Mediascapes journal 27/2026

Exploring the Zombie Internet: Anatomy of Three Deceptive Information Operations on Facebook*

Giada Marino**
Università degli Studi di Urbino Carlo Bo

Fabio Giglietto***
Università degli Studi di Urbino Carlo Bo

Anwesha Chakraborty****
Independent Researcher

Massimo Terenzi*****
LUISS Guido Carli

Samuel Olaniran******
University of the Witwatersrand

This study advances the concept of deceptive information operations as an analytical framework that moves beyond traditional false-news taxonomies to capture a broader spectrum of manipulative and harmful communicative activities on social media. Drawing on Starbird et al.'s (2019) work on strategic information operations, Chadwick and Stanyer's (2022) theory of deception and META's Coordinated Inauthentic behaviour (Gleicher, 2018), we examine three cases discovered through an automated detection workflow that continuously monitors coordinated sharing behaviour on Facebook: a state-aligned pro-Putin propaganda network, a coordinated campaign promoting online gambling, and a network of poorly moderated public groups distributing adult content and fraudulent links. Each case is analysed through three dimensions: the deceptive nature of the content, the exploitation of platform affordances, and the participatory behaviours that sustain and amplify these operations. We used an inductive grounded coding approach applied to account-level metadata and the 500 most-viewed posts per network retrieved via the Meta Content Library. The findings reveal shared mechanisms across all three operations, e.g. coordinated behaviour to exploit Facebook's algorithmic ranking and simulate organic popularity, and the deliberate mixing of benign and harmful content to obscure intent and evade moderation. These dynamics resonate with the notion of the "Zombie Internet", an information ecosystem in which the boundary between human and automated, toxic behaviour becomes increasingly difficult to maintain. Together, the cases expose structural vulnerabilities in platform governance that allow deceptive operations to persist, scale, and evade detection. The study contributes an empirically grounded conceptualisation of deceptive information operations and identifies critical gaps in platform governance frameworks that focus narrowly on content inauthenticity rather than addressing the structural conditions that enable manipulation at scale.

Keywords: Facebook, strategic information operation, deception, coordinated inauthentic behaviour, gambling, propaganda, adult content

* Article submitted on 26/02/2026. Article accepted on 22/06/2026.
** giada.marino@uniurb.it
*** fabio.giglietto@uniurb.it
**** anweshachakrabortyunibo@gmail.com
***** mterenzi@luiss.it
****** psalmuel35@gmail.com

In recent years, social media platforms have been increasingly flooded with clickbait, ragebait, and deceptive content, including AI-generated imagery, phishing schemes, and manipulative visuals designed to provoke emotional reactions or sway audiences (DiResta & Goldstein, 2024; Donovan et al., 2020). Although taxonomies of problematic information are well-established (Jack, 2017; Wardle & Derakhshan, 2017), these frameworks often fail to capture the full scope and complexity of contemporary information operations (Starbird et al., 2019). These operations extend beyond merely spreading falsehoods; they deliberately pollute the digital information ecosystem by flooding social media with low-credibility content – such as clickbait, divisive memes, and emotionally charged posts – thereby eroding public trust and fostering cynicism.

Previous studies, journalistic investigations and interviews with political strategists have documented this ecosystem's intentional pollution. King's "spaghetti approach" (2024) and Bannon's "flood the zone" strategy (Broadwater, 2025) demonstrate methods of overwhelming social media with deceptive content to test the effectiveness of online communication and drown out opposition. These tactics do not merely mislead but also create confusion by blurring boundaries between false and legitimate information, ultimately undermining epistemic foundations of public discourse (Molina, 2023).

These strategies transform platforms like Facebook into digital ecosystems where the boundaries between malicious human and artificial agents blur, a development Koebler (2024) refers to as the Zombie Internet. These environments are characterised by automated conversations, manufactured controversies, and deceptive identities that dissolve what was once a shared consensus about reality. This represents a fundamental shift from earlier concerns about isolated instances of misinformation to a more pervasive challenge to the integrity of online information itself.

While current efforts to preserve ethical digital information environments often focus narrowly on false news, our research addresses more pervasive deceptive communication activities that have significant yet understudied impacts. Unlike false news campaigns, these information operations rely on iterative testing and adaptation of messages, often deploying a mix of benign and harmful content to obscure intent and evade detection (Chadwick & Stanyer, 2022). These tactics function as gaslighting strategies (Jack, 2017) that systematically manipulate targets into questioning their own perception of reality and judgment of events.

To build on this discussion, we introduce the concept of Deceptive Information Operations (DIOs) to better capture these chaotic, boundary-testing communication activities. We draw on three complementary bodies of theory: the concept of deception in mis/disinformation (Chadwick & Stanyer, 2022) which outlines the intention behind these communication activities, the work on strategic information operations (Starbird et al., 2019), and Meta’s concept of Coordinated Inauthentic Behaviour (CIB) (Gleicher, 2018), which captures both the exploitation of platform affordances and — authentic or simulated — participatory behaviour dimensions of these operations. This integrated approach expands the focus beyond disinformation as circulating false content, encompassing intentional communicative acts carried out with the “intention to mislead, resulting in attitudinal or behavioural outcomes

that correspond with the prior intention" (Chadwick & Stanyer, 2022, p. 2). By articulating DIOs as a broader category, we offer conceptual tools capable of identifying harmful communicative activity that platform labels such as Coordinated Inauthentic Behaviour (CIB) or Coordinated Social Harm (CSH) (Gleicher, 2018) address only partially.

To explore this concept empirically, we analyse three DIOs active on Facebook. Despite its declining popularity among younger audiences, Facebook remains widely used globally, particularly among older and less digitally confident users who may be more exposed to confusing or deceptive content (Guess et al., 2019). Our study draws on a monitoring tool developed within the vera.ai research project to track accounts that repeatedly share content flagged as problematic by Meta's third-party fact-checkers. We also rely on the Meta Content Library (Meta Platforms, Inc., 2025), making this one of the first studies to employ this dataset at scale. Each case is examined through three dimensions: the operation's deceptive features, and the exploitation of platform affordances and the participatory behaviours used to spread it.

The paper is structured as follows: We begin by defining the theoretical background on which DIOs are built, followed by a presentation of the VeraAI Alert and analytical framework used to detect and study these operations. We then examine three case studies, each illustrating different types of networks and distinct deceptive intentions: a network disseminating pro-Putin propaganda, a network promoting online casinos and gambling, and a network of poorly moderated public groups spamming adult content with fraudulent links. We conclude by discussing the broader implications of our findings for content moderation on platforms and future research.

Theoretical Background

Research on online manipulation has generated a rich but fragmented conceptual landscape. Three bodies of theory have done substantial analytical work: Deception Theory, which examines how misleading content reshapes beliefs and behaviour (Buller & Burgoon, 1996); Strategic Information Operations (SIOs), which analyse the exploitation of platform affordances and participatory user behaviours (Starbird et al., 2019); and Coordinated Inauthentic Behaviour (CIB), which identifies deceptive actors and network structures (Gleicher, 2018). Each captures an important dimension, yet none is comprehensive. Deception Theory was not designed for algorithmically amplified environments and, in Chadwick and Stanyer's formulation, focuses narrowly on mis/disinformation. SIO scholarship analyses strategic political intent but remains largely silent on other forms of harm circulated through the same techniques. CIB frameworks offer essential network-level vocabulary but risk reducing a communicative phenomenon to a technical policy category.

These limitations are sharpest in the context examined here. Facebook represents what Koebler (2024) has termed the "Zombie Internet": an environment saturated with synthetic content, inauthentic accounts, artificial engagement signals, and widespread user fatigue.

In this congested ecosystem, the conditions that once enabled relatively straightforward viral spread have eroded. Malicious actors can no longer rely on a single deceptive lever and reaching audiences at scale demands a more sophisticated operational apparatus. The result is a marketplace of deception in which political, financial, and socially motivated actors converge on a shared toolkit — dynamics EU regulators have independently recognised as platform-level systemic risks (European Board for Digital Services & European Commission, 2025) — making motivational distinctions analytically secondary to structural logic, and a unifying theoretical lens an empirical necessity.

This paper proposes deceptive information operations as that unifying framework, synthesising the three bodies of theory into a single analytical lens. DIOs extend beyond the traditional focus on mis/disinformation — typically understood as wholly or partially false information shared intentionally or unintentionally (Jack, 2017) — to encompass organised communicative activities on social media designed to circulate information with the intent to cause harm (Weedon et al., 2017).

Deceiving Social Media Users and Companies

The notion of "causing harm" is a recurring theme in discussions of mis/disinformation. Such operations are deliberate actions intended to inflict harm in pursuit of personal, corporate, or institutional gain (Romero-Vicente et al., 2024). These operations generally fall into three broad categories. The first is political influence, both domestic (Vargo et al., 2018) and foreign (U.S. Senate Select Committee on Intelligence, 2019), in which actors seek to shape public opinion, destabilise societies, or undermine adversaries in support of specific agendas. The second category consists of issue-based campaigns, involving coordinated efforts to promote particular causes, such as conspiracy theories or anti-vaccine narratives. Although often grassroots in appearance, these campaigns share important strategic similarities with politically motivated operations (Huth et al., 2020). Finally, lucrative information operations are financially motivated, relying on clickbait, traffic redirection to monetised websites, and the exploitation of online advertising systems (Silverman & Alexander, 2016).

