Journal of Cyberspace Studies
eISSN: 2588-5502
http://jcss.ut.ac.ir
University of Tehran

Disinformation as weaponised ignorance: A hybrid teleological–epistemic framework

Ramin Rambod, Hadi Samadi*, Shahla Eslami
Department of Philosophy, SR.C. Islamic Azad University, Tehran, Iran.
(*Corresponding author: samadi@srbiau.ac.ir, https://orcid.org/0000-0002-6898-5606)

Article Info

Original article

Main Object: Humanities & Social Sciences, Epistemology of Information, Disinformation

Received: 15 October 2025
Revised: 13 December 2025
Accepted: 31 December 2025
Published online: 13 January 2026

Keywords:
bullshit, disinformation, epistemic authority, ignorance-generating content, intentional misleading, knowledge-first epistemology, LLMs.

Abstract

Background: Contemporary debates on disinformation are dominated by two influential approaches: Fallis's functional (teleological) model and Simion's purely epistemic Disinformation as Ignorance-Generating Content (DIGC) model. Their respective strengths and weaknesses become salient in real-world settings such as the Backfire Effect, the spread of bullshit, and disputes about the epistemic status of Large Language Models (LLMs).
Aims: We aim to (i) critically evaluate the explanatory and classificatory utility of the functional and purely epistemic models across these scenarios, (ii) diagnose key failure modes (especially DIGC's over-generation and the functional model's difficulties with non-intentional sources), and (iii) propose a more extensionally adequate framework.
Methodology: We conduct a comparative conceptual analysis of both models and test their classifications against several cases. In particular, we use empirical findings on the Backfire Effect to examine whether a purely consequence-based criterion misclassifies accurate, well-intentioned scientific information. We also incorporate Frankfurt's distinction between lying and bullshit to refine how epistemic malice is characterised.
Findings: Fallis's functional model captures complex forms of disinformation (including true and adaptive disinformation) by tying disinformation to a misleading function, but it struggles to classify outputs from non-intentional sources such as autonomous AI. DIGC broadens coverage by removing intentionality and focusing on dispositions to increase ignorance, yet this purely epistemic stance yields an Over-generation Problem: under Backfire conditions, it can wrongly classify accurate and well-intentioned scientific communication as disinformation. To address these limitations, we propose a hybrid teleological framework, Functional-Contextual Disinformation (FC-DIGC), which combines DIGC's consequence criterion with a teleological constraint requiring a misleading function. This synthesis better separates malicious deception (disinformation) from unintended epistemic harm (contextually harmful misinformation) and helps clarify how LLM outputs should be categorised.
Conclusion: A hybrid teleological approach improves extensional adequacy by preventing over-generation while retaining coverage for non-intentional systems. FC-DIGC provides a principled way to distinguish disinformation from contextually harmful misinformation and, by integrating the lying–bullshit contrast, captures a broader spectrum of epistemically motivated malice relevant to contemporary information environments, including AI-mediated communication.

Cite this article: Rambod R, Samadi H, Eslami Sh. (2026). "Disinformation as weaponised ignorance: A hybrid teleological–epistemic framework". *Cyberspace Studies*. 10(2): 493-512. doi: https://doi.org/10.22059/jcss.2025.403904.1190.

Website: https://jcss.ut.ac.ir/| Email: jcss@ut.ac.ir |
EISSN: 2588-5502
Publisher: University of Tehran

Cyberspace Studies, Vol 10, No 2, July 2026

1. Introduction: Disinformation as an epistemic threat

1.1. The disinformation crisis and practical implications

The global information crisis (often described as an *infodemic* by the World Health Organization) forces a re-examination of the concepts of information, misinformation, and disinformation (Simion, 2024a; Fallis, 2015). Disinformation is not only a practical challenge that threatens democratic processes, public health, and social trust. It is also a first-order epistemological challenge. It reshapes how knowledge is acquired, maintained, and undermined within complex information environments. The consequences of disinformation extend far beyond the formation of isolated false beliefs. They often involve resistance to evidence. Agents come to discredit or ignore reliable and accurate information (Simion, 2024a; 2024b). Understanding this mechanism is essential for building resilient epistemic communities and trust networks. Traditional dictionary-style definitions treat disinformation simply as false information spread deliberately to deceive. Such definitions are no longer adequate. They fail to capture contemporary phenomena, including true disinformation, algorithmically amplified content, adaptive conspiracies, and high-volume bullshit (Fallis, 2014; 2015; Frankfurt, 1988).

