Discover Global Society

Research

The disinformation lifecycle: an integrated understanding of its creation, spread and effects

Kimberley Kruijver[1] · Neill Bo Finlayson[2] · Beatrice Cadet[3] · Sico van der Meer[2]

Received: 3 February 2025 / Accepted: 20 May 2025
Published online: 16 June 2025

Abstract
The proliferation and development of social media platforms in recent years has contributed significantly to the spread of disinformation. Police Authorities around Europe have observed that harmful or criminal behaviour, stemming from social unrest, hate speech, and violent disorder are regularly preceded by disinformation campaigns. This begs the question: How can practitioners be better prepared for the real-world consequences of malign disinformation activities and to potentially even mitigate any criminal consequences? The first step in properly countering disinformation is to enhance the understanding of the complex phenomenon. Therefore, this article puts forth a new theoretical framework, called the 'C5 Interaction Model', that explains the creation, spread and impact of disinformation, synthesising academic theory to provide practical guidance on disinformation dynamics. The multidisciplinary model represents a lifecycle and contains five main elements: Context, Causes, Content, Consequences, and Cycle of Amplification. They are each organised into two further layers of (sub)factors, which were developed to provide a comprehensive overview and breakdown of the important elements of disinformation. The C5 Interaction Model represents one of the first concerted efforts to bring diverse insights together into a comprehensive integrative framework. The complexity of the model shows that this process is non-liner and that there are a multitude of factors determining the lifecycle of disinformation, making it a highly complex phenomenon to research. A key contribution of this article is the focus on the interaction between different elements that influence the process of disinformation—from creation to consequences. Importantly, the lifecycle route is predominantly influenced by the social context in which it exists.

1 Introduction

The rapid development of social media platforms and online social networks in the past decade has changed the way people communicate with each other. Users rely on these tools to share information, connect with other people, and stay informed about trending events. Despite the potential benefits of this, social media platforms have also contributed to an explosive growth in the amount of false and inflammatory information being spread in the world. Nowadays, the presence of such falsehoods online—referred to as disinformation from hereon—is not only disruptive or distracting for everyday users, but it can also have an impact on the individual and behavioural level, contributing to harmful or even criminal behaviour on the part of the receivers of the disinformation messages. For example, during the COVID-19 pandemic, false information was widely shared that undermined trust in public institutions across several countries. This not only led to (sometimes unlawful) protests, but also to vandalism of, for example, critical infrastructure [1]. Research

✉ Kimberley Kruijver, kimberley.kruijver@tno.nl; Neill Bo Finlayson, neill_bo.finlayson@tno.nl; Beatrice Cadet, Beatrice-cadet@hotmail.fr; Sico van der Meer, sico.vandermeer@tno.nl | [1]Netherlands Organisation for Applied Scientific Research (TNO), The Hague, Netherlands. [2]TNO, The Hague, Netherlands. [3]Air France KLM, Paris, France.

into the German 'Reichsbürger' movement showed that exposure to disinformation narratives can even impact an individual's justification of the use of violence against the out-group [2]. Disinformation can thus have a profound impact on both individuals and wider society.

Firstly, it is important to explain what is meant by disinformation. The European Union (EU) High Level Group on Fake News and Online Disinformation defines it as "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit" [3, 4]. Debate continues in academic literature about the definition, conceptual boundaries and relevance of this distinction between mis- and disinformation. However, a practical distinction may be that disinformation is generally created and spread for economic gains or for political or ideological goals, and it can be exacerbated by how different audiences and communities receive, amplify, and engage with disinformation [3]. This is different to falsehoods being spread in good faith by people concerned about health risks, vaccines or technology, for example. The risk of disinformation includes threats to democratic political processes and values, which can specifically target a variety of sectors, such as health, science, education, finance and more. What makes disinformation especially harmful is that it can take the form of both a single message as well as multiple interrelated messages, such as a targeted disinformation campaign.

Police Authorities (PAs) have observed that criminal activities such as hate speech, violence, riots, and terrorist attacks are regularly preceded by disinformation campaigns. An infamous example is the Capitol riots in the United States (US) in January 2021. A disinformation campaign that increasingly gained attention on Facebook, claimed that former President Donald Trump was working against a so-called 'criminal deep-state' [5]. The disinformation alleged that members of this deep state, elite associates of the US government, media, and business, were actively trying to undermine him. Within this context, it was claimed that President-Elect Joe Biden had 'stolen' the election from his opponent (and then President) Donald Trump by election fraud. This narrative eventually led to a crowd of rioters storming the Capitol building in Washington DC during a joint session of Congress to formalise Biden's election victory. Five people were killed in the riots, including a police officer, and many more people were injured, including 135 police officers. Afterwards, questions arose about whether the police were adequately prepared for the riots [6].

The impact of disinformation in stoking public disorder and violence in the US has been felt across the world. Based on interviews with five European PAs, a shared need was formulated to improve readiness for the harmful or even criminal effects of disinformation on the local level.[1] At present, PAs tend to be short on staff with relevant knowledge on the disinformation process, which is fundamental when seeking to counter its damaging effects on social order, safety and security. Therefore, the first step in aiding not only these PAs, but practitioners tasked with handling the real-world consequences of disinformation more generally is to enhance their understanding of the complicating dynamics around disinformation, grounded in academic theory, while at the same time providing practical guidance.

This article was written within the context of the EU funded research project VIGILANT which aims to develop an integrated platform of advanced disinformation identification and analysis tools to cover disinformation from major sources.[2]

1.1 Literature review

Attempts have been made to provide guidance and clarity on the threat of disinformation for practitioners, who are often tasked with dealing with its real-world impact. For example, the RAND Corporation published guidelines for PAs on how to combat *misinformation* particularly around the COVID-19 pandemic [7], while the EU issued a handbook to provide local and regional authorities with recommendations on how best to counter disinformation [8]. These publications focus on countering mis- or disinformation, but provide little understanding of its underlying mechanisms, how it spreads, and its consequences. Furthermore, the guidance offered in these publications is rather practical and therefore does not describe the necessary insights from academic literature to adequately understand the behavioural and cognitive drivers of disinformation. An academic handbook on disinformation, edited by Arcos, Chiru and Ivan [9], does offer such insights into the dynamics of disinformation. However, for all that it offers in knowledge on the subject, it in turn misses the required accessibility and applicability for use by practitioners. As such, this provides an opportunity to build

[1] These interviews were carried out within the context of the VIGILANT project.

[2] The EU VIGILANT project aims to address these issues by developing an integrated platform of advanced disinformation identification and analysis tools to cover disinformation from major sources, in all modalities, and in multiple languages. Functioning as a thorough theoretical base of the complete VIGILANT project, this article develops a conceptual framework of disinformation, which provides conceptual input for the technological tools that are developed in other work packages of the project. The model may also help prevent overlooking relevant aspects of disinformation in later phases of the project. For more information about VIGILANT, see: https://www.vigilantproject.eu/.

on this work by developing accessible and practical guidance on the dynamics of disinformation that strikes a balance between theory and practice.

There are numerous frameworks and models that attempt to make sense of the disinformation phenomena in academic literature. Owing to the relative novelty of this field of research, there is a strong line of research on developing taxonomies and typologies for concepts relating to disinformation (e.g. [10]), while others have sought to provide frameworks for the multimodal nature of disinformation content [11]. Indeed, much of the recent academic research on disinformation has focused on the technical aspects of detection, whether it is employing language models to design multimodal detection systems (e.g. [12]) or more theory-driven detection models that analyse writing features of content (e.g. [13]). Furthermore, steps are being taken to develop more sophisticated computational [14] or automated [15] frameworks and methods for detection. Although these are useful technical solutions for detecting and analysing disinformation content, they do not provide a satisfactory theoretical explanation of the dynamics that underpin the creation, spread and effect of disinformation.

One of the most comprehensive frameworks on disinformation is DISARM (DISinformation Analysis & Risk Management): an open-source framework designed to identify and counter disinformation [16]. It provides both a conceptual and practical understanding of disinformation by comprising a structured methodology for identifying, analysing, and mitigating the tactics and techniques used by proponents of disinformation [17]. Similar to this, Kozyreya et al. [18] developed a toolbox of individual-focused interventions which provides a comprehensive conceptual overview of how to counter *misinformation*. Both frameworks achieve an effective synergy of theory and practical guidance. However, the focus remains on the *what* (detecting and countering disinformation) and not the *why* (underlying mechanisms of disinformation spread). An understanding of the latter is precisely what practitioners need in order to more effectively carry out the former.

There is a considerable body of research in the social and behavioural sciences that provides frameworks for analysing the underlying drivers of disinformation. For the most part, however, these relate to the dissemination or diffusion of disinformation in online social networks (e.g. [19]), often employing computational methods (e.g. [20]). That said, Froehlich [21] developed a framework that gives insight on the creation, dissemination and effects of disinformation. Yet, he focuses specifically on the legitimation and sustainment of disinformation,how people's own critical thinking and the power of cognitive authorities can sustain belief in 'fake news'.[3] Similarly, Arayankalam and Krishnan [22] attempted to synthesise this rather fragmented body of research on disinformation through a systematic literature review, focusing particularly on the "psychosocial antecedents of its spread" and subsequent impact. The authors conclude, based on their review, that further research is needed to establish a much-needed theoretical and methodological grounding for understanding disinformation, highlighting the importance of taking into account behavioural, social and environmental factors [22].

The framework developed by George, Gerhart and Torres [23] does provide a comprehensive multi-disciplinary theoretical grounding for the underlying dynamics of the creation (including the underlying motives) and subsequent spread of disinformation for 'fake news research'. The authors focus on (1) distinct relationships that cause the message to accelerate, perpetuate and eventually cause societal impact,and (2) the roles that different actors—like creators and consumers—play. Although this framework helpfully illustrates the main elements of a general disinformation process, some important gaps were identified. The importance of context in the process was lacking. According to Hameleers [24], context is key when it comes to understanding the actors, intentions, and techniques behind disinformation. For instance, it is important to consider the role that different threats, political challenges, media ecosystems and specific events can have on the spread and impact of disinformation such as, for example, the current hybrid threat environment [25].[4] Moreover, the description and framing of the real-world (often criminal) implications of disinformation in George et al.'s [23] framework miss the nuance and detail required to adequately explain the escalatory reinforcing relationships between the factors that underpin the dynamics of disinformation.

Evidently, multiple disinformation frameworks exist from various disciplines such as communication studies, psychology, political science, sociology, and computer science. This plethora of research carries the risk of fragmentation, as researchers often focus on isolated aspects of the problem. There is a gap in the current literature on disinformation theory for a truly multidisciplinary conceptual or theoretical framework for understanding the causes, contents and

[3] Disinformation is often referred to as 'fake news', but this is an inadequate term because disinformation often involves content that is not completely 'fake' but fabricated information blended with facts and practices that go well beyond anything resembling 'news' [3]. Nevertheless, the present article includes the review of literature using the term 'fake news' in addition to other terms for disinformation precisely because it is often incorrectly used as a term for disinformation in general.

[4] For more information on context, see chapter 3.1 Context.

consequences of disinformation that incorporates behavioural, cognitive and communication science perspectives. Furthermore, as mentioned previously, there is a need for such a framework to be practicable and applicable for laymen or practitioners, particularly those on the front line of safety and security work. Given these gaps, the purpose of this paper is to develop a comprehensive framework that explains the creation, spread and impact of disinformation, synthesising academic theory to provide practical guidance for practitioners.

1.2 Research problem

There is need for greater understanding, knowledge and expertise about disinformation within the authorities tasked with dealing with the real-world consequences of the phenomenon. Based on the practical and theoretical discussions as described above, this article sets out to answer the following research question: "Which conceptual elements constitute the disinformation process that can escalate to harmful or even criminal behaviours?". It aims to contribute to (1) practice, by aiding practitioners to understand and thereby being able to respond adequately to disinformation, as well as to (2) theory, by developing a conceptual model that demonstrates the relationships between the different elements of disinformation and how its impact is affected by diverse factors.

The following chapter explains the research methodology. In Chapter 3, the developed 'C5 Interaction Model' is presented, including a detailed description of its various elements. Chapter 4 discusses the findings, categorised along two lines: its academic contribution and its practical relevance, as well as some limitations. The article ends with Chapter 5, which includes the conclusions of this research report.

2 Methodology

With the aim of collecting as many key elements of the disinformation phenomenon as possible, the developed C5 Interaction Model (discussed in Chapter 3) is primarily based on insights acquired from existing academic literature from various disciplines, including communication and media studies, sociology, and psychology. The first step in the literature review was to build the basic foundations of the model, such as the main factors in the process of disinformation and how they relate to each other.[5] To limit the number of articles to be scanned, the focus of this part of the review was on existing literature reviews and meta-analyses because they already offer a comprehensive review of a certain academic field and ensure a state-of-the-art analysis. This methodological choice can be seen as a limitation because already synthesized data was re-analysed again. However, due to time constraints this was still preferred and the newly combined data was nonetheless able to create something new. Keywords such as 'review disinformation causes', 'disinformation review content' and 'fake news riots' were used to search relevant literature; mainly via Google Scholar. Moreover, when considered useful, stand-alone academic papers and related literature were also included through snowballing, a common activity in literature research that consists of identifying new articles quoted in the papers already analysed. The list of academic articles includes purely theoretical papers as well as more practical ones, including case studies.

The data was then analysed using the Grounded Theory Method [26]. This method ensures that new theories or models are built based on the systematic collection of data. By using a coding system, the literature was thematically assembled into groupings of articles related to the same aspect of disinformation. For instance, an article providing insights on the platforms used to create and/or spread content was coded under "Channels", which was subsequentially categorised under the code "Content". This thematic analysis allowed the authors to break down and group different aspects of the collected literature. To measure inter-rater reliability, the coding was carried out in two main phases, whereby the codes were checked by another author and then confirmed or discussed in order to adjust where appropriate. While no formal inter-rater reliability coefficients (e.g., Cohen's Kappa) were used, coding was iteratively refined through collaborative discussions aimed at achieving consensus, following principles of reflexive thematic analysis [27]. Coding was also facilitated by corresponding it with five initial use cases[6] provided by the PAs to the VIGILANT project, which ensured the coding remained relevant for practical application. A questionnaire was created and organised into four parts: (1)

[5] 51 articles were analysed for specifically the conception of the model. For reporting purposes, a total of 120 articles were used, including the 51 of the initial literature review. Additional references were for instance added to justify the methodology or the background information of the completed work.