Chadwick and Stanyer (2022) developed their framework specifically in the context of political mis/disinformation and its threat to democratic discourse. They propose deception as a key concept for understanding the harm malicious actors inflict on their targets. According to the authors, deception occurs when an actor's intent to mislead produces attitudinal or behavioural changes aligned with that actor's objectives (2022, p. 6). Earlier, Buller and Burgoon (1996) argued that deception depends on the successful execution of misleading intentions during communication. This perspective emphasises the interaction between deceivers and the deceived, framing deception as a communicative act aimed at fostering false beliefs or misunderstandings (Chadwick & Stanyer, 2022, p. 3). By focusing on the theoretical relationship between intent, process, and outcome (Rubin, 2016), this

approach offers an empirical framework for analysing information operations through the "attributes and actions of deceptive entities" (Chadwick & Stanyer, 2022, p. 14).

This study treats the DIO framework as an adaptation of communicative deception theory, whose core analytical logic — intent to mislead, communicative execution, and resulting attitudinal or behavioural change — is not inherently limited to the political domain. The deceptive mechanism remains structurally consistent whether the harm is epistemic (undermining shared reality), economic (defrauding users through scam content), or social-normative (evading community standards). What varies is motivational context and harm type.

Deception operates at two interdependent levels. At the user level, operations manufacture false impressions of organic popularity, community consensus, and account authenticity. At the platform level, malicious actors manipulate algorithmic amplification and evade policy enforcement. The two levels are operationally linked: platform-level deception provides the infrastructure through which user-level deception is scaled.

Platform Affordances Exploitation and User Participation to Amplify Deceptive Information

The spread of deceptive content has been examined from multiple perspectives. Meta and other platforms initially framed these operations through the distinction between authenticity and inauthenticity, defining Coordinated Inauthentic Behaviour (CIB) as a phenomenon in which groups of individuals work together to mislead others (Graham, 2024). In addition to CIB, Meta introduced the concept of Coordinated Social Harm (CSH), referring to networks of users who systematically violate platform policies in order to inflict harm (Rogers & Righetti, 2024).

Common across these forms of manipulation is the strategic exploitation of algorithm-driven information flows to maximise resilience and reach (Starbird et al., 2019). Through these techniques, information operations achieve rapid dissemination and demonstrate adaptability across technological contexts and audiences (Bradshaw & Howard, 2018). Social media platforms provide tools originally designed for advertisers — such as audience targeting and impression management — that malicious actors can use to deliver content to carefully selected publics. Combined with algorithms prioritising engagement, these affordances enable the rapid circulation of content and create self-reinforcing feedback loops that amplify the visibility and impact of deceptive operations (Gillespie, 2021).

Moreover, social media platforms blur the boundaries between news, entertainment, and opinion within a single feed, collapsing contextual distinctions (Newman et al., 2025). This genre collapse makes it more difficult for users to discern the intent behind a given piece of content, particularly in a post-broadcast information environment characterised by the multiplication and fragmentation of channels (Prior, 2014). Exploiting this ambiguity,

malicious actors frequently combine accurate and misleading information or adopt fabricated identities to obscure their true intentions (Wardle & Derakhshan, 2017).

Algorithms play a central role in governing the dissemination and consumption of information on social media. Platforms use complex ranking systems to sort, filter, and prioritise content, optimising primarily for engagement and time spent on the platform (Fowler, 2016). Scholars have argued that, by personalising content according to users' interests and behaviours, these algorithms may create "filter bubbles" that reinforce biases and limit exposure to diverse viewpoints (Gillespie, 2021). Algorithms tailor content and search results on the basis of users' interests, past behaviours, and geographic location. Although the internet theoretically expands access to information, algorithmic curation often constrains the free circulation of ideas necessary for healthy democratic deliberation. Content that provokes outrage, confirms pre-existing beliefs, or drives engagement is more likely to be prioritised, facilitating rapid dissemination regardless of informational accuracy (Al-Rawi, 2019). Consequently, while access to information has expanded in the contemporary media ecosystem, exposure to critical or diverse perspectives may have become more limited.

The exploitation of platform affordances does not operate in isolation; rather, it is systematically amplified by coordinated user behaviour. Social media users with similar goals are able to coordinate their activities without centralised control (Jenkins et al., 2015). In the context of information operations, such coordination may serve explicitly malicious purposes (Giglietto et al., 2020a). Scholars and platforms have described these collaborative dynamics using concepts such as dark participation (Quandt, 2018), CIB (Gleicher, 2018), and coordinated influence operations (O'Donovan, 2024). The most widely recognised definition remains Meta's concept of CIB, which refers to networks of inauthentic accounts working together to deceive users, platforms, or systems in order to influence public discussion, commit fraud, or evade policy enforcement (Romero-Vicente et al., 2024).

Coordination plays a central role in the success of SIOs. It enables the systematic production of misleading content while amplifying dissemination through the exploitation of platform algorithms (Gleicher, 2018; Graham, 2024). Because actions on social media are publicly visible and quantitatively measurable — and because these metrics shape how algorithms prioritise content — coordinated behaviour can substantially increase the perceived credibility and impact of deceptive content. Visibility thus becomes a proxy for influence, making coordinated participation particularly effective in manipulating public discourse (Allen, 2022).

The Present Study

Taken together, these theoretical strands converge in the concept of deceptive information operations. Crucially, DIOs need not rely exclusively on false content: misinformation may circulate within them but is neither a necessary nor a sufficient condition. What defines them

is the convergence of orchestrated misleading practice, exploitation of platform affordances, and (simulated) user coordination — a combination that the research questions below are designed to examine empirically. An overarching general question frames the inquiry; three sub-questions operationalise each dimension in turn, enabling systematic analysis of the cases examined.

RQ1. How do deceptive information operations integrate intent to deceive users, platform affordances exploitation, and coordinated behaviour within digital ecosystems?

RQ1a. How are the operations qualitatively deceptive in their intent and execution?

RQ1b. Which platform-specific features are exploited to automate and amplify this deception?

RQ1c. How does coordinated participatory behaviour influence the reach and perceived legitimacy of these operations?

Method

This study investigates three cases of deceptive information operations on Facebook discovered by a tool implementing a workflow designed for monitoring lists of problematic actors and their activities (Giglietto et al., 2023). The workflow builds upon existing methodologies for detecting information operations by integrating automated content collection, behavioural pattern analysis, and iterative account tracking. The workflow and analytical approach are displayed in Figure 1.

The three cases were selected purposively (Palys, 2008) to represent the three categories of harm identified across the 17 networks detected by the workflow: epistemic harm (pro-Kremlin political propaganda), economic harm (gambling promotion), and social-normative harm (the circulation of unsolicited adult material). They are not statistically representative of the broader population of deceptive operations but are analytically illustrative of structurally distinct operation types, each of which mobilises the same convergent logic — deceptive intent, affordance exploitation, and coordinated participation — in service of different malicious objectives.

Vera.ai Workflow: Design and Implementation

The vera.ai workflow operates through a multi-step process that systematically detects, analyses, and updates lists of coordinated social media accounts. Unlike static network analysis approaches that rely on one-time data collection, this workflow utilises an adaptive methodology to track evolving coordination patterns over time (Giglietto et al., 2023).

The process begins with the identification of a curated seed list of social media accounts flagged as potentially problematic. These accounts are selected based on entries in verified fact-checking databases, investigative reports as well as academic research. Starting with this high-confidence baseline ensures that the initial sample is both credible and analytically

robust. Once the seed list is established, the system initiates automated data collection. Scheduled queries are employed to systematically retrieve posts from the seed accounts at predetermined intervals. The system prioritises content that exhibits high engagement – measured through shares, reactions, and comments – as these indicators often signal broader influence or coordinated activity.

To detect patterns of coordinated behaviour, the workflow utilises CooRTweet , which is specifically designed to identify multiple forms of synchronous sharing (Rogers & Righetti, 2025). These include the simultaneous posting of identical URLs across multiple accounts, similarities in image text and post blurbs, and temporal clustering in content distribution. Such behaviours often point to coordinated efforts to amplify specific narratives and distort public discourse.

Following detection, the workflow moves into the feature extraction and network expansion phase. Here, elements like URLs, images, and text snippets are analysed across the platform to identify other accounts that exhibit similar patterns of dissemination. Accounts that consistently engage in such activities are incorporated into the monitoring pool, allowing the system to adapt, in real time, to shifts in coordination strategies.