1.2. From information to disinformation: Factivity and intentionality

A common strategy in the philosophical literature defines *information* first and then treats misinformation and disinformation as its defective variants (Floridi, 2011: 260-261; Kelp & Simion, 2025). These accounts often build factivity into the concept. Information is taken to entail truth. On this view, misinformation is false content that purports to inform, and disinformation is false content that purports to inform and is produced with an intention to deceive (Fetzer, 2004; Mahon, 2016).

Two problems follow. The first is the factivity constraint, which struggles with cases where factually correct content misleads. Paradigmatic instances of true disinformation work in precisely this way (Fallis, 2014). The Athanasius case ('He is not far away') shows that a speaker can use a true statement (through implicature) to induce a false belief. The second is the intentionality constraint. It fails to capture non-intentional but systematic forms of deception. Adaptive disinformation is one clear example. Sincere believers may maintain and propagate conspiratorial content that benefits them or their groups without any explicit intention to deceive (Fallis, 2015; Skyrms, 2010). These limitations suggest that factivity and individual intention cannot be the only determinants of disinformation. We need a framework that accounts for epistemic consequences, structural functions, and the broader social and technological context.

1.3. Aim and structure of the paper

Our aim is to develop a hybrid teleological account of disinformation that combines the strengths and avoids the weaknesses of two recent approaches. The first is Fallis's functional model, which defines disinformation through a misleading function (Fallis, 2015). The second is Simion's Disinformation as Ignorance-Generating Content (DIGC), which defines disinformation in terms of its disposition to generate ignorance (Simion, 2024a). In Section 2 we introduce the knowledge-first background and present both models. Section 3 extends the analysis to Frankfurt's notion of bullshit and shows how DIGC captures some epistemic harms yet generates an Over-generation Problem. Section 4 sets out the FC-DIGC hybrid teleological framework, explains its internal unity, and addresses major objections. Section 5 tests FC-DIGC on several cases. These include the Backfire Effect, adaptive disinformation, true disinformation, black-box AI, honest journalism, failed lies, and LLMs as putative epistemic authorities. Section 6 draws policy implications and shows how FC-DIGC generalises to electoral, health, and algorithmic contexts. Section 7 concludes and outlines directions for future research.

2. Knowledge-first epistemology and two models of disinformation

2.1. Knowledge-first epistemology and information as possible knowledge

Simion's DIGC model is rooted in knowledge-first epistemology. In this approach, knowledge is the central epistemic primitive (Williamson, 2007; Kelp & Simion, 2025). Classical probabilistic accounts face two familiar difficulties:

- First, there is the absolute probability problem. Assigning probability 1 to any empirical proposition seems inappropriate, yet many accounts of knowledge appear to require such maximal confidence.
- Second, the reference class problem makes it hard to assign a single correct probability to a token event, since the event belongs to many reference classes (Hájek, 2007).

Knowledge-first epistemology bypasses these problems. It takes knowledge and its norms as basic and then defines epistemic notions such as evidence and information in relation to knowledge. On this view, information is possible knowledge. In the tradition of Dretske's information-theoretic account of knowledge (Dretske, 1983), it is content whose function is to generate knowledge in normal conditions (Kelp & Simion, 2025). Symmetrically, disinformation is content whose uptake tends to destroy, block, or prevent knowledge. This background underwrites DIGC's core idea. We should identify disinformation by its epistemic effects and not by its truth value or the producer's mental state.

2.2. Fallis's functional model: Teleology of deception and varieties of disinformation

Fallis (2015) proposes a functional analysis that shifts attention from the mental states of individual deceivers to the function of content in communication systems. In rough terms, a piece of representational content is disinformation if (and only if) it has the function of misleading. It exists or is sustained because it tends to produce false or unwarranted beliefs in some audience (Fallis, 2009; 2014; 2015). This functional approach has deep roots. Augustine classified lies according to their moral purpose and not merely their surface form. What matters is whether the speaker uses words in a way that conflicts with the proper aim of speech (Fallis, 2014). Chisholm and Feehan (1977) also emphasise the epistemic goal of communication. Deception makes a hearer epistemically worse off than they could have been.