[6] Five PAs are a members of the VIGILANT project: Catalonia (Spain), Estonia, Germany, Greece and Moldova. Each of them provided a use case of a realistic scenario in which they encountered disinformation.

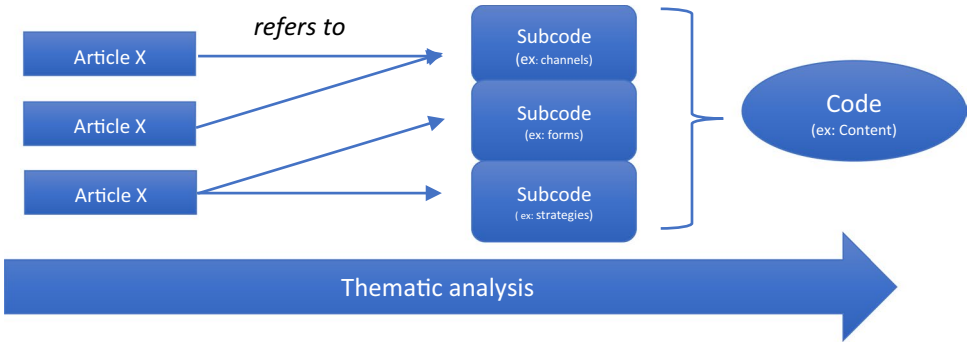

Fig. 1 The coding process

description about the use case, (2) questions about relevant factors, (3) the relationship to criminal behaviour, (4) the investigative process, and (5) the (technical) support they would have wanted. In addition, two use cases in collaboration with the Catalonian police were written following the same questionnaire as well as three additional use cases were created based on open-source information about (1) The Dutch Willem Engel Court case, (2) The US Capital riots, and (3) a New Zealand court case on disinformation. These ten distinct and structured use cases fed information to the authors' understanding on different disinformation processes. This iterative process, completed as part of the Grounded Theory Method, allowed the authors to gradually construct the conceptual model. Once this had been completed, steps were taken to validate the model by applying it to a disinformation use case developed in conjunction with the Catalonian Police within the VIGILANT project.

The main codes became clear quite early in the analysis process as articles and codes were thematically grouped, and later also became the main elements of the C5 Interaction Model: Context, Causes, Content, Consequences, and Cycle of Amplification. Causes, Content, and Consequences had already been suggested by the experts engaged in the proposal phase of the VIGILANT project, which were reflected in meta-analysis articles like George et al. [23] and Arcos et al. [25]. As a result of the iterative literature review, the codes Context and Cycle of Amplification were identified as main contributing factors. First, Context was added as multiple articles referred to contextual factors such as social context [28], the hybrid threat environment [25] and national policy and legal differences with regards to disinformation [29]. Since the context is often neglected in disinformation research [22], even though it is vital in influencing the path of disinformation messages, it is one of the two codes that were chosen to emphasize more. The second code is the Cycle of Amplification. This was mainly inspired by George et al. [23], who created a framework to help drive future research on disinformation. After finding what George et al. [23] refer to as the cycle of amplification, the authors recognized (parts of) that element in other articles as well. For example, when Zhang and Ghorbani [28] refer to social context, they actually argue that it heavily influences to what extent messages are disseminated, hence to what extent it is amplified.

Once the basic foundation of the model was constructed (i.e. the five 'Cs' and the general process of disinformation), a second literature review of empirical studies was then conducted to identify, develop and explain the subfactors for each main factor. This was necessary because although the meta-analyses provided useful broad overviews of the phenomena of disinformation, a more granular approach was needed to better understand and describe the specific mechanisms of each sub-factor and why it is relevant to the overall model. For instance, under the main code Causes, the following subcodes were used: Creators and Motives. In turn, they could have their own subcodes as well. For example, under the subcode Motives was a lower level of codes: Political and Financial. Although a distinction was made between the use of overarching literature reviews and empirical studies in this process, there is some overlap between the two types of literature reviewed for this paper in explaining and supporting certain subfactors. This was done in order to simultaneously provide the necessary level of detail required to understand the specific mechanisms of disinformation in the subfactors, while also ensuring it is understood within the wider context of disinformation and its impact on society. Therefore, when explaining and justifying (sub-)factors in the main body of this paper, theoretical and empirical evidence is at times intertwined.

By using this layering in the coding, a useful overview and step-by-step breakdown of the important elements of disinformation could be identified. Moreover, the model was refined by turning the analysis around; the authors zoomed out, undertook a high-level analysis of all the subcodes breakdown and were able to detect the overarching themes: the main C5 elements and subsequent factors could then be either re-affirmed or altered (Fig. 1).

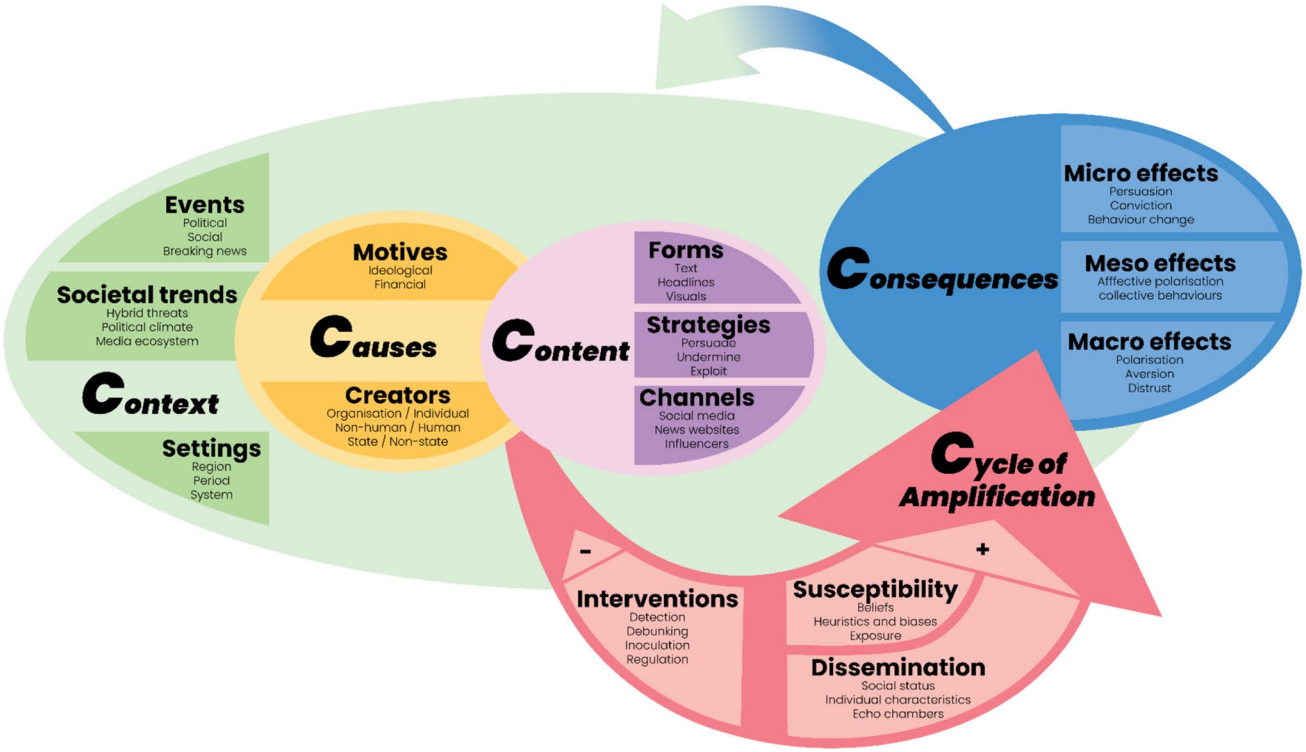

Fig. 2 The C5 interaction model

The main codes correspond to so-called 'C5 elements' in the model, while the subcodes correspond to factors and subfactors. At times, it was necessary to rework multiple aspects of elements based on the authors' own discretion, which itself was based on findings in the literature. The extent to which codes and subcodes were cross-referenced and their occurrence in the literature also influenced whether they were integrated into the model. As the figure above demonstrates, one article could be relevant for multiple (sub-)codes, as it can discuss different topics. On the other side, some initial subcodes were deleted because they were not reflected strongly enough in the literature. Finally, it should be noted that the list of factors is not exhaustive, and their descriptions are not all-encompassing because disinformation, along with the online information environment, is constantly evolving and highly dependent on contextual circumstances. This article focuses on the most important and thematically relevant factors and their interrelationships.[7]

3 The C5 interaction model

The coding of relevant literature highlighted five main aspects of disinformation that constitute the C5 elements: Context, Causes, Content, Consequences, and the Cycle of Amplification. Although the role of the Cycle of Amplification has been identified in multiple recent articles, emphasising the importance of it to the development of disinformation is one of the main contributions of this model. This model (see Fig. 2) outlines the interactions and interrelationships between these five factors and its subfactors. The model also seeks to emphasise that, due to the nature of the relationships between these factors, the same piece of content can have different consequences, depending on which factors are at play and to what extent. This is important when trying to develop an understanding of how disinformation campaigns can lead to harmful or even violent consequences.

The C5 Interaction Model (shown in Fig. 2) depicts the five elements that play a role when exposure to disinformation content leads to cognitive and behavioural effects:

[7] All initial, the way they were found and subsequently coded articles for the conception of the model were organised in an Excel sheet which will be made available upon request.

1. The Context (social, cultural, political, or economic factors, important events, or relevant trends).
2. The Causes (creators and their motives).
3. The Content (the tailored piece of disinformation).
4. The Consequences (short and long-term effects on the consumer and society).
5. The Cycle of Amplification (the interaction of receiver susceptibility, dissemination factors, and the possible interventions to counter the cycle).

The following sections describe each of the main elements of the C5 Interaction Model, its factors and subfactors, in more detail. Each section starts with a visualisation to present an overview of the breakdown of each element. The corresponding factors and subfactors are only meant to display that they are grouped together in different layers: each element (C in the darkest colour), can be categorised in a layer of broad factors (slightly lighter colour), which in turn can be categorised into more detailed subfactors (lightest colour). For instance, looking into C1 Context could mean analysing 'Setting' which can be done by investigating the 'region' and/ or 'period'. This list is indicative and non-exhaustive. It covers the main elements of disinformation. Moreover, many interrelations exist between the factors and subfactors. The most relevant ones are described in the following sections.

3.1 Context

As previously discussed, disinformation is considered to be a 'context-bound phenomenon'; which is to say that the context in which the disinformation exists is fundamental to understanding the actors, intentions, and techniques behind the manipulation [24]. Indeed, as argued by Zhang and Ghorbani [28], social context is a crucial determinant of the extent to which messages are disseminated, to what extent they are amplified and what effects they may have. Despite this, however, the importance of context is often overlooked in academic literature on disinformation [22]. Context refers to anything related to the social, cultural, political, or economic setting or environment, including important events or relevant trends, in the wider society (see Fig. 3).

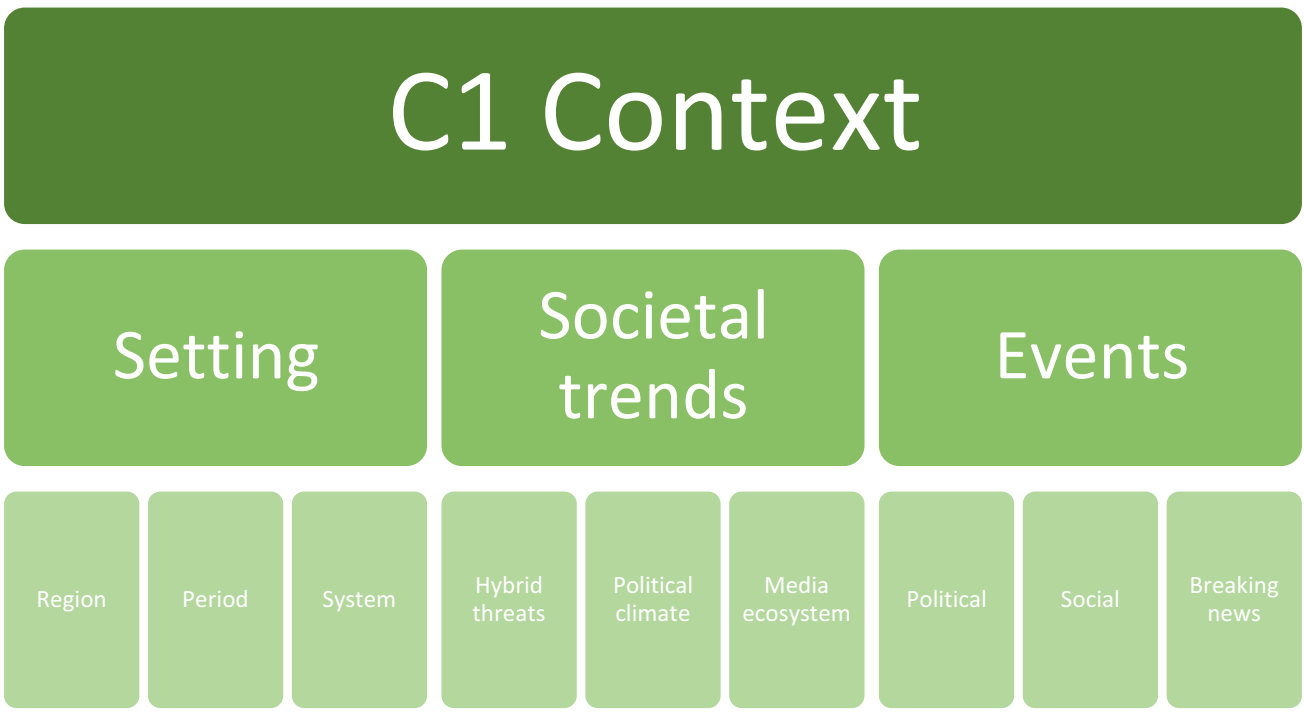

Fig. 3 A breakdown of the C1 element context

3.1.1 Setting

This is perhaps the most important contextual factor that influences the potency of disinformation is the setting in which it is sent and received. The geographical region or period in which a disinformation campaign is waged will determine what the disinformation looks like, what it says, how it is presented, and who it targets. Setting, therefore, intrinsically influences the creation, dissemination, amplification, and effects of disinformation.