Between October 2023 and August 2024, a CrowdTangle-based implementation of this dynamic tracking approach named VeraAI Alert System demonstrated substantial scalability in detecting coordinated information operations. The system was initiated with a seed dataset of 1,225 Facebook accounts – comprising both Pages and public groups – each strategically selected based on documented dissemination of misinformation, specifically having shared in a coordinated manner a minimum of four URLs from a corpus of 36,091 web pages identified as false by Meta's third-party fact-checking partners between 2017 and 2022 (Messing et al., 2020). During the 10-month operational period, the system identified 10,681 unique coordinated links, revealing the underlying content infrastructure of influence operations, and discovered 2,126 previously undetected accounts beyond the original seed list. The incorporation of these newly identified accounts into the monitoring framework established a self-reinforcing detection mechanism that ultimately captured 7,068 coordinated posts displaying synchronised sharing patterns. Analysis revealed 17 distinct networks pursuing varied operational objectives across multiple geographic regions. These networks encompassed anti-vaccine discourse, migration-related narratives, and regionally targeted campaigns, with notable activity clusters in Southeast Asia, Eastern Europe, and Latin America. System operations ceased on August 14, 2024, following Meta's deprecation of the CrowdTangle platform.

Case Studies Selection

We selected a purposive sample (Palys, 2008) from the 17 identified networks to examine a representative cross-section of DIOs. This approach enabled us to examine multiple dimensions of coordinated behaviour, including the rapid, synchronous dissemination of identical or near-identical links and images within a single group, across multiple groups by

networked actors, or by users exploiting weak moderation to amplify illicit content. The three cases were selected on the basis of three criteria: (a) typological diversity — they represent distinct operational objectives (political influence, financial fraud, and content-based harm); (b) actor diversity — they involve different types of principals, including state-aligned actors, commercially motivated automated networks, and weakly governed user communities; and (c) coordination mechanism diversity — they illustrate centralized seeding, bot-driven amplification, and exploitation of moderation gaps respectively. In terms of geographic scope, the pro-Putin network is transnational in nature, with content in Russian, Arabic, and English and administrators identified as Bulgarian and Russian; the gambling network targets primarily English-speaking audiences; and the unmoderated groups display political content tied to South African and Pakistani contexts, alongside internationally circulated entertainment material.

Our sample encompasses a variety of actors engaged in coordinated sharing, such as individual users, automated accounts, organised groups, and pages operated by vested interests. This diversity illustrates that coordination is not limited to a specific type of actor but involves participants with different motivations and strategies. The cases also reflect the wide array of misleading objectives pursued through coordinated sharing, including spreading misinformation, promoting engagement bait, conducting political propaganda, amplifying scams, and distributing spam and pornographic content.

After identifying these cases, we imported the lists of accounts into the Meta Content Library (the new tool designed by Meta in lieu of CrowdTangle) to streamline data retrieval and analysis. This step enabled the systematic observation of relevant posts, engagement metrics, and associated dissemination patterns of these accounts. However, the adaptation to MCL also presented challenges, such as the severe limitation on the number of downloadable posts, which necessitated conducting observations within the platform.

The Analytical Approach

Our analytical approach followed an inductive category development process informed by constructivist grounded theory (Charmaz, 2014; Starbird et al., 2023). While the three analytical dimensions — identity management, behavioural coordination, and content manipulation — were theoretically motivated by our framework, the specific categories within each dimension (such as brandjacking, narrative mixing, and synchronised posting bursts) emerged inductively from close reading of the data rather than being imposed a priori. The analysis scheme is available in the Appendix.

The online dimension of deceptive information operations enables the study of motives, tactics, and narratives through the digital traces they leave on platforms – traces that, while incomplete (Starbird et al., 2019), are nonetheless observable and measurable. We analysed a dataset combining account-level metadata and behavioural patterns with the top 500 most-viewed posts for each case study within the time period between October 2023

and August 2024, tracked on the Meta Content Library (Meta Platforms, Inc., 2025) between November 2024 and January 2025.

Our analysis involved an iterative movement between micro – and macro-level perspectives: systematic open coding of individual posts and close reading of emerging patterns. Through constant comparison, analytical categories emerged organically from the data and were subsequently interpreted through the theoretical frameworks of Starbird et al. (2019) and Chadwick and Stanyer (2022), as discussed in the previous section.

Each case study was assigned to an author for primary coding. To ensure consistency, the lead author cross-coded 30% of the dataset, and all coding decisions were reviewed collaboratively to refine the emerging categories. Operational definitions for all categories are provided in Table A1 in the Appendix.

The analysis revealed three main types of patterns. The first concerns how accounts manage their identity, e.g. through name changes, thematic inconsistencies, and traces of administrator activity. The second relates to coordinated behaviour, e.g. through synchronised posting bursts and content shared simultaneously across multiple groups. The third involves content manipulation, where, for example, AI-generated images, impersonation of legitimate brands, and the deliberate mixing of harmless and harmful narratives were all observed.

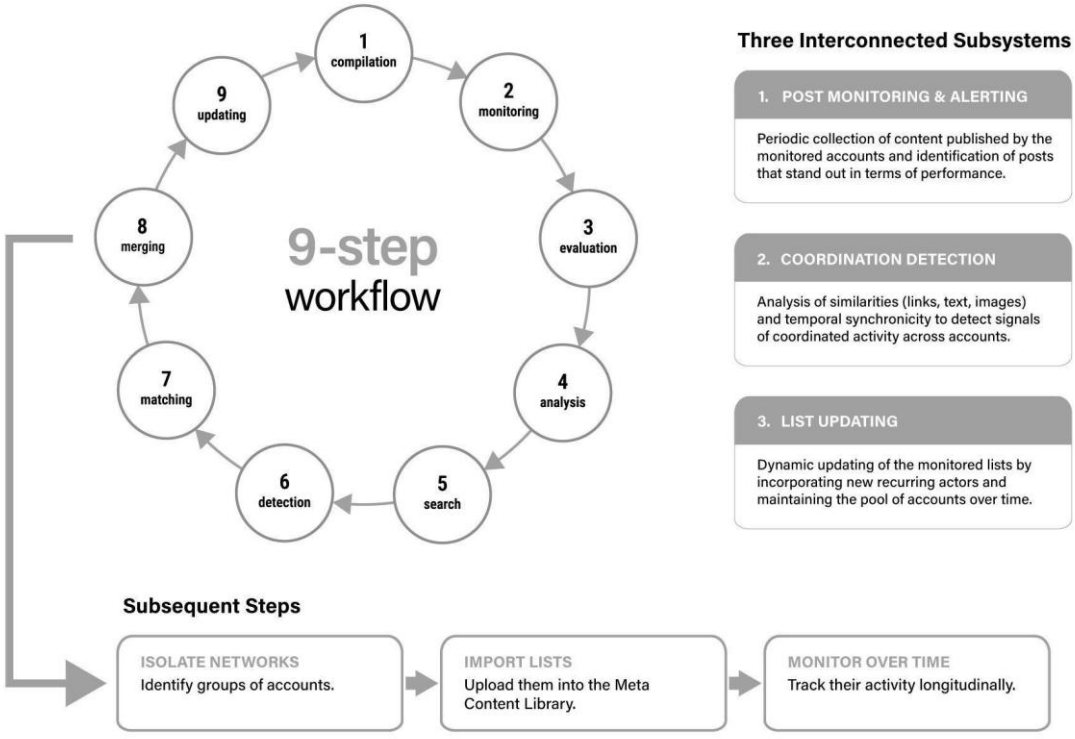

Figure 1. Workflow and analytical approach.

Findings

This section illustrates the three case studies selected from a map of account networks exhibiting coordinated sharing behaviour identified by the VeraAI Alert System. The selected cases illustrate distinct types of deceptive information operations: political propaganda, online gambling promotion, and the dissemination of a mix of misinformation, memes, and adult content.

The subsequent sections will analyse the case studies in depth. Following our theoretical frame, the analysis of each deceptive information operation follows a threefold structure guided by the three research questions: [1] describing deceptive content, [2] investigating how Facebook affordances are exploited, and [3] observing users' participatory behaviours.

Pro-Putin Propaganda

Spreading Deceptive Content for Political Influence

The first case study identified 27 public groups on Facebook that exhibit strong pro-Putin affiliations. This is the most traditional form of information operation to promote political propaganda, resonating with Chadwick and Stanyer's (2022) framework of deception driven by political motivations. The deliberate integration of apparently benign material, such as posts celebrating Moscow's urban development, alongside more explicitly politically polarising content, such as mocking the Ukrainian President, exemplifies the gaslighting nature of these information operations, fostering confusion and advancing a specific agenda (Jack, 2017). These groups display notable structural similarities, including the frequent replication of naming conventions, identical group descriptions, and synchronised name changes preceding key geopolitical events. For instance, the group THE BEST PRESIDENT, VLADIMIR VLADIMIROVICH PUTIN, which has approximately 35,000 members, includes a profile description outlining community rules that explicitly prohibit abusive language and spam while also not overtly expressing a pro-Putin stance. Other groups with nearly identical names and descriptions, which have twice as many members, also exhibit patterns of coordinated behaviour. Notably, several groups underwent name changes in early 2022, just weeks before Russia invaded Ukraine, suggesting strategic adaptation in response to geopolitical developments.