Fallis reformulates these ideas in a modern functional idiom. He distinguishes between design functions (intentional aims built into a signal by a designer) and etiological functions (roles that a signal plays because those roles have historically benefited the signal producer or group). This model handles a range of contested cases. It excludes jokes, obvious sarcasm, and implausible lies, since they are not plausibly misleading in normal conditions (Fallis, 2015). It includes visual disinformation such as doctored photographs or misleading graphs, because it does not require language. It also includes true disinformation, where factually correct statements generate false beliefs through implicature (Fallis, 2014). And it includes adaptive disinformation, where content is sustained because it benefits its producers even when individual propagators sincerely believe it (Fallis, 2015; Skyrms, 2010).

However, the functional model faces two central problems. The first is the disjunctive analysis problem. Function is defined in terms of either design intention or systemic benefit. This may undermine theoretical unity (Kingsbury & McKeown-Green, 2009). The second is the problem of non-functional agents such as black-box AI systems. These systems generate ignorance-producing content without clear intent or benefit for human agents (Simion, 2024a). These difficulties motivate a move toward a purely epistemic criterion.

2.3. Simion's DIGC model: Disinformation as ignorance-generating content

Simion (2024a) proposes that we define disinformation directly through its epistemic effects. On DIGC, content (r) is disinformation if, in normal conditions, its uptake decreases a relevant audience's warranted evidential probability for a true proposition (p). Formally, this means that ($\Pr(p \mid r) < \Pr(p)$). The focus is not on the producer's mental state or the historical function of the content. It is on the disposition to generate ignorance in a given context. DIGC identifies several

mechanisms through which disinformation can generate ignorance (Simion, 2024a). These include spreading false belief, misleading defeat where undercutting evidence strips knowledge, and inducing epistemic anxiety where raising irrelevant error possibilities or stakes causes agents to lose knowledge. DIGC also recognises confidence-defeating disinformation, where justified confidence in a truth drops enough to break knowledge, and it covers pragmatic mechanisms where true assertions carry false implicatures. One of DIGC's strengths is that it can classify black-box AI output and well-intentioned but epistemically harmful communication as disinformation. The model ignores intent and tracks only ignorance generation. It also naturally accommodates Frankfurt's notion of bullshit, as we see in the next section. At the same time, the purely epistemic focus creates Over-generation and operational problems. These problems motivate our hybrid FC-DIGC framework.

3. Bullshit, epistemic harm, and the limits of a purely epistemic account

3.1. Lying, bullshit, and attitudes to truth

The philosophical literature usually treats lying and deception as conceptual opposites of truth-telling (Adler, 1997; Mahon, 2016). Harry Frankfurt (1988) complicates this picture by introducing the category of bullshit. For Frankfurt, the liar is still concerned with truth. The liar knows, or takes themselves to know, the truth and seeks to hide it or induce belief in its negation. Lying is an act of opposition to truth. The bullshitter is different. The bullshitter is indifferent to truth. The aim is not to represent facts correctly or incorrectly, but to shape an audience's attitudes. These may be attitudes toward the speaker, toward a group, or toward some practical project. Bullshit may accidentally be true or false. What defines it is not its content but the producer's lack of concern for accuracy. Frankfurt argues that this indifference is often more corrosive than ordinary lying. It undermines the value of truth itself and erodes the norms that structure sincere communication. Contemporary information environments contain not only lies but also high volumes of bullshit. This raises a natural question. Can a purely epistemic account like DIGC capture the harms of bullshit without reference to attitudes such as indifference?

3.2. Bullshit, DIGC, and epistemic anxiety

DIGC seems able to answer this question in the affirmative. Bullshit is often produced without adequate evidence and in violation of Grice's maxims, especially the second maxim of quality ('Do not say that for which you lack adequate evidence'; Grice, 1989; Fallis, 2014). High-volume bullshit can overwhelm the evidential background that agents use to assess testimony. It can create epistemic anxiety and misleading defeat (Simion, 2024a). Agents who previously knew (p) may fall into

doubt or confusion. They lose knowledge because the steady supply of bullshit inflates perceived alternatives or undermines source reliability.