The specific *region* in which the disinformation is disseminated is a fundamental subfactor to consider. As Humprecht [30] demonstrated, the content and style of disinformation changes from country to country. For example, online disinformation in the US and the United Kingdom (UK) is predominantly politically partisan, whereas in Germany and Austria sensationalist stories predominate over political content [30]. This can also be explained by the unique political systems in the US and the UK, discussed in more detail below for political system. Furthermore, Humprecht [30] found that online disinformation in English-speaking countries tends to target political actors, whereas in German-speaking countries, the main focus is immigrants. As such, other region-specific indicators are important to consider when assessing the contextual conditions of disinformation, such as how resilient a society is, levels of populism, polarisation, media trust, time spent on social media platforms, media literacy, and the strength of the public broadcasting service. Furthermore, the cultural context of a region can be a useful determinant of how effective disinformation will be, how susceptible people are and the extent to which it can spread (see Chapter 3.1.5).

Aside from regional differences, the specific *period* in time in which disinformation is disseminated is an important contextual subfactor. Overall, disinformation campaigns that are being waged today can be considered to take place in the so-called disinformation or 'post-truth' age [31] which refers to the present era of corrupt information environments [32]. Within this era, more contentious periods take place that feed disinformation, like the COVID-19 pandemic [7]. The WHO (World Health Organisation) even called the increase in disinformation due to COVID-19 an international 'infodemic'. Since there was a lot of insecurity around COVID-19, which scientists could not provide an immediate response to, it was fertile ground to create and spread disinformation [33].[8] In response, the WHO launched a special platform for policymakers, academics and public health professionals to come together with media organisations, social media platforms and civil society to devise a new framework on how to deal with the COVID-19 infodemic [34].

Perhaps the most important subfactor for setting is the political *system*. This broadly refers to the formal, constitutionally enshrined institutions, processes and structures that constitute a state and its political order [35]. This is distinct from the subfactor political climate which refers to more temporal political discourses at a particular time, rather than the more static and permanent regime of institutions and processes that comprise a political system.[9] As comparative research shows, the type of political system (non-democratic versus democratic regimes) is a strong determinant of the likelihood of disinformation spread [36]. Such research on different types of democracies specifically is, unfortunately, scarce.[10] Nevertheless, it should be acknowledged that, for instance, the rigid two-party system used in the US, and to a lesser extent in the UK, is unique compared to other advanced democracies in Europe which generally operate under a multi-party system. Similarly, the UK's first-past-the-post electoral system is also unique compared to the proportional representation systems favoured by other European states. Therefore, the way in which disinformation manifests, spreads and takes effect may be vary depending on the particular type of democratic system.

3.1.2 Societal trends

Similar to the setting, societal trends also play a significant role in determining the nature of disinformation. The concept of Societal Trends refers to a more specific, temporal set of trends in society that can influence how disinformation is created, disseminated, amplified, and takes effect. For instance, although EU countries may have a similar setting (region and period), each country—and within that, each social group or political class—will be facing different types of threats, be exposed to different political challenges, and be part of different media ecosystems, thereby influencing the nature of disinformation to which they are exposed.

[8] More about the amplification process can be found in C5 3.1.5 Cycle of Amplification.

[9] More about political climate can be found in C1 3.1.1.2.5.

[10] However, considerable research has been conducted measuring phenomena such as polarisation in different political contexts, such as European multi-party systems (e.g. [132]).

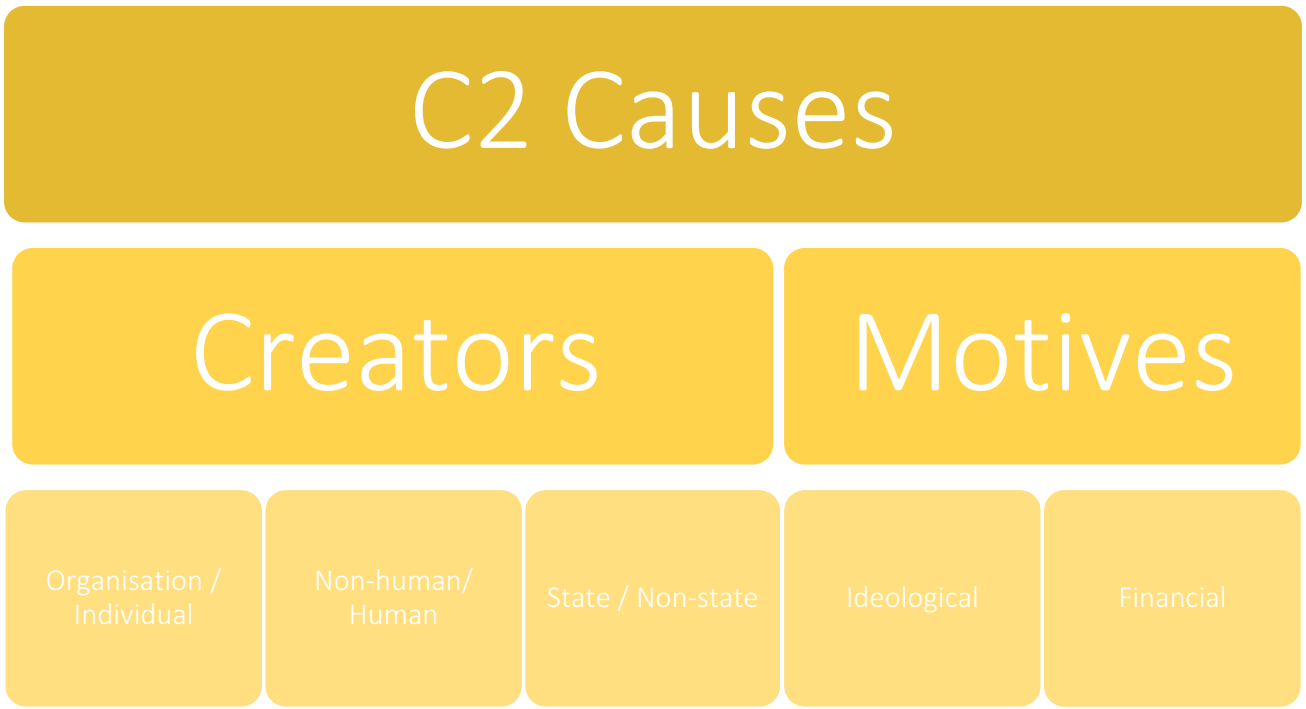

Fig. 4 A breakdown of the C2 element causes

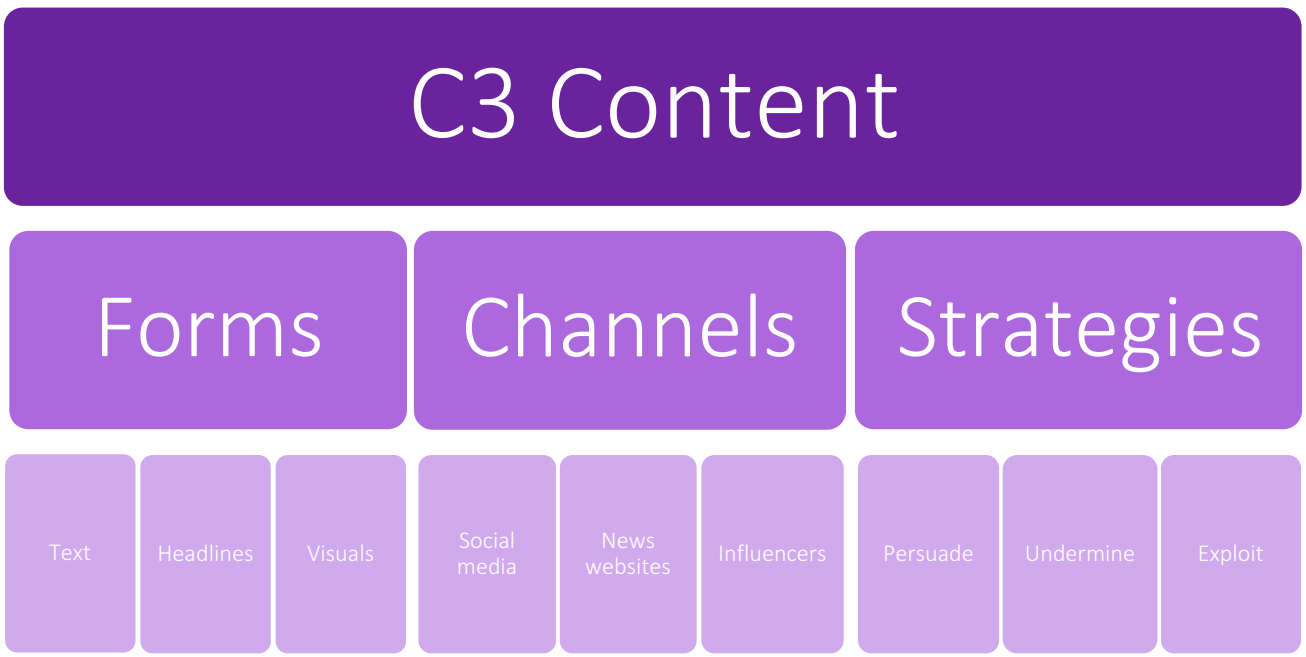

Fig. 5 A breakdown of the C3 element content

The targets of disinformation are often unaware that such illegitimate uses of digital media are in fact related to wider *hybrid threats* and information warfare activities [25]. Hybrid threats refer to activities of an adversarial that are difficult to detect, attribute and hence to counteract and remain below the threshold of war. In that context disinformation campaigns are popular to sow discord amongst populations through for example proxies to make it difficult to see the true source. In recent years, foreign actors and their proxies have begun weaponising information to fulfil their nefarious motives and forcefully alter public beliefs or perceptions about a certain person, event, or body [37]. This is often referred

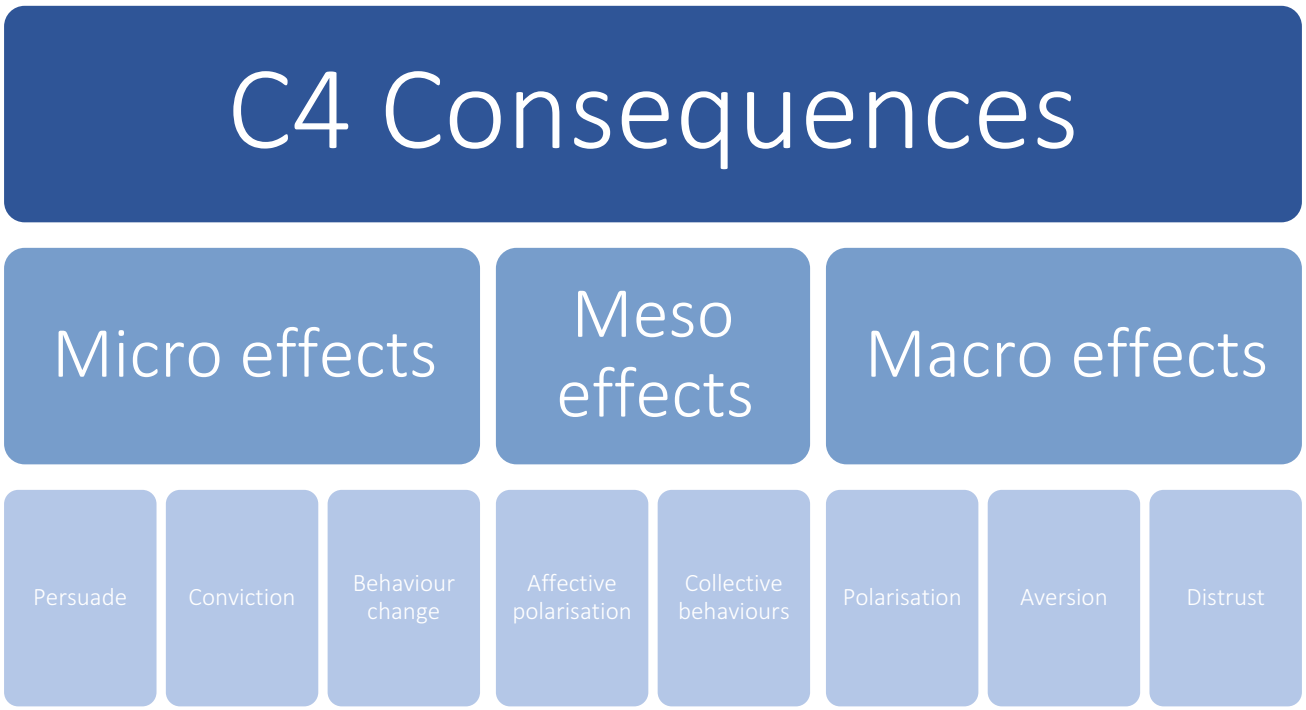

Fig. 6 A breakdown of the C4 element consequences

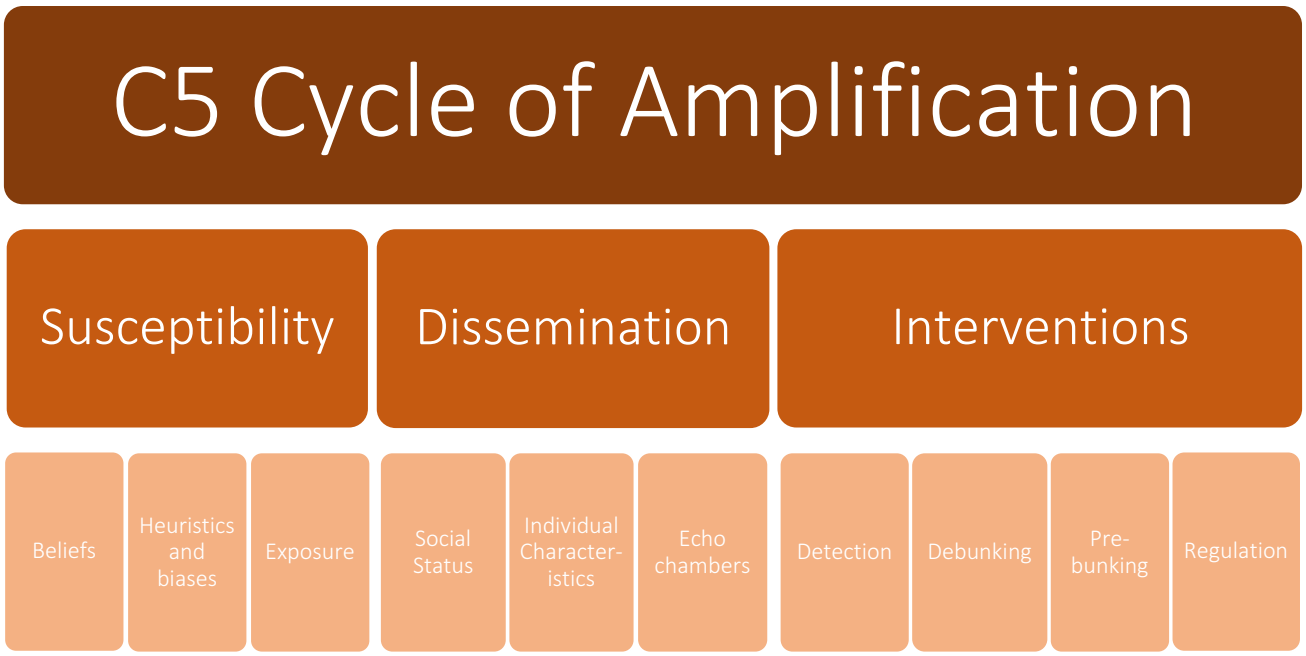

Fig. 7 A breakdown of the C5 element cycle of amplification

to as Foreign Information Manipulation and Interference (FIMI) [38]. Extensive research shows that hybrid actors, such as Russia, exploit digital media and online environments to test and deliver their messages—a growing trend that poses a major security threat to democracy [25].