The content disseminated within these groups can be grouped into two primary thematic categories: first, posts that disparage Western leaders and Ukraine, and second, content that seeks to bolster the image of Putin and Russia. Among the most widely circulated posts in the observed period was a meme targeting Volodymyr Zelenskyy, shared in coordination across multiple pro-Putin groups within a tightly constrained time frame. On February 12, 2024, between 9:12 and 9:18 CEST, six groups posted an identical image portraying the then U.S. President Joe Biden in a Soviet military uniform, sternly holding a cat with

Zelenskyy's face, accompanied by the caption: "Dad, don't leave me... Russia has won, I don't need you anymore" (see Figure 2). The timing of this coordinated posting coincided with significant battlefield developments, including a Russian drone strike in Kharkiv that resulted in civilian casualties. This suggests that the disinformation network actively leveraged real-world events to reinforce the narrative of Russian military superiority while portraying Western support for Ukraine as diminishing.

In contrast to the satirical posts, another subset of content was designed to project a positive image of Putin and Russia. Several widely circulated posts sought to depict Moscow as a superior urban centre compared to Western cities. In one instance, two posts prominently featured American journalist Tucker Carlson, who apparently described Moscow as "a more pleasant city than anywhere in the U.S." during his speech at the World Government Summit in Dubai. Another post claimed that a Canadian farming family had relocated to Russia to preserve traditional values, reinforcing a broader narrative of Russia as a bastion of moral integrity in contrast to perceived Western decadence. Additionally, several AI-generated images depicted Putin in various heroic poses, including one surrounded by young women, seemingly reinforcing his appeal to younger generations. Other images drew on nationalistic symbolism, such as a Russian bear wielding a Kalashnikov rifle in front of a national flag or a lion juxtaposed against hyenas bearing the flags of Ukraine, the EU, and the US. These visual representations functioned as implicit ideological statements, reinforcing a binary opposition between Russia and the West.

Exploiting Facebook's Affordances for Algorithmic Amplification

The systematic amplification of content within these groups reveals a deliberate exploitation of Facebook's platform affordances to maximise visibility and engagement. In particular, coordinated sharing behaviour is employed to game Facebook's ranking algorithms, prioritising content that generates engagement. By ensuring that identical posts appear in multiple locations within short intervals, the actors behind these operations effectively created an illusion of organic popularity, increasing the likelihood that the content would be further highlighted by the platform's recommendation algorithms. A good example demonstrating a high degree of coordination in its dissemination of content is WorldRusWorld, managed by Bulgarian and Russian moderators and one of the most active content producers within this network. On October 1, 2023, for instance, WorldRusWorld posted an identical pro-Putin image in ten different groups within a three-minute window (16:21–16:24 CEST). The near-simultaneous posting pattern suggests an organised effort to amplify reach and visibility, indicative of a coordinated campaign.

The structured timing of content dissemination aligns with previous studies on coordinated influence operations, where automated or semi-automated systems play a key role in content propagation (O'Donovan, 2024). Furthermore, while most of the content was in Russian, some of the most widely viewed posts appeared in Arabic and English, indicating an effort to target non-Russian audiences (see Figure 2 for an example).

Coordinated Participation and Broader Implications

The engagement metrics associated with these posts reveal significant variation in effectiveness. Some failed to attract substantial engagement, with posts generating only a few hundred views and minimal interaction. However, others demonstrated considerable reach, with the most successful posts exceeding 100,000 views and accumulating thousands of reactions, comments, and shares. For instance, the Zelenskyy meme campaign in Figure 2, which displayed clear signs of coordination, accumulated over 1.1 million views, underscoring the extensive reach and potential impact of these efforts.

While the initial coordination of content is orchestrated by a relatively small group of actors, the amplification of these messages often depends on the actions of ordinary users who unwittingly participate in spreading misleading content. This interplay between orchestrated coordination and organic participation reflects the broader challenges posed by contemporary disinformation campaigns, where the distinction between authentic and inauthentic engagement becomes increasingly blurred.

Overall, this first case study highlights a state-aligned operation promoting pro-Putin propaganda. Coordinated Facebook pages and groups create the illusion of widespread grassroots support for Russia's geopolitical goals. Automated tools synchronise content dissemination, exploiting Facebook's features to amplify narratives supportive of the Russian government. This behaviour aligns with previous findings on using social media affordances for propagandistic purposes (Giglietto et al., 2020b).

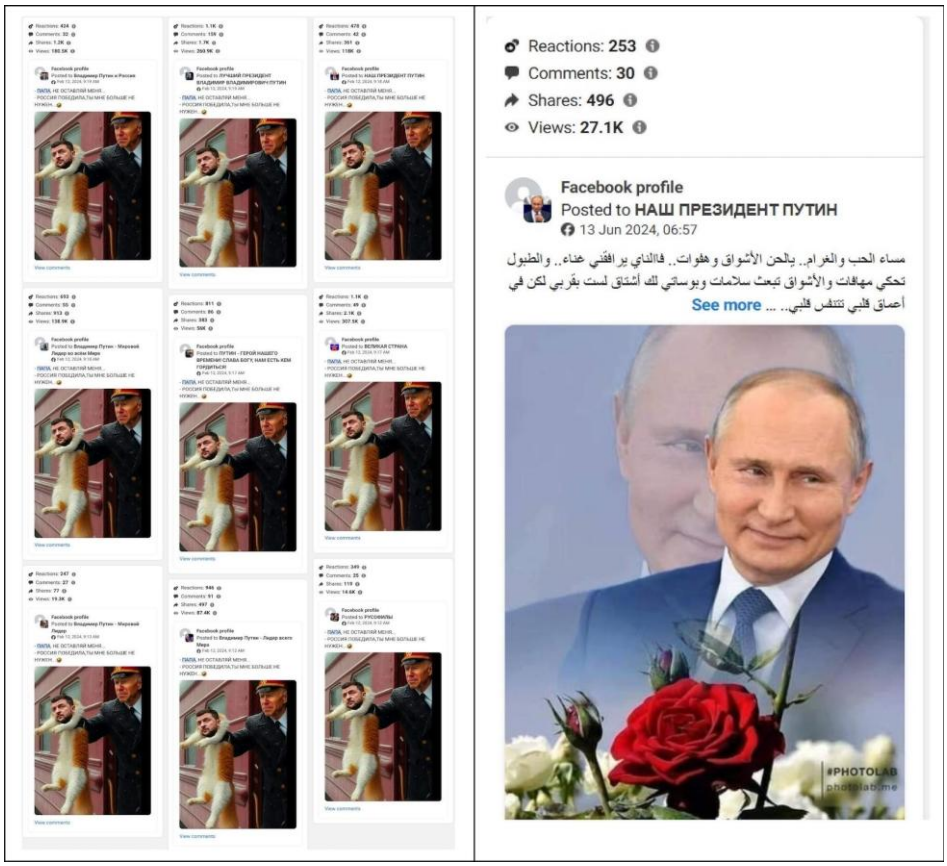

Figure 2. On the left: Meme mocking Zelenskyy with the satirical caption shared simultaneously across multiple pro-Putin accounts; on the right: Post in Arabic in poetic language to express emotions of longing.

Online Casino Engagement Bait

Spreading Deceptive Content for Gambling Promotions

The second case study focuses on a coordinated campaign promoting online casinos and games through 260 Facebook groups. These groups, with memberships ranging from 30,000 to nearly 600,000, use automation and generative AI to produce and share content promoting a set of digital casino websites. The consistent thematic branding in group names reflects a deliberate strategy to attract users with gambling incentives. Notably, among the 500 most-viewed posts from the network, 368 were no longer accessible at the time of the analysis, likely due to systematic policy violations leading to content takedowns. This suggests a strategy designed to maximise exposure before removal.

These groups signal deception through branding tactics, prominently featuring gambling brands like Orion Stars and Juwa – appearing 111 and 90 times, respectively – without verified affiliations. For example, the third-largest group, FREE PLAY ALL GAMES (JUWA CITY ORION STARS / FIRE KIRIN / GAME VAULT) (429,332 members), illustrates how these brands gain visibility within the network. By leveraging well-known names, these groups promote lesser-known casino apps through brandjacking (Du Plessis, 2018).

Beyond misleading branding, these operations cause harm by luring users into scams under the guise of legitimate gambling platforms, exploiting vulnerabilities, and fostering economic harm. They also employ synthetic content, such as AI-generated visuals and comments, further polluting the digital environment (DiResta & Goldstein, 2024) and fuelling distrust. The operation's ability to test and refine policy circumvention strategies amplifies its impact and lays the groundwork for future campaigns. The accumulation of social media assets and networked groups increases their potential for repurposing in other deceptive operations (Donovan et al., 2020), including disinformation or fraud, posing a broader threat to digital integrity.

Exploiting Facebook's Affordances for Algorithmic Amplification

Automation plays a key role in these groups for maintaining high posting frequency and consistency. Some groups produced up to 10,000 posts each in a single month (Figure 3), a volume that would be difficult to sustain manually. The near-simultaneous distribution of identical promotional content across multiple groups suggests the use of scheduling tools and bots, ensuring continuous and widespread publication. This tactic amplifies visibility within Facebook's algorithmic feed, increasing user exposure.