In such cases, bullshit satisfies DIGC. Uptake of the content tends to reduce agents' closeness to knowledge. It does so either by inducing false beliefs or by stripping existing knowledge. This shows that a purely epistemic consequence model can, at least in principle, capture a phenomenon that Frankfurt originally framed in attitudinal terms. However, once DIGC is applied more broadly, and especially once it is combined with empirical work on Backfire effects, it begins to classify as disinformation some communications that are normatively very different from disinformation. On Simion's view, therefore, high-volume bullshit that predictably generates such epistemic anxiety will straightforwardly qualify as disinformation. FC-DIGC will later refine this verdict by distinguishing between bullshit that is embedded in a misleading function and bullshit that merely pollutes the epistemic environment.

3.3. Over-generation and the backfire effect

The most serious difficulty for DIGC is the Over-generation Problem. Since DIGC classifies disinformation solely in terms of epistemic effect, it can label some accurate and well-intentioned communication as disinformation. This is not just a theoretical puzzle. It arises in public health contexts where conceptual clarity is crucial. Empirical work on vaccine hesitancy documents the MMR vaccine Backfire Effect (Nyhan et al., 2014; Pluviano et al., 2017). In these studies, vaccine-sceptical parents received scientifically accurate and evidence-based information about the safety of the MMR vaccine. Instead of revising their beliefs toward the evidence, many participants (especially in specific subgroups) emerged more convinced of the false claim that vaccines cause autism. In probabilistic terms, for these recipients the conditional probability of the true proposition (p) ('MMR vaccines do not cause autism') given the corrective content (c) was lower than their prior probability. So, ($\Pr(p \mid c) < \Pr(p)$) (Ausili, 2025).

On Simion's criterion, this is enough to classify the corrective content as disinformation. The content has the disposition (in this context) to undermine warranted confidence in a true proposition and therefore counts as ignorance-generating content (DIGC+). The model is committed to the claim that carefully vetted and scientifically accurate public health messaging, designed to combat disinformation, is itself disinformation (Simion, 2024a; Ausili, 2025). This verdict is both counterintuitive and normatively troubling. It collapses the distinction between malicious epistemic sabotage and well-intentioned communication that misfires in a hostile environment. From a policy perspective, this creates serious confusion.

If we treat such corrective campaigns as disinformation, we lose the normative difference between actors who weaponise ignorance and

actors who attempt (and sometimes fail) to reduce it. The Backfire case therefore exposes a structural limitation in a purely epistemic-consequence model. Epistemic effect alone cannot define disinformation. The FC-DIGC framework arises from this tension. It preserves DIGC's focus on epistemic harm and accepts that Backfire-style interventions do generate ignorance in some audiences. At the same time, it adds a teleological condition. The ignorance-generating disposition must be part of a non-accidental misleading function. In Backfire cases, this second condition is not satisfied. The design and history of the campaign aim at countering deception and promoting public health and not at gaining from the spread of falsehoods. FC-DIGC therefore classifies such content as contextually harmful misinformation and not as disinformation.

4. The FC-DIGC hybrid teleological framework

4.1. Definition and motivation

The previous discussion suggests that neither Fallis's functional account nor Simion's DIGC account is sufficient on its own. Fallis is right that disinformation involves a teleological dimension. It is not merely misleading content but content that exists or is sustained for the sake of misleading. Simion is right that the distinctive harm of disinformation is epistemic. Disinformation generates ignorance, anxiety, or unwarranted doubt.

The 'Functional-Contextual Disinformation (FC-DIGC)' framework combines these insights. In simplified form:

FC-DIGC: A token of content (*X*) in context (*C*) is disinformation if and only if:

- *X* has a disposition in *C* to generate or increase ignorance in a relevant audience (the *DIGC condition*), and
- This ignorance-generating disposition is part of a misleading function of *X*, realised through design intention, foreseen intent, or systemic and etiological advantage (the *Function condition*).