Another societal trend that can influence the nature of disinformation is the *political climate* of any given context. This refers to the temporal nature of political discourse and debate at a particular moment in time in a particular political setting. For example, research shows that in Portugal, one of the main targets of online disinformation during the 2019

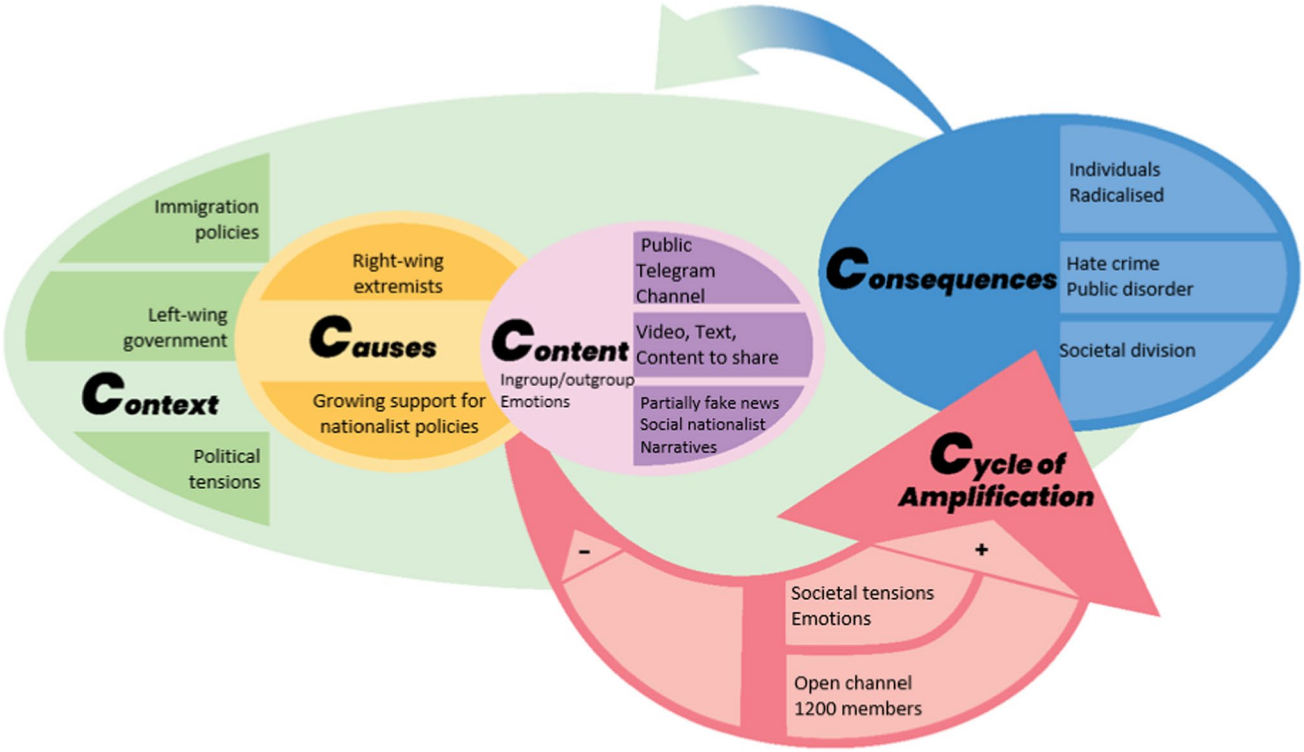

Fig. 8 C5 Interaction model applied to a use case

election was corruption [39]. Corruption was the most pressing political issue at the time with the former Prime Minister, José Sócrates, being accused of corruption and tax fraud which meant that the presiding Prime Minister, António Costa, who was seeking re-election, became the main target for disinformation campaigns related to corruption [40]. The political climate was clearly an important contextual factor that determined the content of disinformation during those elections.

The present-day **media ecosystem** has helped facilitate the "extensive, unrestrained and sometimes systematic distribution" of disinformation [41]. Bokša [42] demonstrates how in the case of Russia disinformation moves through a so-called "amplification pyramid" in which state-run media initiate messages intended for circulation. These messages are then distributed by experts or political representatives, after which the disinformation moves into disinformation websites, online social networks and to its users, trolls and bots. Bokša [42] argues that disinformation slowly transforms to misinformation, as people unknowingly are spreading false information.[11] As societies continue to move towards digital dependency, it is more difficult to find trustworthy media sources (Alcott and Gentzkow 2017). It is argued by some that we are living in a 'post-factual' information setting [43]. This fragmented media ecosystem presents serious challenges to countering disinformation, mostly due to the sheer volume and breadth of the reach of disinformation which makes it hard for fact-checkers and news organisations to keep track and debunk [25].

3.1.3 Events

Disinformation is often predicated on the occurrence of specific events that hold a certain significance to the target audience. Such events draw attention from the media and the audience and are often surrounded by ambiguous information—ambiguity that purveyors of disinformation seek to exploit so they can dominate the narrative and frame the event in a way to further their agenda [38]. This is effective in stoking divisions as it taps into the so-called 'hostile media effect' which is the phenomenon whereby people on opposite sides of a divide perceive identical news coverage of an event as biased against their side [25, 44].

[11] More about the amplification process can be found in C5 3.1.5 Cycle of Amplification.

Political events such as elections, referenda, or protests, are often the focus of disinformation campaigns. Research shows that disinformation spikes during election cycles [24, 45], famous examples of which may be the 2016 US presidential election [46], the 2016 Brexit referendum in the UK [47] and more recently the EU parliamentary elections [48]. Political protests can also act as a trigger for disinformation campaigns, for example, when the Chinese state upped its disinformation output to discredit the 2019–2020 protests in Hong Kong [49]. This subfactor refers to specific political events that can be used as leverage for disinformation campaigns, rather than the political mood or climate of a country, which is more abstract, affective and fluid.[12] (political climate is covered above in Chapter 3.1.1.2).

Similarly, *social events*, such as public, national, or religious holidays, celebrations, or festivals, can act as triggers for disinformation campaigns. For instance, Russian disinformation outlets sought to stoke fears over the 'Islamification' of Europe by falsely claiming that Christmas lights were being banned by the EU to appease immigrants and that snowmen were banned because they are a racist symbol [50]. Commemorations of 9/11 have also become trigger events for disinformation campaigns [51, 52].

Other prime targets for disinformation are *breaking news events*. The most prescient example of this was the initial outbreak of war in Ukraine. In its research on Russian disinformation, the European External Action Service [38] documented various key incidents in the Ukraine war and mapped them to related disinformation campaigns instigated by the Kremlin. Cases of disinformation were often in response to breaking events in the conflict, but also often pre-empted conflict events in the hope of framing the events in a certain light before they took place [38]. Another example is the Russian disinformation campaigns that artificially hyped up the bedbug scare in Paris in the summer of 2023 [53] or which promoted counter-narratives after the MH17 disaster to paint a false picture of the Kremlin's involvement [54].

3.2 Causes

The element Causes includes two factors that help define the source of the content creation: the creators themselves and their motives. Causes interacts with Context as contextual factors may create opportunities for creators and their motives. In other words, creator's motives may involve the exploitation of a specific opportunity that arises from favourable contextual occurrences, such as an event or social trend (see Fig. 4). For example, the growing trend of the hostile media effect (C1 Context), may represent a sufficient opportunity which then motivates a creator (C2 Causes) to exploit it with disinformation content.

3.2.1 Creators

Different types of creators have been identified in the literature and can be characterised based on three overlapping categories, as described below. For example, the creator can be a fake 'news agency' and therefore be characterised as a non-human, organisation, and non-state actor. For example, a head of state spreading disinformation on social media platforms or online social networks would be characterised as a human, individual, and state actor.

Disinformation content can be the direct creation of a *human* agent who initiates and hands-on publishes the content on the chosen platform. It can also be produced by a *non-human* agent, most often cyborgs or social bots [55]. Non-human agents are most likely automated and allow not only for the creation but also for the mass sharing of content, thus providing a greater capability potential than humans.

The literature identifies both isolated *individuals* and *organisations* as creators. Organisations can be companies as well as organised networks. Budak [56] has identified several organisations, acting as what he calls 'fake news agencies', during the 2016 US elections.

A third and last distinction can be made to characterise actors in relation to whether or not they are a state-actor or related to one. Disinformation campaigns are often related to hybrid warfare and fulfil geopolitical motives [25]. For instance, Russia and China engage in so-called FIMI to manipulate global discourse in a way to "sow divisions in the societies, denigrate democratic processes and institutions and rally support for its imperialist policies" [38].

[12] More about political climate can be found in C1 3.1.1.2.5.

3.2.2 Motives

Motives refer to the underlying reasons for individuals and organisations to engage in the creation of disinformation [23]. This factor is explicitly about the motive to construct disinformation and not about the consumption or sharing of it. The former of the two is addressed as part of 'susceptibility', whereas the latter is analysed as 'dissemination'.[13] Lazer et al. [57] state that all traits of disinformation aim to draw users' attention to read and share it, with two main objectives: financial or ideological gain. Of course, disinformation content or campaigns can be motivated solely by social or interpersonal conflicts, such as defamation, spite, revenge, character attacks, bullying or ridicule [58]. However, the scope of this research is focused on disinformation that leads to harmful or even criminal behaviour that has a society wide impact and requires PA involvement. Such cases of disinformation tend to be on a much smaller scale and, although they may arise from interpersonal conflicts, are nonetheless often intertwined with political or financial motivations [59].

Financial motives can include revenue generation from advertising opportunities in combination with low operating costs and the fact that few means are required to fabricate disinformation [23, 57]. Specifically, Lazer et al. [57] argue that fraudulent strategies exist to create and share disinformation that gets as much traction as possible just to gain such financial revenues. For example, the driving force behind a for-profit troll farm in Veles, Macedonia, which was set up during the 2016 US Presidential election, was financial revenue rather than political influence [60]. There may be instances of purely economic-motivated disinformation that seeks to disrupt economies, take control of markets or undermine competitors. Although there is a growing concern over this [61], there is little evidence to suggest that disinformation has a direct effect on economic performance and in any case, it is usually politically motivated.

Governments and other (aspiring) political actors can create disinformation for *ideological* motives [23]. Politically related fabrications can be carried out with a wide variety of aims including the persuasion of a population towards a partisan political option,to deceive or confuse a target; or to sow mistrust against government authorities, institutions or other communities [25]. A prime example of politically motivated disinformation is the type of FIMI disinformation carried out by Russia and China (see chapter 3.1.2.1 above on state actors). For instance, Russia wants to be seen as a great power, so it serves the Kremlin to spread messages highlighting weaknesses in Western political institutions while at the same time gaining support for its own policies [38]. Religious or cultural differences, factions or conflicts are another important consideration here. Like political conflicts, religiously motivated disinformation campaigns are often initiated with an ideological goal in mind (such as inciting hatred against social groupsor undermining civic institutions) [62].

3.3 Content

Content relates to the constituent parts of any given piece of disinformation (see Fig. 5). For instance, what form does the disinformation take (e.g., text-based or visual)?; how is it being disseminated?; what action or strategy is being propagated?; and which narratives and other storytelling-devices is the disinformation exploiting?

3.3.1 Forms

Disinformation can take on many different forms, such as videos, images, or text-based articles or a combination of them. It is important, therefore, to grasp what these different forms look like, how they differ from each other, and how they can be deployed to effectively communicate disinformation. The rapid expansion and proliferation of artificial intelligence (AI) techniques have ushered in a new era of disinformation which poses "formidable threats to the integrity of information ecosystems worldwide" [63]. As such, the forms of disinformation detailed in this section (text, headlines and visuals) can all be subjected to mass-scale automated production and dissemination using Artificial Intelligence (AI) technologies.

Research shows that there are certain qualities and patterns that can be commonly found in the *text* used in disinformation articles that can demarcate them from real news. For instance, the lexicon used in disinformation tends to be more informal and simpler; the text tends to lack substantial information, containing lots of redundant language which is less analytical and more personal than real news [64]. Furthermore, fake 'news' articles tend to be a lot shorter than real news articles, and feature smaller words, less punctuation, fewer quotes, and have an absence of technical words and logical argumentation [64]. The simple yet compelling features of text-based disinformation messages not only make the information more persuasive, but also more attractive for repeated and widespread dissemination [133, 134]. In their study, Horne and Adali [64] found that fake 'news' articles required a slightly lower education threshold to read them.

[13] See C5 3.1.5 Cycle of Amplification.