The dataset's most-viewed posts show a stark imbalance, with disproportionately high comment count relative to reactions and shares. One post, for instance, amassed 169.1K views and 44.1K comments but only 7 reactions and 0 shares – an anomaly suggesting artificial comment inflation. These comments often contain generic praise ("great platform," "awesome gameroom"), random alphanumeric strings ("k4," "oyy," "flq"), or unspaced words

("Awesomekeepuptheamazingworkandkeepdoingyourbest"), aligning with synthetic engagement strategies (Cresci et al., 2016). By fabricating popularity, these campaigns mislead users into perceiving widespread endorsement, a tactic consistent with broader concerns about increasing automation in digital interactions as within the broader discussions on the 'Dead Internet Theory' (Muzumdar et al., 2025) positing that much of the online content and interactions are generated by bots rather than humans.

Group administrators play a crucial role in sustaining this automation-driven strategy. Many administrator accounts show minimal non-automated activity, resembling compromised profiles used in spam networks. The analysis suggests these accounts are centrally managed, facilitating seamless gambling content distribution.

For instance, the largest group, ORION STARS, JUWA, GAME VAULT, FIRE KIRIN, MOOLAH - FREEPLAY (595,839 members), is managed by 6 administrators and 5 moderators. Except for the page "OrionStars Freeplay 24h/7," the remaining admins have low engagement, only a few followers, and very little content. Many label themselves as "digital creators," yet their profiles lack meaningful activity. Some administrators, like "Dm me 25$ Freeplay" and "Message me for 30Freeplay," exist solely to promote entrance bonuses, lacking original content or followership.

Shared content analysis reveals frequent low-engagement posts, emphasising promotion. Posts often advertise external casino pages offering bonuses or invite users to join private chats that include reward mentions in their titles.

[Synthetic] Coordinated Participation and Broader Implications

Although user participation in these gambling groups appears extensive, much of it is artificially generated through AI visuals and engagement bait tactics that simulate activity. This deceptive model blends synthetic and real user interactions, making genuine users unknowingly complicit in amplifying misleading content.

Our analysis of the 500 most-viewed posts reveals centralised coordination. AI-generated imagery plays a key role, enhancing visual appeal while streamlining content production. These graphics – often depicting slot machines, lions, sharks, and regal female figures – serve to attract attention while enabling rapid, cost-effective content creation. A pattern of image recycling is evident, with the same promotional graphics reappearing across multiple groups, such as the AI-generated shark image (Figure 4) repeatedly disseminated across the network on the same day, often by anonymous Facebook profiles posting in near-simultaneous intervals (e.g., multiple posts on August 6, 2024, at 01:40 AM).

This follows an established pattern where content from central Facebook pages is shared by private or public profiles in these groups. Public profiles engaging in this behaviour typically describe themselves as digital creators, have at least 1,000 followers, and blend casino promotions with investment and life-coaching content. One such profile, for example, mirrors behaviours observed in the following case study on exploiting unmoderated groups for deceptive content. These accounts remain largely inactive over time, apart from periodic profile image changes. Engagement incentives, such as free play bonuses for likes and

comments, encourage real users to interact with manipulated posts. This boosts visibility, attracting further engagement and reinforcing the content's presence in the platform's algorithmic ranking.

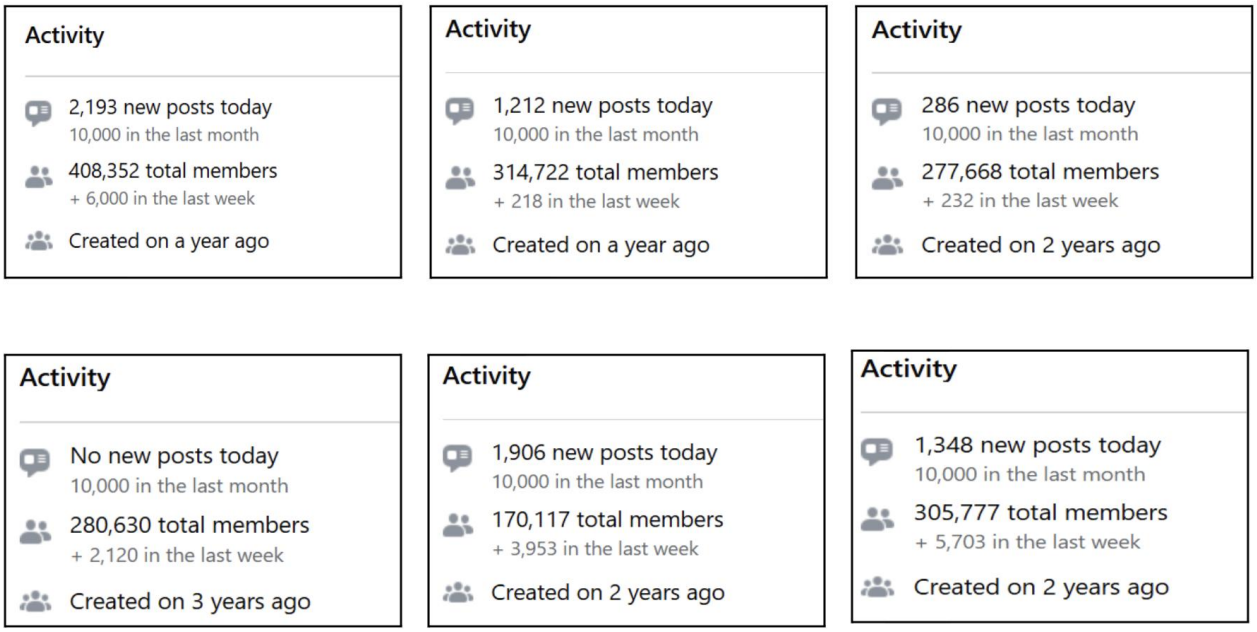

Figure 3. Boxes showing the activity of some of the accounts in the network.

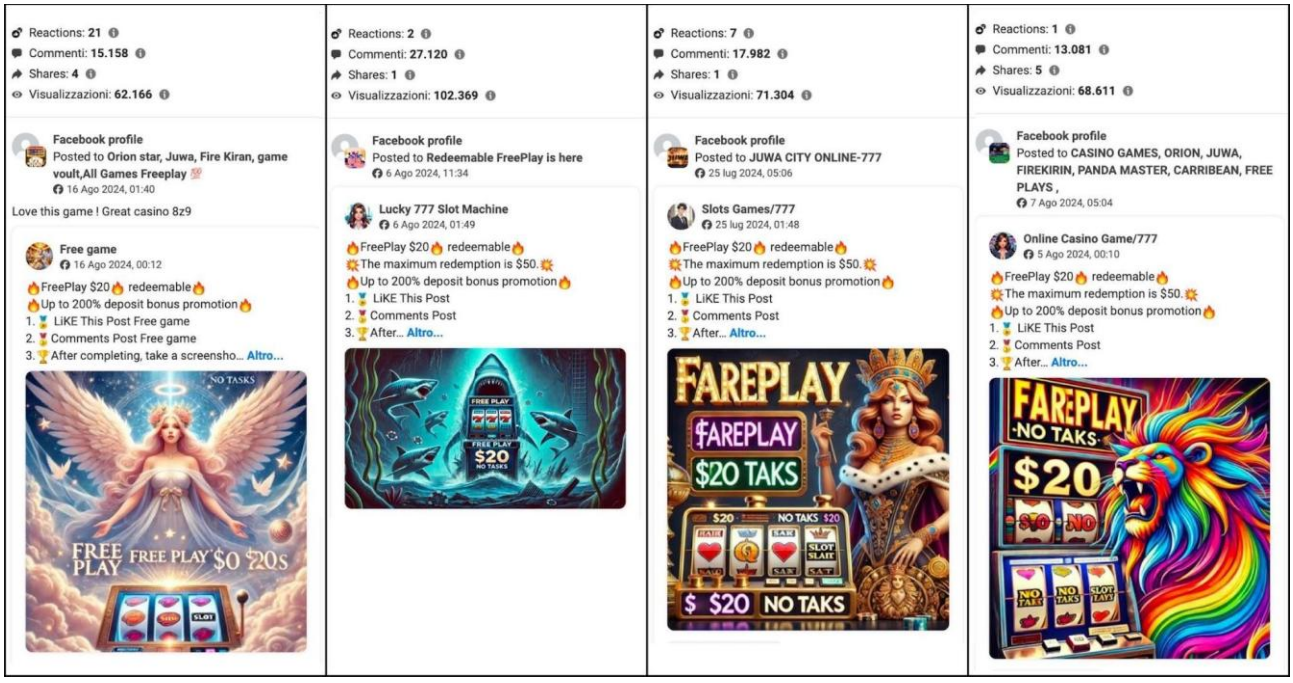

Figure 4. Four examples of posts shared from the network accounts displaying AI-generated visuals.

