russia and Ukraine*. ArXiv.org. https://doi.org/10.48550/arXiv.2306.11390