However, this is changing, and fake 'news' now more closely mimics the same language used by respectable journalists. This makes it much more difficult to ascertain whether the information is real or not.[14]

Often the *headlines* that accompany disinformation articles are just as critical to their successful dissemination as the content itself. Headlines tend to be bombastic and controversial, laced with emotive language that exaggerates and sensationalises the topic [45, 135]. As Munger et al. [65] suggested, attention-grabbing headlines are for enticing people to read and spread disinformation. Such headlines must appeal to the emotions, feelings, and identity of the audience, while exploiting certain emotions and partisan cues [65]. The aim of these 'clickbait' headlines is to grab readers' attention, exploit their heuristics and push them to disseminate with as little scrutiny of the article as possible [64, 65].

A significant proportion of disinformation is disseminated through *visual* imagery. According to Weikmann [136], visual disinformation differs from text-based disinformation in relation to: *production* (the level of sophistication of the visuals matters more with visual disinformation and varies significantly); *processing* (greater realism and visual stimuli can affect attitudinal and behavioural change more directly); and *effects* (leads to stronger and possibly different effects than text). Visual disinformation can be classed along two dimensions: modal richness (whether still or moving images are used) and manipulative sophistication (the level of creation technique used; Weikmann 136). Multimodal disinformation (i.e., text and visual) is seen as more credible than mere textual disinformation because an image provides a more direct representation of reality and lends "an inherent evidential quality to a story" [63, 66] Fabricating images and videos is not a new form of disinformation, but recent technological advancements have supercharged the speed and breadth of production and dissemination. Early research suggests that generative AI at least contributes to people's cynicism and uncertainty of news sources, which can lower trust in news articles on social media platforms [67].

3.3.2 Channels

Channels are the means through which to communicate disinformation to audiences. Different channels allow for different types of dissemination, effects, and reach. Of course, this list is not exhaustive, and disinformation is often spread using other types of websites and online social networks, such as message boards, chat forums, and blogs.

Social media platforms—such as Facebook, Twitter (now X), TikTok, and Snapchat—are ripe for the dissemination of disinformation [68]. Information can be spread instantaneously through online social networks around the world via these platforms, transcending culture and geography to reach global audiences. Recent advancements in the algorithms that underpin social media platforms mean that the communication speed and reach of online social networks on these platforms is now matched with a highly sophisticated system of personalised targeting [69]. These characteristics contribute significantly to the proliferation of disinformation [70, 71]. Online social messaging services, such as WhatsApp or Telegram, are also being used to effectively disseminate false information. For example, both in India and Brazil, messaging services such as WhatsApp, are now the main avenues by which disinformation is spread [72, 73].

The disinformation that is spread on social media platforms and messaging services often originates from *news websites*. This includes state-controlled channels that have an affiliation with a state actor, such as RT[15] (formally Russia Today), as well as channels that have no official link to state actors but still toe the state line, such as Red [74]. News websites also refer to independent inauthentic news websites posing as real news agencies or websites that peddle disinformation, and conspiracy theories [38]. For example, during the 2016 US Presidential election, more than 100 pro-Trump fake news websites were registered at a for-profit troll farm in Veles, Macedonia [75]. Another category can include websites that are impersonating actual news websites, using their authority to share untruthful information and making it difficult for an internet user to see that they are consuming disinformation [76]. Lastly, authentic news agencies or 'mainstream media' could contribute to the spread of disinformation simply by reporting on it, which creates more attention for that specific campaign and thus can unintentionally contribute to the cycle of amplification [31]. See chapter 3.1.5 below on the C5 Cycle of Amplification.

Online *influencers* are now being recruited and targeted by malign actors to channel their disinformation to new or bigger audiences. These actors mostly operate on social media platforms but can also be employed in other forms of broadcasting media, such as tv-shows, radio broadcasts, or podcasts. Influencers can wield significant influence over their audience—perhaps more than other political or state actors—via their chosen medium, whether it is a blog, video-sharing

[14] Information from a panel during the Community for European Research and Innovation for Security (CERIS) workshop on Disinformation, Fake News and Hate Speech, held on 11 May 2023.

[15] RT is a Russian state-controlled international news channel.

platform, or social media platform and can therefore amplify the spread and impact of disinformation (e.g., [23]). China is broadening its use of social media influencers to counter criticisms of the country's human rights record [38], while in the Philippines, populist former President Rodrigo Duterte recruited influencers to mobilise his supporters—so much so that most disinformation in the country is perceived to come from influencers, bloggers and vloggers [77, 78].

3.3.3 Strategies

Several strategies for disinformation messaging exist that are each closely related to its creators' motives. In this context, strategies are defined as the general means by which one would achieve one's overarching goal with disinformation. In general, most creators want to persuade the receiver of something, or at least cast enough doubt on perceived reality to make persuasion possible. Persuasion can be achieved by using strategically formulated messaging that uses language that undermines truth and plays on emotion; perhaps even connecting with the audience by signalling authority and homophilic characteristics (i.e., shared features between the disinformation creator or message and the receiver) [23]. Of course, these strategies are all interlinked and are often employed in tandem or sequentially as part of the same campaign.

The objective of disinformation is often to *persuade* an audience; either to shift attitudes towards a certain standpoint or to simply shift attitudes away from a conventional truth. Arcos et al. [25] argues that both rational and emotional persuasion can be used. Wiggins [133] considers that the sensationalist and attractive way in which most disinformation is presented fits into the peripheral route of persuasion (emotional), which "implies focusing on those components not central to the argument or message, but paying more attention to how the message is presented" 133, as opposed to the central route (rational). Moreover, since the effect of persuasion usually decreases over time, actors aiming to persuade an audience will want to do this while the public's interest is at its highest, and before other narratives take over [23].

Another key strategy of disinformation is to *undermine* democratic processes and perceptions of truth. In other words, disinformation campaigns often seek to undermine such things as democratic institutions and processes, political actors and the political process; or even undermine perceived reality and truths of a certain audience. These activities are closely related to the subfactor political motives (C2 Causes) and can also include denigrating and misleading the public. The ways in which actors seek to achieve this can include, but are not limited to, conspiracy theories, deepfakes (photo, video, sounds and text), other manipulation of images, micro-targeting, the leak of apparently official documents, and the general alteration of history [25].

In a similar vein, another strategy of disinformation is to *exploit* audience's cognitive vulnerabilities for financial gain. For instance, financial motives for creating disinformation can be connected to a fraudulent strategy to make money from (online) advertisements, baits, views, and clicks [57]. Even though the content of a message might be political, the underlying motive could be maximum financial gain for the creator and disseminator.

A core element of all these strategies is the narrative. Narratives are broader than one single message or campaign, and are akin to a storyline or a theme that helps bolster the effectiveness of strategies designed to persuade, undermine or exploit. For example, a common Russian narrative is to position themselves as a victim of the West [79]. Disinformation seeks to fuel such narratives by adding contextual embellishment to the main storyline. The most common types of narratives are often related to political topics or sensationalist rumour [45]. Narratives can also be based on myths, such as the historical myths propagated by German Nazis in their propaganda narrative to justify violence against certain people [2]. Finally, most narratives play on emotions as they often encourage strong feelings such as anger, despair or excitement which are linked to higher reader/user engagement [80–83]. Narratives often use exaggeration to enhance the emotional and sensational component of the content to attract and engage the news consumer more intensively [84].

3.4 Consequences

Consequences relate to the factors that are the direct or indirect results of exposure to disinformation, especially visible in the behaviour of its recipients. In other words, consequences are the impact of disinformation (see Fig. 6). This can happen either at an individual level (micro effects), which then emerge at the group level (meso effects related to in-group and in-group dynamics), or even escalate to the societal socio-psychological level (macro effects leading to public disorder). There is a hierarchical and escalatory relationship between the effects of exposure to disinformation at the micro, meso and macro levels, which can lead to harmful or potentially criminal offline behaviour that requires involvement from law enforcement authorities. For example, although increased affective polarisation is not directly the concern of police authorities, the offline behavioural consequences of the effect of polarisation, such as violent protest

or hate speech, does fall under the jurisdiction of law enforcement. Furthermore, these consequences can in turn lead to changes to the context in which the disinformation takes place, creating a feedback loop between Consequences and Context, as denoted in Fig. 2.

3.4.1 Micro effects

Micro-level effects relate to cognitive effects experienced at the individual-level as a result of exposure to disinformation, which could tap into shared grievances. In other words, the psychological impact of disinformation on an individuals' cognition. As seen in Chapter 3.1.3, the strategy of a disinformation campaigns is often to undermine populations, casting doubts and sowing division to allow for the potential of persuasion, whether that be shifts in attitude or behaviour. Considering that the scope of this paper is about disinformation that can lead to harmful or even criminal behaviours, the following section is written on that assumption that the initial process of casting doubts and undermining audiences has already taken place, therefore attitudes and behaviour change may then be possible to influence. It is important to caveat this with the fact that it is extremely difficult to ascertain causal links between disinformation and attitudinal and behaviour change; however, they remain important considerations when analysing possible consequences of disinformation.

A fundamental goal of disinformation is *persuasion*. Research shows that disinformation, disseminated via online social networks in particular, is effective in affecting political attitude change in individuals [85]. The ways in which beliefs become established in individuals are numerous (identity, political knowledge etc.), however people often rely on cognitive shortcuts to form beliefs and make sense of the vast information environment [23]. For instance, an oft-cited explanation for the persuasive power of disinformation is confirmation bias: an individual's tendency to react positively to information that matches their prior beliefs [23]. Another explanation of persuasiveness concerns the Elaboration Likelihood Model theory which posits that individuals process information through two paths: (1 a central path of argument evaluation that requires high cognition,and (2 a peripheral path that relies on heuristics and requires less cognition [86, 87]. The extent to which information is persuasive depends on which path the information is evaluated through.

From persuasion comes *conviction*. This refers to "the incorporation of a given fake news narrative into an individual's mental model as a deeply held belief" [23, 23].

Ultimately, one of the main desired effects of disinformation is often *behaviour change*, directly or indirectly. Recent research by Bastick [88] found that even limited exposure to disinformation (less than five minutes) was enough to alter individuals' unconscious behaviour. In a political context, it has been shown that exposure to disinformation—particularly if it was politically charged and intentionally fabricated—had a positive effect on electoral support for populist parties, regardless of prior support [89]. Populist parties are not the only ones who engage in disinformation, although it does tend to stem mainly from parties on the extremities of the political spectrum [46, 90, 91]. Although disinformation alone cannot explain growths in populism, for example, there are signs that disinformation has the potential to influence unconscious behaviour change, which can therefore lead to changes in conscious behaviours such as voting and demonstrating.

3.4.2 Meso effects

Individual-level effects can aggregate to generate group-level effects, otherwise known as 'meso effects'.

As disinformation spreads and pollutes an information environment, effects at the individual level will eventually impact groups and communities, potentially leading them to engage in rumour, conspiracy or even seek out extreme viewpoints. Because of the divisive nature of disinformation [92], people will tend to connect even more with their in-group ('their own',people with similar beliefs and experiences) and widen the gap with the out-group ('the other'). This is called *affective polarisation* [93]. Susceptible members of groups may reinforce and legitimise each other's beliefs. Social dynamics can influence and strengthen in-group/out-group perception, which lies at the core of polarisation processes, can lead to organised extremism, and is common in disinformation campaigns. The group becomes a factor of influence on the cognitions and behaviours of its members, leading to a potential threat to individual autonomy and neutralising individual critical thinking [88].

Furthermore, the collective polarisation and radicalisation process provides psychological safety and further motivational triggers for individuals to mobilise and engage in *collective behaviours*, including civil unrest and potentially violence. It is well established that communication flows and information controls are "indispensable ingredients of violent conflict" [94], and there is growing evidence that false or conspiratorial information can be a catalyst for group violence. For example, research has shown that the spread of disinformation on WhatsApp in India could be a contributing factor

to an increase in lynchings and violence as users were being predisposed to hate a certain group and encouraged to engage in violence either motivated by prejudice or by rumour [95]. Furthermore, an analysis of data from more than 150 countries worldwide showed that the dissemination of disinformation is among the drivers of domestic terrorism [96]. That said, as mentioned previously, the causal links between disinformation and offline unrest or violence are difficult to grasp and there are countless mitigating factors that must be considered when assessing behaviour influencing and change. Collective behaviour can also mean inaction. Audiences exposed to a lot of disinformation may experience the phenomenon of "information learned helplessness" whereby people are so engulfed in false information that they simply give up trying to ascertain the truth [97].

3.4.3 Macro effects

Individual-level effects can aggregate to generate group-level and then societal-level effects. This is referred to as 'macro effects'.

As individual beliefs become entrenched and news consumption becomes more partisan, the prospect of *polarisation* in society becomes more likely. The mindset of 'us vs. them', often initiated at the individual-level through persuasion and conviction, is easily extrapolated to the societal-level once people form groups with other like-minded individuals based on a shared distinct belief [23]. Polarisation occurs when these groups position themselves in opposition to other belief groups, which creates division and cleavages. Polarisation most often relates to political ideology or social identity, but it can be predicated on numerous factors.

The next step from polarisation is *aversion*, defined as "the complete repudiation of opposing views and those that hold them" [23]. Individuals who are strongly convicted in their beliefs and are members of polarised groups can quickly transition from benignly avoiding oppositional viewpoints to actively attacking those who hold opposing views, or supply information, contrary to their own [23]. Aversion therefore goes beyond mere polarisation (forming of partisan groups), and is the process by which these groups proactively antagonise and resist their opponents.

Polarisation and aversion can lead to a climate of *distrust* in society, in which the divergence between oppositional groups causes an erosion of trust. At a societal level, distrust can negatively impact a multitude of factors, such as political institutions, the rule of law, the media, or society itself [98]. An example of this that relates to disinformation is the phenomenon of hostile media effects, discussed above, which is the culmination of polarisation and aversion to create widespread suspicion and distrust of mainstream news agencies [23].

These effects, whether at the micro or macro level, are the potential consequences of disinformation. As mentioned above, these consequences can in turn influence or change the nature of the context in which the disinformation takes place, thereby creating somewhat of a feedback loop between the Consequences and Context factors in the C5 Model. For instance, if a consequence of disinformation is that it influencers individual attitudes, collective action or levels of polarisation in society then this impacts the nature of the context, whether that is in regard to the setting, societal trends or even specific events that have been instigated as a result of the disinformation.