Unmoderated Groups Flooded with Explicit Imagery

Spreading Deceptive Content for Circulating Explicit and Illicit Material

The third case study examined 222 public Facebook groups, which formed two distinct communities: one with 187 accounts and another with 35. The findings reveal deceptive practices exploiting large, poorly moderated spaces. The deceptive nature of these operations manifests itself in two distinct ways. First, multiple accounts are misleadingly renamed, shifting between political and entertainment themes. For example, a group originally named EFF Juju Nation was later renamed DSTV Premier League Soccerzela and changed names 13 times. Before the South African elections in May 2024, it was temporarily rebranded as 2024 Elections South Africa, then reverted to a soccer-related name post-election. Despite these shifts, its core content remained apolitical, suggesting a strategy to manipulate public perception while evading detection. Similarly, The Golden Girls Fan Club was previously linked to the Pakistan Peoples Party (PPP), posting pro-PPP content before going dormant for nearly a year. It later rebranded as a fandom group for The Golden Girls while mixing sitcom-related posts with spam and celebrity death hoaxes, exposing users to misinformation when they are most susceptible – while consuming entertainment content.

Second, the lack of moderation allows misinformation on wide-ranging topics, adult content with fraudulent links, fabricated job ads, and illegal services like follower sales to spread unchecked. A notable example of misinformation is the false claim that Coca-Cola contains alcohol. Despite being debunked by fact-checkers, it circulated widely, accumulating 3.5 million views and 37,300 shares at the time of analysis (Figure 5).

Exploiting Facebook's Affordances for Algorithmic Amplification

The VeraAI Alert detected a sophisticated deception technique within this information operation on Facebook. The operation begins when private users share a "listening to" activity status update inside these groups. This status includes a link to a Facebook page categorised as Musician/Band. However, these pages are typically empty, with profile images that change periodically to clickbait visuals—riddles, pornography, or emotionally charged images of deceased individuals or people with visible disabilities. We hypothesise that this tactic helps evade Meta's detection of CIB.

This is not the only circumvention method observed. For instance, adult content is often disguised under unrelated blurbs narrating the history of BMW, the automobile company, all posted with a set of hashtags, including #carporn (see Figure 5). These posts are strategically coordinated to maximise visibility and mask the real link destination under seemingly safe domains, including, ironically, facebook.com. An anonymous participant shares a post in a public group originally created by another user. It takes the form of a "link" status update (e.g., "reading," "listening", "attending") and appears automotive-related, referencing BMW's history. However, its preview image contains explicit content. The post

links to an external website through tinyurl.com link shortener. This formula appeared repeatedly across the dataset, suggesting a templated approach to content disguise rather than isolated incidents.

Many of these posts follow a formula that maximises engagement by leveraging sensationalism, humour, or controversy. Engagement bait – such as provocative memes and misleading headlines – ensures sustained visibility. A particularly viral example featured a female teacher in a classroom with the caption “Teachers nowadays are 🔥,” (Figure 5), generating massive engagement despite its apparently mundane nature.

Coordinated Participation and Broader Implications

The user base of these accounts plays a key role in orchestrating information operations, often with or without administrators' knowledge. The lack of moderation in these groups allows deceptive content to proliferate. Among the 500 most-viewed posts, 38 surpassed one million views, with the top three exceeding 3 million. One widely circulated example was a manipulated meme of Phineas from Phineas and Ferb, humorously questioning his shirt-wearing habits. Though seemingly harmless, its viral spread across X, YouTube, Reddit, and TikTok (confirmed via Google Lens) illustrates how user engagement drives visibility, regardless of accuracy or intent.

Some groups adopt misleading appearances in order to attract more members. For example, the group Sons and Daughters of Apostle and Dr. Lizzy Johnson Suleiman presents itself as a religious community, yet users there primarily share adult content (DiResta & Goldstein, 2024).

Broadly, this case study underscores how deception is deeply embedded in Facebook’s information operations and CIB. By exploiting platform affordances – such as name changes, AI-generated content, and algorithmic amplification – these networks manipulate public discourse while avoiding detection. High user engagement further fuels their spread, ensuring the continued circulation of deceptive content.

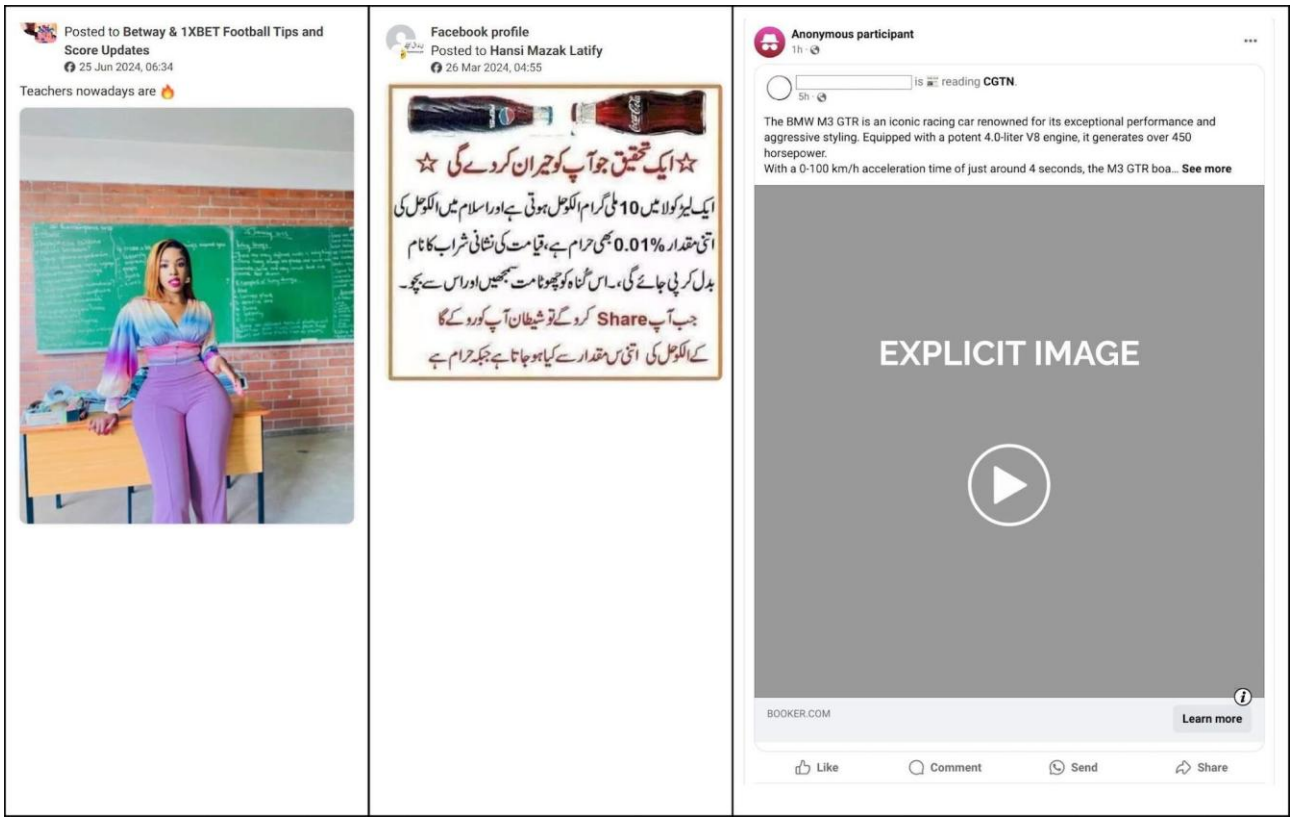

Figure 5. On the left: A widely circulated image of a female teacher in a classroom, shared with a provocative caption; In the center: A viral piece of misinformation claiming that Coca-Cola contains alcohol. On the right: An anonymous participant shares a post in a public group originally created by another user. The post blurb briefly tells BMW's history. However, its preview image contains explicit content.

Cross-case Synthesis

The three cases – anti-democratic political propaganda, gambling promotion, and adult content – show both shared strategies and important differences in how deceptive information operations function on Facebook. Across all cases, actors conceal their intent behind seemingly harmless content, rely on coordinated sharing behaviour, and use synthetic or AI-generated materials to amplify messages while maintaining an appearance of organic activity. In doing so, they exploit the platform's engagement-driven ranking system and adapt to the evolution of moderation policies.

Where they diverge is in how they engage users, depending on each network's goals. The pro-Putin operation blends coordinated seeding with genuine user engagement, allowing authentic interactions to further extend the reach of these narratives. The gambling network relies almost entirely on synthetic activity, creating the illusion of engagement with minimal real-user participation. Unmoderated groups occupy a middle ground: genuine – or presumed genuine – users interacting with entertaining content inadvertently amplify deceptive posts, including adult material but also various forms of disinformation.

These dynamics produce different effects: international visibility for political propaganda, high-volume but shallow engagement in gambling networks, and the circulation of deceptive

content under the guise of innocuous posts. Taken together, the cases expose structural vulnerabilities in Facebook's architecture that make DIOs particularly difficult to detect and disrupt.