The DIGC component captures the epistemic profile of disinformation. It ensures that disinformation is not just content that happens to be false or misleading. It is content whose uptake typically dislodges knowledge or blocks its acquisition. The FC component introduces a teleological and normative asymmetry. Disinformation is not only epistemically harmful. It is also non-accidentally embedded in practices or systems that exploit this harm for some benefit.

This hybrid analysis addresses the main problems noted earlier. By requiring the FC condition, FC-DIGC avoids Simion's Over-generation Problem. Backfire-effect cases satisfy DIGC but not FC. They are therefore classified as contextually harmful misinformation rather than disinformation. At the same time, by retaining the DIGC condition, the framework avoids reverting to a purely intentional or function-only

account. It ensures that disinformation remains tied to serious epistemic threat within an information environment. The misleading function is specified independently of the concept of disinformation, through design intentions, foreseen use, and stable benefit patterns in communicative practices. The framework also preserves the advantages of functionalism for non-paradigmatic cases. It can treat certain AI-mediated campaigns, algorithmic systems, or institutional strategies as disinformation even where no single agent intends deception. In such cases, the misleading function is realised at the level of a broader socio-technical system. The system's design history and benefit patterns mirror the core features of Fallis-style disinformation (Fallis, 2015; Simion, 2024b: 178-191). This does not mean that any ignorance-generating AI output counts as disinformation. Purely autonomous black-box systems that produce misleading content without being embedded in such human-guided practices are best treated as sources of epistemic pollution rather than disinformation.

4.2. Epistemic threat and normative priority

FC-DIGC is best viewed as an account of a specific kind of epistemic threat. The starting point is not moral condemnation or political illegitimacy. The central question is what distinguishes disinformation from other forms of misleading or defective communication within an epistemic ecology. On this proposal, disinformation is a practice that systematically undermines the prospects for knowledge and is organised around this effect. The DIGC condition isolates the threat. The FC condition explains why the threat is not accidental. This ordering has theoretical and policy consequences. In public debate, there is a temptation to use 'disinformation' as a catch-all label for morally or politically objectionable speech. FC-DIGC resists that temptation.

Because the DIGC condition is epistemic rather than moral, the framework allows that some morally repugnant speech is not disinformation. It also allows that some disinformation may arise in morally mixed contexts. Because the FC condition is teleological, the framework holds that not every epistemic harm (however serious) qualifies as disinformation. The term is reserved for harms that arise within practices whose function is to make epistemic agents worse off. This focus on epistemic threat clarifies the relation between disinformation and neighbouring notions such as misinformation, epistemic pollution, and bullshit. Contextually harmful misinformation is DIGC+ but FC−. It generates ignorance without being part of a misleading function.

Epistemic pollution is a broader category. It includes content and structural features that degrade the information environment. Some instances may satisfy one or both of the DIGC and Function conditions at the level of particular tokens, but without forming part of a stable

practice of weaponising ignorance. Bullshit, in Frankfurt's sense, can fall under disinformation or pollution. If indifference to truth is used as a stable misleading function, then bullshit is disinformation. If that function is absent, bullshit is a source of pollution (Frankfurt, 1988). Contextually harmful misinformation is one important source of epistemic pollution, but not the only one. Structural features such as ranking algorithms or high-volume bullshit can pollute the infosphere even when particular tokens do not meet both disinformation conditions.

FC-DIGC thus offers a taxonomy centred on epistemic threat. It aims at a concept of disinformation that is both normatively robust and extensionally disciplined. It preserves the moral and political urgency of the topic by recognising patterns of exploitation and domination. At the same time, it grounds this urgency in an explicitly epistemic diagnosis. Disinformation is a way of weaponising ignorance.

4.3. Objections and replies

Several objections arise once teleology is reintroduced.