3.5 Cycle of amplification

The Cycle of Amplification refers to the relationship between dissemination, or propagation, and persuasion, which is usually the overarching goal behind disinformation (see Fig. 7) [23]. As such, this element demonstrates the interaction between the first four elements: within a certain context (C1), disinformation messages are created (C2), and based on its content (C3), the susceptibility of its receivers, (C5) and the dissemination (C5), effects are created (C4), which in turn can become causes in itself for new disinformation messages or campaigns. To counter the effects of the cycle of amplification, practitioners can employ interventions (C5). This refers to any measure that is designed to mitigate or prevent the spread and impact of disinformation on society. Here, we are not concerned with (psychological or sociological) effects of disinformation on society but instead with potential interventions that are implemented to counter the dissemination of disinformation or (in)direct threats posed by disinformation.

3.5.1 Susceptibility

It is important to note that susceptibility is not a set variable and can constantly evolve. This also means that it is difficult to establish exact profiles of individuals who would never be susceptible to disinformation, because it can change over time. Moreover, in general people get information through testimony and tend to believe information by testimony,

meaning that they only tend to doubt at a later stage [99]. Susceptibility comes from a combination of traits that are more permanent, like personality traits [100], but also from variations of mood, emotions, and state of mind. All individuals are, to some extent, led and affected by emotions which influence perception and reaction to the processed information [80]. Also, context plays a large role in affecting one's psychology. For example, some contextual amplifiers may have made more people vulnerable to disinformation during the COVID-19 lockdowns such as the constant uncertainty or the overload of information [79]. In sum, not every individual will be triggered by the same piece of disinformation and if they are, they will not be triggered in the same way.[16]

One's *beliefs* are ideological attachments, in other words, core thoughts that influence the rational and emotional information process of an individual [68]. Or, in other words: that which an individual believes to be true. In the case of disinformation, Galeotti [99] argues that beliefs influence the susceptibility of the individual to accept or reject a piece of new information, as they will tend to accept ideas that are aligned with existing beliefs. This can mean that disinformation about something that previously was not particularly important to a person (e.g., how vaccines were made in the context of COVID-19), suddenly generates traction online. Research indicates that a low trust in established governmental and media institutions turns consumers to alternative media sources and makes them more susceptible to disinformation [101]. Galeotti [99] also highlights that the more conservative the beliefs, the more the individual may believe in disinformation. Lastly, reinforced extreme beliefs tend to influence polarisation and binary thinking (false vs true, wrong vs right).

Another factor of influence for susceptibility is found in *heuristics and biases* [86, 102]. The human brain constantly makes a high number of quick decisions, most of them unconscious. To reduce the cognitive effort, the brain uses heuristics, or shortcuts, to make quick, deeply rooted, and almost automated decisions [103]. When people are confronted with a piece of information, they are generally biased to believe in its validity, which is referred to as the truth bias [104] and is similar to the testimony belief referenced to earlier [99]. Heuristics and biases are mostly influenced by beliefs, which they also reinforce. An example of this are the earlier discussed confirmation bias and the Elaboration Likelihood Model theory.[17] The strength of the confirmation bias is also closely related to contextual factors and individual characteristics [105].

Even a single *exposure* to disinformation can increase the perception of its accuracy. This is also referred to as the 'illusory truth effect' and can pertain even though it is flagged as disinformation at a later stage [106]. Exposure to disinformation in itself is not a determinant of susceptibility, but when it occurs it strengthens the effects of one's beliefs and their biases. By increased exposure, beliefs, polarisation, or even radicalisation can be reinforced. Similarly, one's relationship to its social network, and most often the need to belong can increase the susceptibility of gossip consumption and engagement and therefore exposure and the risk of consuming disinformation [107].

3.5.2 Dissemination

Dissemination or propagation refers to the distribution of disinformation [23]. This is enabled by online social networks since they make it easier for users to share disinformation. High virality metrics, including numbers of likes or shares, and group norms further amplify the dissemination [23]. This factor is closely related to the context as well. For example, growing distrust in media—as a contextual feature of the political climate in a given environment—is seen as an important influencer of the consumption and dissemination of disinformation [108].

Several authors confirm that people's *social status* is strongly related to their propensity to share news content [109–111]. Specifically, the user can feel that their social reputation is reinforced by showing their social network (which can exist of several groups) that they are up to date on the latest news. New and impressive information, which is mostly consistent with the main characteristics of disinformation, can lead to greater group acceptance [99].

Certain *individual characteristics* influence to what extent someone is vulnerable to not only consuming disinformation but sharing it as well. Guess et al. [100] examined which individual-level characteristics were relevant with regards to sharing disinformation on Facebook during the 2016 US presidential campaign. After controlling for other demographic characteristics, the authors found that political affiliation and age significantly influenced who shared disinformation. Conservatives and older people were found to share more [100]. Various other studies replicated the finding that

[16] In addition, individual characteristics are discussed under 3.1.5.2 Dissemination and contextual factors are analysed under 3.1.1 Context.

[17] See 3.1.3.1 Micro effects.

ideologically and politically speaking, right-wing oriented people are more likely to consume and spread disinformation [68, 101, 112–114]. Other individual characteristics that influence disinformation consumption are being widely challenged in academia, including age, gender, and internet usage [68].

Echo chambers are created when members of a community share disinformation with each other in specific (online) groups, which leads to other members reading and sharing that same message as well. This phenomenon contributes significantly to the spread of disinformation online [70, 115]. Rini [116] disinformation reaches a user through the testimony of another person, who shared it after accepting it as being true. The testimony is transmitted, especially on social media, often in a biased way, since it comes from someone who has just shared an ideology or expressed a party attachment. This information will be accepted and shared by a recipient who agrees with these social values [116]. Moreover, the spreaders in this echo chamber might unintentionally contribute to the spread of disinformation, since they might just be eager to participate in their (online) community [28]. Disinformation has found increased exposure on Facebook and Twitter (now X) due to the segregation of groups and highly advanced recommendation algorithms that are now key features of online social networks—although users themselves still play a crucial role in the creation of so-called 'filter bubbles' or 'echo chambers' [70].

3.5.3 Interventions

Interventions refer to efforts to mitigate the amount, spread, or impact of disinformation. This includes the disruption of the amplification cycle [23].

Disinformation *detection* is the task of assessing the truthfulness of a certain piece of news. This can be done manually by fact-checking experts but can also be done through automated analyses using data-mining and/or machine learning, although these techniques are in their infancy [28]. Fact-checking forms part of the detection process by verifying the information contained in an alleged piece of disinformation which, again, can be done manually or through automated means [117]. The fact-check is then usually advertised by means of flagging, which simply alerts people to information falsehoods and stimulates critical thinking, as well as lowering the likelihood of users' "intentions to share the article" [25]. A specific type of flagging is source rating**,** which involves the provision of additional information online related to the sources of information contained in supposed disinformation. Source ratings influence the believability of articles, which makes readers more sceptical of news stories on social media platforms and online news sites regardless of the source's credibility, while low source ratings lowered the believability of the article and reduced reader engagement [118].

Going beyond mere detection, a valuable intervention in response to disinformation is *debunking*. This is the process whereby disinformation is detected, flagged, and the false information contained in disinformation is corrected or rebutted [25]. Research shows that debunking disinformation by providing a rebuttal and introducing corrective information is far more effective than merely labelling the article as disinformation [119]. However, in general, the effects of debunking can vary considerably and depend on many factors, including the level of detail included in the debunking information, the level of reasoning behind people's belief in the information, and the time between the disinformation publication and the rebuttal [119, 120]. AI can also be used in debunking efforts [121]. A standout example of debunking in action is the 'Debunk EU' initiative which incorporates AI tools, volunteers and journalists to research and debunk disinformation in the Baltic States and beyond [122].

A more proactive countermeasure to disinformation is *pre-bunking*, sometimes referred to as inoculation. The goal is to build societal resilience against the dangers of disinformation, in order to pre-empt its effects. If people are educated about the threat of disinformation—for example, through increased media literacy [123]—and forewarned that they may be targeted, they will become immunised against disinformation [124, 125]. This also depends on various other key indicators of societal resilience-building, such as levels of populism, polarisation, media trust, time spent on social media platforms and the strength of the public broadcasting service [126]. It proves that interventions can also influence the context eventually. Furthermore, generating social norms around disinformation reporting and detection can lead to higher rates of such reporting by individuals [127].

Another intervention for disinformation is greater *regulation* in online environments. As the threat of disinformation has grown in recent years, so too have calls for more regulation. However, this is a controversial intervention that is fraught with challenges and has sparked considerable debate as policymakers rankle with developing regulatory frameworks at national or regional levels [128, 129]. The EU's Strengthened Code of Practice on Disinformation 2022 compels signatories to: demonetise the dissemination of disinformation; guarantee transparency of political advertising; improve cooperation with fact-checkers; and facilitate greater access to data for researchers [130]. This comes as part of greater regulation of disinformation by the EU via the Digital Markets Act 2022, the Digital Services Act 2022, and the forthcoming

Artificial Intelligence Act. Such regulation is important as it dictates the legal and normative boundaries within which online social networks, for example, can operate and the extent to which they are required to curate their content—the greater regulation, the more online social networks will be required to curate their content to remove disinformation.

4 Discussion

Recent examples of social unrest around the world, like the storming of Capitol Hill in the US in 2022, have illustrated how disinformation can fuel division and incite violence. It is therefore vital for authorities tasked with dealing with the real-world implications of disinformation to understand the process of how it is created, disseminated, and takes effect in society. The present article sought to outline the disinformation process, from creation to effects, to illustrate the process by which the spread of disinformation can escalate to harmful or even criminal behaviours. By synthesising academic literature on disinformation from various fields or research, the aim was to provide practitioners with an accessible and practical guidance on the dynamics of disinformation that strikes a balance between theory and practice. The C5 Model outlines the interactions and interrelationships between the five main factors and its various subfactors. This model can be used in a versatile way by different users. Moreover, the C5 Model contributes to a gap in the academic literature by combining a diverse set of insights on disinformation into one cohesive framework. Another novel feature of this model is that it emphasises that, due to the nature of the relationships between these factors, the same piece of content can have different consequences, depending on the context. The following sections elaborate on both the practical and academic contribution of this article.

4.1 Practical contribution

Based on interviews with five European PAs, a shared need was identified to improve countermeasures against disinformation that can lead to criminal behaviours.[18] According to the PAs, the first step to improving their counter-disinformation capabilities is to develop an in-depth understanding of the phenomenon. None of the PAs that the authors consulted have a dedicated team at their disposal to understand, prevent and counter the harmful and potentially criminal consequences of disinformation. Some PA departments might have one dedicated officer, but it is the norm that these duties are conducted alongside other unrelated police duties. By developing a practical framework that breaks down the complex concept of disinformation into constituent elements, this research will help facilitate a deeper understanding of the phenomenon. It encourages systematic thinking prompting not only the interviewed PAs, but practitioners in general, to consider all stages of the disinformation lifecycle rather than focusing on isolated aspects, and understand how these aspects fit into the broader picture of a campaign.

Practically, the C5 Interaction Model has been developed specifically for the use of practitioners throughout the different phases of the counter disinformation process. First, it can be used *preventively*. The model can be employed as course material to teach fellow colleagues about the workings of the disinformation lifecycle. This way, more officers will be competent in spotting and acting on potentially harmful disinformation messages. Moreover, an understanding might be created about recurring events (nationally or internationally) that inspire disinformation (e.g. elections or events revolving around polarising topics in that specific society). This can enable practitioners to anticipate disinformation and mitigate its effects on society. Second, the C5 Interaction Model can be used on an *ongoing basis* as a decision-support system. If a piece of disinformation is spotted, the model might help the officer to think of important elements that are related to that specific message and therefore aid in deciding whether it would be worthwhile to further investigate. Third, the C5 Interaction Model can be used *retrospectively* to analyse the disinformation process after events have already occurred. This can help with post hoc analysis of specific situations to improve future readiness and responsiveness.

4.1.1 Use case summary

To exemplify the efficacy of the C5 Interaction Model, it was applied to a use case from Catalonian authorities. The application illustrates how the model can be used to identify and understand the various processes of a specific real-world case of disinformation that led to harmful or criminal behaviour (see Fig. 8).

[18] These interviews were carried out within the context of the VIGILANT project.

The Catalonian use case concerns a public Telegram channel which spreads hate narratives linked to far-right ideologies among approximately 1200 followers. The administrators of the Telegram group used disinformation as a tool with the aim of polarising the receivers and generating a climate of hatred towards various groups.

The figure above summarises how the model was applied to understand the Catalonian case.[19] The breakdown of the elements in the illustration shows that the Catalonian disinformation campaign took place in a context of political tensions. The creators used the social power of the community and its social network, contained in the Telegram channel. The political context was probably the catalyst for like-minded people to join the Telegram channel in the first place, thus creating a favourable environment for the disinformation to be received, and most likely spread even outside of the group, thus recruiting more members. The content, which aligned with the members' beliefs, was most likely a trigger to further radicalisation of the group.

4.2 Academic contribution

The field of disinformation research is expanding rapidly, with insights emerging from disciplines such as communication studies, psychology, political science, sociology, and computer science. However, this exponential growth also carries the risk of fragmentation, as researchers often focus on isolated aspects of the problem. The C5 Interaction Model represents one of the first concerted efforts to bring these diverse insights together into one comprehensive, integrative framework. While previous frameworks in academic literature have explained aspects of the disinformation spread (e.g. [19]), sustainment of effects (e.g. [21]), and amplification of messaging (e.g. [23]), the present article has sought to connect all these factors into one comprehensive framework, while simultaneously demonstrating the interrelationships and fluid interconnectedness of the entire process of disinformation creation, spread, and effects. As far as known, no similar comprehensive framework has been developed yet.

The C5 Interaction Model shows that there is not one fixed route that a piece of disinformation takes, making it a highly complex phenomenon to research. Moreover, the route is predominantly influenced by the social context in which it takes place, a factor somewhat neglected in prior literature [22]. This provides a more nuanced appreciation of the underlying mechanisms of disinformation which goes beyond conventional models of communication that denote a linear relationship between sender and receiver (e.g., [131]). The C5 Model puts together disparate pieces of the enormous puzzle of disinformation research. An example of this is the incorporation of the Cycle of Amplification, a recent addition to the body of disinformation literature [23], as a main factor in the model explaining how the impact of disinformation is mitigated by the amplification cycle. Furthermore, as highlighted in the literature review, there is a lack of work that explores the links between disinformation and potentially harmful or criminal behaviour, such as social unrest, public disorder and hate speech, in a comprehensive and practical manner.