Discussion and Implications of Findings for Platforms Governance and Research

Across the three case studies of what we call deceptive information operations, the same dual mechanism enabled large-scale circulation of harmful content: algorithmic ranking systems amplified engagement-driven material, often intensified by participatory or simulated-participatory behaviours such as coordinated sharing, while moderation was inconsistently enforced. The state-backed political operation created coordinated pages and groups, using automated sharing and AI-generated images to trigger algorithmic amplification of pro-Kremlin narratives, blending in seemingly harmless content to reduce the risk of detection. The gambling operation mass-produced promotional materials with generative AI and bot-driven engagement, overwhelming the platform more quickly than moderation systems could respond. The operation distributing adult material repurposed large, seemingly benign Facebook groups for pornographic scams, exploiting spaces with minimal or absent moderation and embedding harmful content alongside unrelated posts to avoid bans. Across all three cases, Facebook's enforcement framework failed to reliably distinguish automated from genuine participation, or deceptive material from legitimate content — producing information asymmetries that blurred the line between authentic and manipulated activity.

The cases also diverge in important ways, most visibly in their deception profiles. The pro-Putin propaganda network primarily produces epistemic harm, distorting public understanding of geopolitical events and degrading the quality of political discourse. The gambling network produces economic harm, luring users into fraudulent schemes under the guise of legitimate platforms. The unmoderated groups produce social and normative harm, exposing users — often in ostensibly civic or entertainment spaces — to exploitative and illegal content. The actor configurations differ as well: a state-backed network, a profit-driven commercial operation, and a more diffuse set of operators exploiting community infrastructure. What unites them is structural rather than substantive — a point that calls for a framework capable of recognising operations whose surface features differ widely but whose underlying logic does not.

That framework requires a step beyond existing accounts. Starbird et al. (2019) develop their account of information operations through cases of political manipulation; Chadwick and Stanyer (2022) frame their typology around disinformation. Meta's CIB category (Gleicher, 2018) is broader in ambition — it is the dominant enforcement frame for deceptive activity at scale — but its operational definition is narrow, recognising harm only at the intersection of coordination and inauthentic actors. Each lens, then, draws its boundary too tightly — around political manipulation, around disinformation, or around a specific

configuration of coordination and inauthenticity — leaving structurally similar operations unaccounted for. We advance the concept of DIOs to occupy this broader space, defining such operations by the convergence of three dimensions: deceptive intent expressed in content and identity practices, strategic exploitation of platform affordances, and multi-account coordination. Each dimension varies independently; it is their convergence, not any single feature, that marks an operation as deceptive in the sense developed here.

These dynamics describe what Koebler (2024) calls the "Zombie Internet": an ecosystem condition in which the distinction between human and automated participation becomes increasingly difficult to sustain. Operations are the unit of activity; the Zombie Internet is the systemic condition their proliferation both depends on and reproduces. When algorithms reward engagement irrespective of authenticity, and moderation cannot reliably separate coordinated from organic participation, manufactured interactions and synthetic identities accumulate into a feedback loop that lowers the cost of further operations. The affordances implicated here — algorithmic curation, engagement-driven visibility, and group autonomy — are not unique to Facebook. Although our evidence is platform-specific, the condition it points to plausibly is not. This is the deeper tension the cases expose: the features that make social media commercially successful are the same ones that make it hospitable to manipulation (Schroeder et al., 2026). The Zombie Internet, in this reading, is less an aberration than an emergent property of engagement-oriented design operating at the limits of governance.

These conceptual limits have direct governance consequences. The gambling and adult-content cases illustrate the point: both produced significant harm yet would be difficult to address under CIB as currently defined, since the operational logic that made them effective is not reducible to coordination-plus-inauthenticity. Our reporting of these findings to Meta in October 2023 and November 2024 was met with acknowledgement but limited follow-up action — a response consistent with an enforcement category that does not readily recognise such operations as harmful in its terms. The recent scaling back of trust and safety investments at major platforms (Kaplan, 2025) suggests this gap is more likely to widen than narrow. The contribution of the DIOs framework, on the governance side, is to name the spectrum of practices that current enforcement categories cannot — a precondition for any enforcement response that aims to match the actual range of deceptive activity on the platform.

The research implications follow directly. Detection of DIOs resembles a persistent arms race in which the goal is not foolproof prevention but raising the operational costs for attackers while making discovery more likely. The automated, cyclical approach presented in this study contributes to that effort, enabling researchers and civil society to monitor coordinated deceptive activity at scale and complement platforms' moderation work. Sustained partnerships between independent researchers and technology companies — including timely platform responses to flagged operations — are essential if such tools are to keep pace with the activity they aim to detect. Without them, the integrity of the information ecosystem will remain precarious.

The contribution of this study, then, is twofold: an empirical demonstration that operations differing in actor, content, and intent share a common structural logic, and a conceptual frame that makes this logic visible to researchers and regulators alike. As social media continues to shape public discourse, the cost of leaving such operations outside our analytical and enforcement categories is unlikely to remain hypothetical.

Acknowledgements

This work has received funding from the European Union under the Horizon Europe vera.ai project, Grant Agreement number 101070093.

Biographical note

Giada Marino holds a Ph.D. from the University of Urbino Carlo Bo, where she is Tenure-track Assistant Professor. Her research focuses on the intersection of information disorder and political polarization, with a particular emphasis on how citizens engage with political content and discussions on social media platforms. She uses a mixed methods approach to analyse these dynamics. Her recent work is published in Platforms & Society, Information, Communication & Society and Polis.

Fabio Giglietto is a Full Professor of Internet Studies at the Università di Urbino Carlo Bo, where he earned his doctorate and teaches *Generative AI and Media* and *Digital Social Network Analysis*. His work sits at the intersection of computational social science, digital platform analysis, and political communication, with a primary focus on information disorders and the detection of coordinated behaviour on social media. His recent work, appearing in venues such as *Platforms & Society* and *M/C Journal* and in the *Routledge Companion to Social Media and Politics*, increasingly examines generative AI in influence operations, and algorithmic governance.

Anwesha Chakraborty is an independent researcher and project consultant working at the intersection of emerging technologies and society. She earlier was part of the University of Urbino team involved with the Horizon Europe project, Vera.ai. her latest publication is a research monograph published by Palgrave titled Designing Informative Technologies: How to Build Transparent, Reliable and Equitable Information for the Social Good (2026).

Massimo Terenzi, PhD, is a postdoctoral researcher at LUISS Guido Carli (MAEDINA project), investigating news avoidance and changing patterns of media consumption. He received his doctorate from the Università di Urbino Carlo Bo, where he contributed to the EU-funded PROMPT project by applying computational methods to map disinformation trajectories across digital platforms. His research sits at the intersection of media sociology, sociocybernetics, digital methods, and platform governance. He primarily focuses on information disorders, coordinated behaviour, algorithmic visibility, and the social consequences of digital regulation.

Samuel Olaniran is a lecturer at the University of the Witwatersrand. His research examines mis- and disinformation, information integrity, and the toxic online environments that increasingly shape public discourse, social values, and political life. His approach treats

digital platforms not merely as sources of data, but as cultural artefacts and spaces of meaning-making that require both computational analysis and humanistic interpretation. His research also explores the societal challenges posed by Geographic Artificial Intelligence, focusing on how spatial narratives are created, contested, and weaponised in an era characterised by growing uncertainty over truth and authority.

References

Allen, J., (2022). Misinformation Amplification Analysis and Tracking Dashboard. Integrity Institute. https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard

Al-Rawi, A., (2019). Viral news on social media. Digital journalism, 7(1), 63–79. https://doi.org/10.1080/21670811.2017.1387062

Bradshaw, S. & Howard, P.N., (2018). Challenging truth and trust: A global inventory of organized social media manipulation, Oxford Internet Institute. https://demtech.oii.ox.ac.uk/research/posts/challenging-truth-and-trust-a-global-inventory-of-organized-social-media-manipulation/

Broadwater, L., (2025). Trump's "flood the zone" strategy leaves opponents gasping in outrage. The New York Times. https://www.nytimes.com/2025/01/28/us/politics/trump-policy-blitz.html?unlocked_article_code=1.904.JWp_.6qWEPiGSsp2o&smid=url-share

Buller, D.B. & Burgoon, J.K., (1996). Interpersonal deception theory. Communication theory: CT: a journal of the International Communication Association, 6(3), 203–242. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2885.1996.tb00127.x

Chadwick, A. & Stanyer, J., (2022). Deception as a bridging concept in the study of disinformation, misinformation, and misperceptions: Toward a holistic framework. Communication theory: CT: a journal of the International Communication Association, 32(1), 1–24. http://dx.doi.org/10.1093/ct/qtab019

Charmaz, K. (2014). Grounded theory in global perspective: Reviews by international researchers. Qualitative inquiry, 20(9), 1074-1084.

Cresci, S. et al., (2016). DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection. IEEE intelligent systems, 31(5), 58–64. http://dx.doi.org/10.1109/MIS.2016.29

DiResta, R. & Goldstein, J.A., (2024). How Spammers and Scammers Leverage AI-Generated Images on Facebook for Audience Growth. arXiv. http://arxiv.org/abs/2403.12838

Donovan, J., Dreyfuss, E., Fagan, K., Faris, R., Friedberg, B., Holmes, C., … Salam, J. (2020). The Media Manipulation Definitions. The Media Manipulation Casebook. https://mediamanipulation.org/definitions

du Plessis, C. (2018, July 19–21). Examining the brandjacking phenomenon in the digital age. In Proceedings of the Communication, Mass Media & Society Conference (pp.