4.3 Limitations and future research

It should be noted that the C5 Interaction Model is not exhaustive. The online environment where disinformation spreads, which also interacts with the physical world, is constantly changing. Therefore, the factors outlined in the C5 Interaction Model should be thought of as thematic labels under which numerous specific examples can be categorised, rather than rigid, definitive concepts. Furthermore, the categorisations and descriptions of the (sub-)factors are based on the authors' current understanding of the disinformation phenomenon, which is ever-changing and growing. As such, continuous research is needed to ensure the conceptual model remains relevant and accurate.

Similarly, further research is required to validate the findings of this article. Steps have already been taken by the authors to test the model, by way of use cases to support its conceptual validity and through workshops with practitioners to ensure its practical relevance. However, more rigorous validation of the model, both in terms of theoretical validity and practical implementation, is required to solidify the findings of this article. A suggestion would be to build on the application of the C5 Interaction Model to the Catalonian use case. More of such examples of practical applications are needed to analyse in which ways the model can fit into current practises and processes of practitioners. However, this would just validate the model's retrospective appliance, whereas—as discussed in the practical contribution section—the model could also be used preventively and on an on-going basis. To test the preventive value of the model, practitioners are encouraged to start using it for educational purposes. These sessions could be observed and analysed by researchers

[19] A full summary of the Catalonian case study can be made available by the authors on request.

and the respective authorities. The same goes for using the model as a decision-support system. Questions that would focus on the practicality, understandability and applicability of the model could guide such evaluations. Also, a vital dilemma remains: when does disinformation become dangerous enough for practitioners to act upon? Future research could analyse this question, whereby it is important to take into account individual societies legal and ethical frameworks.

Another potential critique of this paper is the use of both theoretical and empirical studies in the development of the C5 Interaction Model. As stated in the methodology section, the foundations of the models—in other words, the five main 'Cs' and their ordering in the process of disinformation—were developed using primarily meta-analyses and existing literature reviews. This allowed for a broader analysis of the general disinformation processes. When developing the subfactors, empirical studies were relied upon more. As aforementioned, despite this distinction there remains some overlap between these two types of literature and the weight given to each type of source is not clear. However, this was done in order to simultaneously provide the necessary level of detail required to understand the mechanisms of disinformation, while also ensuring it is understood within the wider context of disinformation and its impact on society. Given the intended audience of this paper and the emphasis on the real-world application of its findings, the need for clarity and efficacy in presenting the most valuable information to practitioners superseded any concerns, albeit valid, over the blurry distinction between theoretical and empirical evidence.

Also, the reliance on evidence from mainly two-party political systems might be seen as a limitation, because this paper is focused on European practitioners who operate in multi-party political systems. However, much of the literature on the phenomena of disinformation is centred around the US or the UK—two distinctively unique political systems that are not wholly generalisable to a European context. This is likely due to the recent high-profile bipartisan political and constitutional events that were cornerstones in manifesting global attention and research on disinformation and its effects, namely the political rise of Donald Trump and the Brexit referendum. Any bias towards US and UK sources, research and case studies is somewhat inherent to disinformation research because these events garnered such international attention and provided such fertile ground for analysis. Furthermore, comparative research on disinformation in different political systems has tended to focus on democratic versus non-democratic regimes rather than differences within democracies, such as two-party or multi-party political systems. As such, this is an obvious issue for which further study is required to understand better how disinformation manifests, spreads and takes hold in varying types of political systems. Nevertheless, this paper endeavoured to provide a more European outlook, utilising a case study from Catalonia at the main case to exemplify the use of the C5 Interaction Model, while also using research from other parts of the EU, such as Portugal, Germany, the Netherlands and Italy.

5 Conclusion

The main contribution of this article is the C5 Interaction Model, which represents one of the first concerted efforts to bring multidisciplinary insights on disinformation together into one comprehensive integrative framework. The model, consisting of five key interrelated elements of disinformation (Context, Causes, Content, Consequences, and Cycle of Amplification), can increase understanding on the dynamics and consequences of the disinformation lifecycle by providing a high-level overview of this complex process. Based on a synthesis of academic theory and literature, the C5 Interaction Model also provides accessible and practicable guidance to practitioners tasked with handling the real-world consequences of disinformation.

Author contributions K.K., NB.F, B.C and S.vd.M. (all authors) wrote the main manuscript text. K.K., NB.F. and B.C. designed the figures. All authors reviewed the manuscript.

Funding The initial research leading to these results received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No: 101073921.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate The Ethics work package from the VIGILANT project, led by the University of Freiburg, has approved the interviews with PAs. Consent to Participate and Consent to Publish were obtained from all participants in the study.

Competing interests The authors declare no competing interests.

References

1. Jamalzadeh S, Mettenbrink L, Barker K, González AD, Radhakrishnan S, Johansson J, Bessarabova E. Weaponized disinformation spread and its impact on multi-commodity critical infrastructure networks. Reliab Eng Syst Saf. 2024;243:1–11.
2. Rathje J. Driven by conspiracies: the justification of violence among "Reichsbürger" and other conspiracy-ideological sovereignists in contemporary dutchy. Perspect Terror. 2022;16(6):49–61.
3. European Commission. A multi-dimensional approach to disinformation: article of the independent High Level Group on fake news and online disinformation. Directorate-General for Communication Networks, Content and Technology. 2018. https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation. Accessed 7 June 2023.
4. Stahl BC. On the difference or equality of information, misinformation, and disinformation: a critical research perspective. Inform Sci. 2006;9:83–96.
5. DiMaggio AR. Conspiracy theories and the manufacture of dissent: QAnon, the 'Big Lie', COVID-19, and the rise of rightwing propaganda. Crit Sociol. 2022;48(6):1025–48.
6. Woodruff Swan B, Lippman D. New Capitol Police document shows how unprepared they were for Jan. 6 riots. Politico. 2021. https://www.politico.com/news/2021/10/29/capitol-police-documents-unprepared-jan-6-riots-517478. Accessed 5 June 2023.
7. Hollywood JS, Harrison B, Matthews M, Donohue RH. Police officers: this article will make you better at combatting misinformation. RAND Commentary. 2020. https://www.rand.org/pubs/commentary/2020/08/how-to-combat-covid-19-misinformation.html. Accessed 28 Oct 2024.
8. Zamparutti T, Jones M, Tugran T, Vona L, Navas L, Sidło K, Chmiel O. Developing a handbook on good practice in countering disinformation at local and regional level. European Committee of the Regions; Commission for Citizenship, Governance, Institutional and External Affairs. 2022. https://doi.org/10.2863/066582.
9. Arcos R, Chiru I, Ivan C, editors. Routledge handbook of disinformation and national security. New York: Routledge; 2024.
10. Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: toward a unified taxonomical framework. New Media Soc. 2021;23(5):1301–26.
11. Alam F, Cresci S, Chakraborty T, Silvestri F, Dimitrov D, Martino GDS, Nakov P. A survey on multimodal disinformation detection. arXiv [Preprint]. arXiv:2103.12541. 2021.
12. Singhal S, Shah RR, Chakraborty T, Kumaraguru P, Satoh SI. Spotfake: a multi-modal framework for fake news detection. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE; 2019. pp. 39–47.
13. Zhou X, Jain A, Phoha VV, Zafarani R. Fake news early detection: a theory-driven model. Digit Threat Res Pract. 2020;1(2):1–25.
14. Fu, D., Ban, Y., Tong, H., Maciejewski, R., & He, J. (2022). DISCO: Comprehensive and explainable disinformation detection. In *Proceedings of the 31st ACM international conference on information and knowledge management (CIKM 2022)* (pp. 4848–4852). https://doi.org/10.1145/3511808.3557202.
15. Santos FCC. Artificial intelligence in automated detection of disinformation: a thematic analysis. Journal Media. 2023;4(2):679–87.
16. DISARM Foundation. DISARM Framework. 2024. https://www.disarm.foundation/framework. Accessed 28 Oct 2024.
17. Terp SJ, Breuer P. Disarm: a framework for analysis of disinformation campaigns. In: 2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). IEEE; 2022. pp. 1–8.
18. Kozyreva A, Lorenz-Spreen P, Herzog SM, Ecker UK, Lewandowsky S, Hertwig R, Wineburg S. Toolbox of individual-level interventions against online misinformation. Nat Hum Behav. 2024. https://doi.org/10.1038/s41562-024-01881-0.
19. Pennycook G, Rand DG. The psychology of fake news. Trends Cogn Sci. 2021;25(5):388–402.
20. Rabb N, Cowen L, de Ruiter JP, Scheutz M. Cognitive cascades: how to model (and potentially counter) the spread of fake news. PLoS ONE. 2022;17(1): e0261811.
21. Froehlich TJ. The role of pseudo-cognitive authorities and self-deception in the dissemination of fake news. Open Inform Sci. 2019;3(1):115–36.
22. Arayankalam J, Krishnan S. The spread and impact of fake news on social media: a systematic literature review and future research agenda. E-Serv J. 2022;14(1):32–95.
23. George J, Gerhart N, Torres R. Uncovering the truth about fake news: a research model grounded in multi-disciplinary literature. J Manag Inf Syst. 2021;38(4):1067–94.

24. Hameleers M. Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. Commun Theory. 2023;33(1):1–10.
25. Arcos R, Gertrudix M, Arribas C, Cardarilli M. Responses to digital disinformation as part of hybrid threats: a systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking. Open Res Europe. 2022;2(8):1–19.
26. Wolfswinkel JF, Furtmueller E, Wilderom CPE. Using grounded theory as a method for rigorously reviewing literature. Eur J Inf Syst. 2013;22(1):45–55.
27. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exercise Health. 2019;11(4):589–97.
28. Zhang X, Ghorbani AA. An overview of online fake news: characterization, detection, and discussion. Inf Process Manag. 2020;57(2):1–26.
29. Fathaigh RÓ, Helberger N, Appelman N. The perils of legally defining disinformation. Internet Policy Rev. 2021;10(4):1–25.
30. Humprecht E. Where 'fake news' flourishes: a comparison across four Western democracies. Inf Commun Soc. 2019;22(12):1973–88.
31. Tsfati Y, Boomgaarden HG, Strömbäck J, Vliegenthart R, Damstra A, Lindgren E. Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. Ann Int Commun Assoc. 2020;44(2):157–73.
32. Carmi E, Yates SJ, Lockley E, Pawluczuk A. Data citizenship: rethinking data literacy in the age of disinformation, misinformation, and malinformation. Internet Policy Rev. 2020;9(2):1–22.
33. Sánchez del Vas R, Tuñón NJ. Disinformation on the COVID-19 pandemic and the Russia-Ukraine War: two sides of the same coin? Humanit Soc Sci Commun. 2024;11(851):1–14.
34. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, Briand S. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res. 2020;22(6): e19659.
35. Heslop DA. Political system. Encyclopedia Britannica. 2025. https://www.britannica.com/topic/political-system. Accessed 9 Apr 2025.
36. Hunter LY. Regime characteristics and online government disinformation. J Inform Technol Politics. 2025;1–20.
37. Meel P, Vishwakarma DK. Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl. 2020;153: 112986.
38. European External Action Service (EEAS). Article on Foreign Information Manipulation and Interferences threats: TOWARDS a framework for networked defence. Strategic Communications, Task Forces and Information Analysis (STRAT.2). 2023. https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en. Accessed 28 June 2023.
39. Cardoso J, Narciso I, Moreno G, Palma N. Online disinformation during Portugal's 2019 elections. ISCTE-IUL, MediaLab CIES-IUL. 2019. https://democracy-reporting.org/en/office/EU/publications/report-disinformation-during-portugals-2019-elections. Accessed 28 June 2023.
40. Cardoso G, Moreno J, Narciso I, Palma N. Social media disinformation in the pre-electoral period in Portugal. CIES e-Working Paper (No. 230/2020). 2020. https://repositorio.iscte-iul.pt/handle/10071/20667. Accessed 28 June 2023.
41. Weikmann T, Lecheler S. Visual disinformation in a digital age: A literature synthesis and research agenda. New Media & Society. 2022. https://doi.org/10.1177/14614448221141648.
42. Bokša M. Russian Information Warfare in Central and Eastern Europe: strategies, impact, countermeasures. The German Marshall Fund of the United States (GMF); 2019.
43. Hameleers M, de Vreese C. Perceived mis-and disinformation in a post-factual information setting: a conceptualisation and evidence from ten European countries. In: Tumber H, Waisbord S, editors. The Routledge companion to media disinformation and populism. Milton Park: Routledge; 2021. p. 366–75.
44. Vallone RP, Ross L, Lepper MR. The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre. J Pers Soc Psychol. 1985;49(3):577–85.
45. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359(6380):1146–51.
46. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. J Econ Perspect. 2017;31(2):211–36.
47. Howard PN, Kollanyi B. Bots, #StrongerIn, and #Brexit: computational propaganda during the UK-EU referendum. arXiv [Preprint]. 2016. arXiv:1606.06356.
48. European Policy Centre. Disinformation ahead of the EU Parliamentary Elections: a snapshot from Bulgaria, Germany, and Italy. 2023. https://www.epc.eu/content/PDF/2023/Disinformation_DP_-_Eiw_and_EMD.pdf. Accessed 9 Apr 2025.
49. La Cour C. Theorising digital disinformation in international relations. Int Politics. 2020;57(4):704–23.
50. Bojovic J. The Brussels conspiracy: narratives of EU-related conspiracy theories in pro-Kremlin media. In Europe: Continent of conspiracies. Routledge; 2021. pp. 214–30.
51. Yablokov I. Conspiracy theories as a Russian public diplomacy tool: the case of Russia Today (RT). Politics. 2015;35(3–4):301–15.
52. Yablokov I. Russian disinformation finds fertile ground in the West. Nat Hum Behav. 2022;6(6):766–7.
53. Reuters. Russian 'disinformation' hyped Paris bedbug scare, French minister says. Reuters. 2024. https://www.reuters.com/world/europe/russian-disinformation-hyped-paris-bedbug-scare-french-minister-says-2024-03-01/. Accessed 30 Oct 2024.
54. Rietjens S. Unraveling disinformation: the case of Malaysia Airlines flight MH17. Int J Intell Secur Public Aff. 2019;21(3):195–218.
55. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM. 2016;59(7):96–104.
56. Budak, C. (2019). What happened? The spread of fake news publisher content during the 2016 U.S. presidential election. In *Proceedings of the 2019 World Wide Web Conference* (pp. 139–150). https://doi.org/10.1145/3308558.3313721.
57. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Rothschild D. The science of fake news: addressing fake news requires a multidisciplinary effort. Science. 2018;359(6380):1094–6.
58. Maftei A, Holman AC, Merlici IA. Using fake news as means of cyber-bullying: the link with compulsive internet use and online moral disengagement. Comput Hum Behav. 2022;127: 107032.
59. Rezayi S, Balakrishnan V, Arabnia S, Arabnia HR. Fake news and cyberbullying in the modern era. In: International Conference on Computational Science and Computational Intelligence, 2018 (CSCI). IEEE. 2018. pp. 7–12. https://doi.org/10.1109/CSCI46756.2018.00010.
60. Hughes HC, Waismel-Manor I. The Macedonian fake news industry and the 2016 US election. PS Political Sci Politics. 2021;54(1):19–23.
61. Petratos PN. Misinformation, disinformation, and fake news: cyber risks to business. Bus Horiz. 2021;64(6):763–74.
62. Apuke OD, Omar B. Fake news proliferation in Nigeria: consequences, motivations, and prevention through awareness strategies. Humanit Soc Sci Rev. 2020;8(2):318–27.