22–29). International Center for Research & Development (ICRD). https://www.researchgate.net/publication/328448876_Examining_the_brandjacking_phenomenon_in_the_digital_age

European Board for Digital Services & European Commission. (2025). First report of the European Board for Digital Services in cooperation with the Commission pursuant to Article 35(2) DSA on the most prominent and recurrent systemic risks as well as mitigation measures. European Commission. https://digital-strategy.ec.europa.eu/en/policies/dsa-brings-transparency

Fowler, G.A., (2016). What if Facebook Gave Us an Opposing-Viewpoints Button. Wall Street Journal. https://www.wsj.com/articles/what-if-facebook-gave-us-an-opposing-viewpoints-button-1463573101

Giglietto, F., Marino, G., Mincigrucci, R., & Stanziano, A. (2023). A workflow to detect, monitor, and update lists of coordinated social media accounts across time: The case of the 2022 Italian election. *Social Media + Society, 9*(3), 1–13. https://doi.org/10.1177/20563051231196866

Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020a). It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society, 23*(6), 867–891. https://doi.org/10.1080/1369118X.2020.1739732

Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020b). Coordinated link sharing behavior as a signal to surface sources of problematic information on Facebook. In *International conference on social media and society* (pp. 85–91). https://doi.org/10.1145/3400806.3400817

Gillespie, T., (2021). Custodians of the internet, New Haven, CT: Yale University Press.

Gleicher, N., (2018). Coordinated inauthentic behavior explained. https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/

Graham, T., (2024). The “inauthenticity” paradox: How platforms profit from and shape coordinated inauthentic behaviour. https://sites.google.com/uniurb.it/csbdetectionconf/home#h.te0xvcsx5g4q.

Guess, A., Nagler, J. & Tucker, J., (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science advances, 5(1), 1-8. http://dx.doi.org/10.1126/sciadv.aau4586

Huth, K., Peters, J. & Seufert, J., (2020). The heartland lobby. correctiv.org. https://correctiv.org/en/top-stories/2020/02/11/the-heartland-lobby/

Jack, C., (2017). Lexicon of lies: Terms for problematic information. Data & Society, 3. https://apo.org.au/sites/default/files/resource-files/2017/08/apo-nid183786-1180516.pdf

Jenkins, H., Ito, M. & Boyd, D., (2015). Participatory Culture in a Networked Era: A Conversation on Youth, Learning, Commerce, and Politics, John Wiley & Sons.

Kaplan, J., (2025). More speech and fewer mistakes. Meta. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

King, M., (2024). The Spaghetti Approach: How Trivial Misinformation Can Cause Significant Damage. https://www.linkedin.com/pulse/spaghetti-approach-mark-king-onnvc?utm_source=share&utm_medium=member_android&utm_campaign=share_via

Koebler, J., (2024). Facebook's AI spam isn't the "dead internet": It's the zombie internet. 404 Media. https://www.404media.co/facebooks-ai-spam-isnt-the-dead-internet-its-the-zombie-internet/

Messing, S., DeGregorio, C., Hillenbrand, B., King, G., Mahanti, S., Mukerjee, Z., Nayak, C., Persily, N., State, Bogdan, & Wilkins, A. (2020). Facebook Privacy-Protected Full URLs Data Set (Version 10). Harvard Dataverse. https://doi.org/10.7910/DVN/TDOAPG

Meta Platforms, Inc. (2025). Meta Content Library API version v5.0. https://developers.facebook.com/docs/content-library-and-api/citations.

Molina, M.D., 2023. Do people believe in misleading information disseminated via memes? The role of identity and anger. New media & society, *27*(2), 847–870. https://doi.org/10.1177/14614448231186061.

Muzumdar, P. et al., (2025). The Dead Internet Theory: A survey on artificial interactions and the future of social media. arXiv. http://arxiv.org/abs/2502.00007

Newman, N., Ross Arguedas, A., Robertson, C. T., Nielsen, R. K., & Fletcher, R. (2025). Reuters Institute digital news report 2025. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025

O'Donovan, K., (2024). Supporting the UK General Election in 2024. Google Blog. https://blog.google/around-the-globe/google-europe/united-kingdom/google-2024-uk-election-support/

Palys, T. (2008). Purposive Sampling. In The SAGE Encyclopedia of Qualitative Research Methods. SAGE Publications, Inc. https://doi.org/10.4135/9781412963909.n349

Prior, M., (2014). Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections, Cambridge, England: Cambridge University Press.

Quandt, T., (2018). Dark Participation. Media and communication, 6(4), 36–48. https://www.cogitatiopress.com/mediaandcommunication/article/view/1519

Rogers, R., & Righetti, N. (2025). Coordinated inauthentic behaviour on Facebook? A typology of manufactured attention. *Platforms & Society*, *2,* 1–13.

Romero-Vicente, A., Miguel, R. & Sessa, M.G., (2024). Coordinated Inauthentic Behaviour (CIB) detection tree, EUDL. https://www.disinfo.eu/publications/coordinated-inauthentic-behaviour-detection-tree/

Rubin, V. L. (2017). Deception detection and rumor debunking for social media. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 342–363). SAGE Publications. https://doi.org/10.4135/9781473983847.n21

Schroeder, D. T., Cha, M., Baronchelli, A., Bostrom, N., Christakis, N. A., Garcia, D., … Kunst, J. R. (2026). How malicious AI swarms can threaten democracy. Science, 391(6783), 354–357. https://doi.org/10.1126/science.adz1697

U.S. Senate Select Committee on Intelligence., (2019). Russian Active Measures Campaigns and Interference In the 2016 US Election, 116th US Senate Congress. https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures

Silverman, C. & Alexander, L., (2016). How Teens In The Balkans Are Duping Trump Supporters With Fake News. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo

Starbird, K., Arif, A. & Wilson, T., (2019). Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. Proc. ACM Hum.-Comput. Interact., 3(CSCW), 1–26. https://doi.org/10.1145/3359229

Starbird, K., DiResta, R. & DeButts, M., (2023). Influence and Improvisation: Participatory Disinformation during the 2020 US Election. Social media + society, 9(2), 1–17. http://dx.doi.org/10.1177/20563051231177943

Vargo, C.J., Guo, L. & Amazeen, M.A., (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society, 20(5), 2028–2049. https://doi.org/10.1177/1461444817712086

Wardle, C. & Derakhshan, H., (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking, Council of Europe. http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

Weedon, J., Nuland, W. & Stamos, A., (2017). Information Operations and Facebook. Meta Newsroom. https://i2.res.24o.it/pdf2010/Editrice/ILSOLE24ORE/ILSOLE24ORE/Online/_Oggetti_Embedded/Documenti/2017/04/28/facebook-and-information-operations-v1.pdf

Appendix 1

Table A1. Codebook for Deceptive Information Operations (derived from grounded coding).

Dimension	Emergent Categories	Description of the Category and/or Indicative Examples
Identity management	Account name changes	Accounts using fluid identities: The group DSTV Premier League Soccerzela changed names 13 times.
	Thematic incoherence	Thematic changes over time or theme declared mismatch with content published. E.g., The Golden Girls Fan Club group was previously linked to the Pakistan Peoples Party (PPP). It later rebranded as a fandom group.
	Indicators of administrator identity	Nationality and profile information of the administrators. For example, ORION STARS, JUWA, GAME VAULT, FIRE KIRIN, MOOLAH - FREEPLAY admin accounts have generally low engagement, few followers, and little content. Many label themselves as “digital creators,” yet their profiles lack meaningful activity.
	Descriptive quantitative data	Number of accounts in the network, Groups members, Pages followers, Number of admins, Creation date, etc.
Behavioral coordination	Bursts of identical posting	WorldRusWorld posted an identical pro-Putin image multiple times.
	Posting near-duplicated content in a short time window	The near-duplicated pro-Putin posts were shared by WorldRusWorld within a three-minute window (16:21–16:24 CEST).

Dimension	Emergent Categories	Description of the Category and/or Indicative Examples
	Cross-posting among accounts	WorldRusWorld posts the same content among 10 different groups.
	Use of cloaked or masked links/posts	Use of link shorteners or activity-status posts to link unrelated pages.
Content manipulation	AI-generated visuals	Synthetic or manipulated imagery designed such as “Shark” casino image or heroic Putin portraits.
	Brandjacking	Unauthorized use of known brand names to imply legitimacy. “FREE PLAY ALL GAMES (JUWA CITY ORION STARS / FIRE KIRIN).”
	Narrative mixing	Combination of benign and misleading materials that obscure intent. For example, lifestyle imagery mixed with anti-Ukraine propaganda.
	Unrelated content mixing	Use of innocuous blurbs to mask the content of the post or the external destination such as in the case of sexually explicit material circulated with the BMW history blurb.