63. Shoaib MR, Wang Z, Ahvanooey MT, Zhao J. Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. In: 2023 International Conference on Computer and Applications (ICCA). IEEE; 2023. pp. 1–7.
64. Horne BD, Adali S. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Paper presented at the Eleventh International AAAI Conference on Web and Social Media, Montreal, QC, Canada, May 15–18. 2017.
65. Munger K, Luca M, Nagler J, Tucker J. The (null) effects of clickbait headlines on polarization, trust, and learning. Public Opin Q. 2020;84(1):49–73.
66. Hameleers M, Powell TE, Van Der Meer TGLA, Bos L. A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. Polit Commun. 2020;37(2):281–301.
67. Vaccari C, Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Soc Med + Soc. 2020;6(1):2056305120903408.
68. Gupta M, Dennehy D, Parra CM, Mäntymäki M, Dwivedi YK. Fake news believability: the effects of political beliefs and espoused cultural values. Inform Manag. 2023;60(2):1–12.
69. Zarouali B, Dobber T, De Pauw G, de Vreese C. Using a personality-profiling algorithm to investigate political microtargeting: assessing the persuasion effects of personality-tailored ads on social media. Commun Res. 2022;49(8):1066–91.
70. Zimmer F, Scheibe K, Stock M, Stock WG. Fake news in social media: bad algorithms or biased users? J Inform Sci Theor Pract. 2019;7(1):40–53.
71. Bernal P. Fakebook: why Facebook makes the fake news problem inevitable. North Irel Leg Q. 2018;69(4):513–30.
72. Khurana, P., & Kumar, D. (2018). SIR model for fake news spreading through WhatsApp. Paper presented at the 3rd international conference on Internet of Things and Connected Technologies (ICIoTCT), Jaipur, India, March 26–27, 2018.
73. Gragnani, J. (2018). Um Brasil dividido e movido a notícias falsas: Uma semana dentro de 272 grupos políticos no WhatsApp. *BBC News*. https://www.bbc.com/portuguese/brasil-45666742.
74. US Department of State. Alerting the world to RT's global covert activities. Office of the Spokesperson of the US Department of State, Fact Sheet. 2024. https://www.state.gov/alerting-the-world-to-rts-global-covert-activities/. Accessed 30 Oct 2024.
75. Posetti J, Matthews A. A short guide to the history of 'fake news' and disinformation. Int Center Journal. 2018;49(3):577.
76. Van der Linden S, Roozenbeek J. Psychological innoculation against fake news. In: Greifender R, Jaffé ME, Newman EJ, Schwarz N, editors. The psychology of fake news: accepting, sharing, and correcting misinformation. Milton Park: Routledge; 2021. p. 147–70.
77. CNN. Survey: most Filipinos see fake news as a problem. CNN Philippines. 2022. http://www.cnnphilippines.com/news/2022/10/11/pulse-asia-survey-fake-news.html. Accessed 13 July 2023.
78. Ong JC, Cabañes JVA. Architects of networked disinformation: behind the scenes of troll accounts and fake news production in the Philippines. University of Massachusetts Amherst, Communication Department Faculty Publication Series. 2018. p. 74. https://scholarworks.umass.edu/communication_faculty_pubs/74/. Accessed 23 July 2023.
79. Hoyle A, Powell T, Cadet B, van de Kuijt J. Web of lies: mapping the narratives, effects, and amplifiers of Russian COVID-19 disinformation. In: Gill R, Goolsby R, editors. COVID-19 disinformation: a multi-national, whole of society perspective. Berlin: Springer International Publishing; 2022. p. 113–41.
80. Martel C, Pennycook G, Rand DG. Reliance on emotion promotes belief in fake news. Cognit Res Princ Implic. 2020;5:1–20.
81. Valenzuela S, Piña M, Ramírez J. Behavioral effects of framing on social media users: how conflict, economic, human interest, and morality frames drive news sharing. J Commun. 2017;67(5):803–26.
82. Harber KD, Cohen DJ. The emotional broadcaster theory of social sharing. J Lang Soc Psychol. 2005;24(4):382–400.
83. Berger J, Milkman KL. What makes online content viral? J Mark Res. 2012;49(2):192–205.
84. Heath C. Do people prefer to pass along good or bad news? Valence and relevance of news as predictors of transmission propensity. Org Behav Hum Decis Process. 1996;68(1):79–94.
85. Gil de Zúñiga H, González-González P, Goyanes M. Pathways to political persuasion: linking online, social media, and fake news with political attitude change through political discussion. Am Behav Sci. 2021. https://doi.org/10.1177/00027642221118272.
86. Chen CY, Kearney M, Chang SL. Comparative approaches to mis/disinformation: belief in or identification of false news according to the elaboration likelihood model. Int J Commun. 2021;15(1):1263–85.
87. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. New York: Springer; 1986. p. 1–24.
88. Bastick Z. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. Comput Hum Behav. 2021;116: 106633.
89. Cantarella M, Fraccaroli N, Volpe R. Does fake news affect voting behaviour? Res Policy. 2023;52(1): 104628.
90. Bennett WL, Livingston S. The disinformation order: disruptive communication and the decline of democratic institutions. Eur J Commun. 2018;33(2):122–39.
91. Farhall K, Carson A, Wright S, Gibbons A, Lukamto W. Political elites' use of fake news discourse across communications platforms. Int J Commun. 2019;13:4353–75.
92. Asmolov G. The disconnective power of disinformation campaigns. J Int Aff. 2018;71(1.5):69–76.
93. Druckman JN, Klar S, Krupnikov Y, Levendusky M, Ryan JB. Affective polarization, local contexts and public opinion in America. Nat Hum Behav. 2021;5(1):28–38.
94. Lewandowsky S, Stritzke WGK, Freund AM, Oberauer K, Krueger JI. Misinformation, disinformation, and violent conflict: from Iraq and the "War on Terror" to future threats to peace. Am Psychol. 2013;68(7):487–501.
95. Banaji S, Bhat R, Agarwal A, Passsanha N, Pravin MS. WhatsApp vigilantes: an exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. London School of Economics and Political Science. 2019. http://eprints.lse.ac.uk/id/eprint/104316. Accessed 23 Apr 2024.
96. Piazza JA. Fake news: The effects of social media disinformation on domestic terrorism. Dyn Asymmetric Conf. 2022;15(1):55–77.
97. Nisbet EC, Kamenchuk O. Russian news media, digital media, informational learned helplessness, and belief in COVID-19 misinformation. Int J Public Opin Res. 2021;33(3):571–90.
98. Grabner-Kräuter S, Bitter S. Trust in online social networks: a multifaceted perspective. Forum Soc Econ. 2015;44:48–68.

99. Galeotti AE. Believing fake news. In: Condello A, Andina T, editors. Post-Truth, philosophy and law. Milton Park: Routledge; 2019. p. 50–75.
100. Guess A, Nagler J, Tucker J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv. 2019;5(1):1–8.
101. Wagnsson C. The paperboys of Russian messaging: RT/Sputnik audiences as vehicles for malign information influence. Inf Commun Soc. 2022;26(9):1849–67.
102. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science. 1974;185:1124–31.
103. Baptista J. Ethos, pathos e logos. Análise comparativa do processo persuasivo das (fake) news. Eikon. 2020;1:43–54. http://ojs.labcom-ifp.ubi.pt/index.php/eikon/article/view/816.
104. Pantazi M, Kissine M, Klein O. The power of the truth bias: false information affects memory and judgment even in the absence of distraction. Soc Cogn. 2018;36(2):167–98.
105. Knobloch-Westerwick S, Liu L, Hino A, Westerwick A, Johnson BK. Context impacts on confirmation bias: evidence from the 2017 Japanese snap election compared with American and German findings. Commun Res. 2019;45(4):427–49.
106. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen. 2018;147(12):1865–80.
107. Talwar S, Dhir A, Kaur P, Zafar N, Alrasheedy M. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. J Retail Consum Serv. 2019;51:72–82.
108. Nielsen, R. K., & Graves, L. (2017). *"News you don't believe": Audience perspectives on fake news*. Reuters Institute for the Study of Journalism. https://www.digitalnewsreport.org/publications/2017/news-dont-believe-audience-perspectives-fake-news/. Accessed 27 June 2023.
109. Bright J. The social news Gap: how news reading and news sharing diverge. J Commun. 2016;66(3):343–65.
110. Duffy A, Tandoc E, Ling R. Too good to be true, too good not to share: the social utility of fake news. Inf Commun Soc. 2019;23(13):1965–79.
111. Lee CS, Ma L. News sharing in social media: the effect of gratifications and prior experience. Comput Hum Behav. 2012;28(1):331–9.
112. Halpern D, Valenzuela S, Katz J, Miranda JP. From belief in conspiracy theories to trust in others: Which factors influence exposure, believing and sharing fake news. In: Meiselwitz G, editor. International conference on human-computer interaction. Berlin: Springer; 2019. p. 217–32.
113. Lewis R, Marwick AE. Taking the red pill: Ideological motivations for spreading online disinformation. In: Understanding and addressing the disinformation ecosystem. Annenberg School for Communication; 2017. pp. 18–22.
114. Mancosu M, Vassallo S, Vezzoni C. Believing in conspiracy theories: evidence from an exploratory analysis of Italian survey data. South Eur Soc Politics. 2017;2(3):327–44.
115. Sunstein CR. Going to extremes: how like minds unite and divide. Oxford: Oxford University Press; 2009.
116. Rini R. Fake news and partisan epistemology. Kennedy Inst Ethics J. 2017;27(3):43–64.
117. Gaozhao D. Flagging fake news on social media: an experimental study of media consumers' identification of fake news. Gov Inf Q. 2021;38(3): 101591.
118. Kim A, Dennis AR. Says who? The effects of presentation format and source rating on fake news in social media. MIS Q. 2019;43(3):1025–39.
119. Chan MPS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol Sci. 2017;28(11):1531–46.
120. Ecker UK, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, Amazeen MA. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol. 2022;1(1):13–29.
121. Costello TH, Pennycook G, Rand DG. Durably reducing conspiracy beliefs through dialogues with AI. Science. 2024;385(6714): eadq1814.
122. Steingartner W, Možnik D, Galinec D. Disinformation campaigns and resilience in hybrid threats conceptual model. In: 2022 IEEE 16th International Scientific Conference on Informatics (Informatics), IEEE; 2022. pp. 287–92.
123. Guess AM, Lerner M, Lyons B, Montgomery JM, Nyhan B, Reifler J, Sircar N. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proc Natl Acad Sci. 2020;117(27):15536–45.
124. Banas JA, Miller G. Inducing resistance to conspiracy theory propaganda: testing inoculation and metainoculation strategies. Hum Commun Res. 2013;39(2):184–207.
125. Lewandowsky S, Van Der Linden S. Countering misinformation and fake news through inoculation and prebunking. Eur Rev Soc Psychol. 2021;32(2):348–84.
126. Humprecht E, Esser F, Van Aelst P. Resilience to online disinformation: a framework for cross-national comparative research. Int J Press/Politics. 2020;25(3):493–516.
127. Gimpel H, Heger S, Olenberger C, Utz L. The effectiveness of social norms in fighting fake news on social media. J Manag Inf Syst. 2021;38(1):196–221.
128. Marsden C, Meyer T, Brown I. Platform values and democratic elections: how can the law regulate digital disinformation? Comput Law Secur Rev. 2020;36: 105373.
129. Pielemeier J. Disentangling disinformation: what makes regulating disinformation so difficult? Utah Law Rev. 2020;4:917–40.
130. European Commission. Strengthened code of practice on disinformation. 2022. https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation. Accessed 23 Apr 2024.
131. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27(3):379–423.
132. Reiljan A. 'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems. Eur J Polit Res. 2020;59(2):376–96. https://doi.org/10.1111/1475-6765.12351.
133. Wiggins BE. Navigating an immersive narratology: Factors to explain the reception of fakenews.International J E-Politics (IJEP), 2017;8(1):16–29.
134. Munger K, Luca M, Nagler J, Tucker J. The effect of clickbait.2018. https://www.semanticscholar.org/paper/The-Effect-of-Clickbait-*-Munger-Luca/9edda403530e9a45dd0b756ea0b938797dce6a82. Accessed 23 June 2023.
135. Silverman C. Here are 50 of the biggest fake news hits on Facebook from 2016.BuzzfeedNews.2016. https://www.buzzfeednews.com/article/craigsilverman/top-fake-news-of-2016. Accessed 23 June 2023.
136. Weikmann T, Lecheler S. Visual disinformation in a digital age: A literature synthesis andresearch agenda.New Media & Society2022;14 614448221141648.https://doi.org/10.1177/14614448221141648