One worry concerns coverage for autonomous AI. The misleading function requirement seems to narrow the range of cases. DIGC can classify content produced by autonomous black-box AI as disinformation (Simion, 2024a), while FC-DIGC often will not. The response is that this is a justified cost. The concept of disinformation, especially for legal and policy use, should preserve a sense of malice or systemic advantage. Content that generates ignorance without such a teleological basis is better treated as epistemic pollution. This calls for platform-level and environmental interventions and not for the sanctions that attach to disinformation (Simion, 2024b: 178-191). In this respect FC-DIGC agrees with Fallis: standalone black-box AI output is pollution, not disinformation. The model nonetheless covers AI-enabled disinformation at the level of human-driven socio-technical systems.

A second objection is the disjunctive analysis problem. FC-DIGC inherits the use of a disjunctive notion of function (design or etiological). Critics argue that this undermines unity (Kingsbury & McKeown-Green, 2009). FC-DIGC replies that both design and etiological functions are species of one broader kind. Each is a case of non-accidental, teleologically grounded ignorance generation. Whether the function is intentionally built in or stabilised by systemic benefit, what matters is that ignorance is exploited and not merely produced by accident.

A third concern is the operationalisation of function. Misleading functions are historically grounded and often opaque. This may make them hard to identify in practice. FC-DIGC mitigates this by giving DIGC a primary role in detection. Once content is flagged as ignorance-generating, function can be assessed using tractable indicators. These

may include direct intent markers (for example internal plans and documents) and systemic benefit patterns (for example financial or strategic gain) (Fallis, 2014; 2015). The hybrid approach shifts the burden from subjective intention to observable patterns of advantage.

A fourth concern is the exclusion of significant contextual harm. Some worry that FC-DIGC excludes serious cases, such as the well-intentioned journalist whose accurate report generates public damage by creating unwarranted scepticism (Simion, 2024a). FC-DIGC accepts that these harms are important. It nevertheless insists on keeping a normative boundary. Such content is contextually harmful misinformation and not disinformation. The correct response is not punishment, but tools such as epistemic literacy, better risk communication, and social interventions that aim to restore trust (Ausili, 2025).

5. Case studies and extensional adequacy

5.1. Backfire and contextually harmful misinformation

We can now return to the Backfire Effect as a test of extensional adequacy. A public health authority disseminates accurate and evidence-based information about the safety of the MMR vaccine. The explicit aim is to reduce ignorance and counter disinformation. In a subset of highly sceptical recipients, entrenched distrust and motivated reasoning lead to an increase in false belief. After exposure, these recipients are more convinced that vaccines cause autism (Nyhan et al., 2014; Pluviano et al., 2017).

On DIGC, the content (c) lowers the warranted probability of the true proposition (p) for this audience. So ($\Pr(p \mid c) < \Pr(p)$). The content is therefore disinformation (DIGC+). The model correctly captures epistemic harm but classifies a paradigm case of sincere and scientifically grounded communication as disinformation (Simion, 2024a; Ausili, 2025).

On Fallis's functional model, the content is not disinformation. It has no misleading function. The communicators do not intend to mislead, and the message does not play a role in a system that benefits from error. The model preserves the intuitive classification but offers less detail about the mechanism of harm.

FC-DIGC combines these strengths. The Backfire message satisfies the DIGC condition. It has a disposition to generate ignorance in the relevant context. It fails the function condition. There is no malicious design, no foreseen intent to mislead, and no benefit pattern that treats the resulting ignorance as a feature rather than a failure. FC-DIGC therefore classifies Backfire-style interventions as contextually harmful misinformation. This preserves a precise explanation of epistemic harm, avoids penalising public health authorities for audience-side bias, and maintains a clear boundary between organised epistemic threat and non-malicious failure (Ausili, 2025).

5.2. Operational failures: Retroactive disinformation and contradictory outcomes

The Over-generation Problem is accompanied by more practical operational failures when DIGC is used in large-scale classification (Ausili, 2025).

The first is retroactive disinformation. Because DIGC defines disinformation only through ignorance generation, it can retrospectively reclassify past communication as disinformation when the epistemic environment shifts. Consider the early stages of a public health crisis. Experts issue cautious statements based on limited evidence. Later, more robust data become available. The earlier statements may then mislead new audiences who encounter them without context. On a strict DIGC reading, the original content now has a disposition to generate ignorance. It therefore counts as disinformation, even though it was responsible and evidence-based when first produced. This retroactive reclassification blurs the line between responsible communication under uncertainty and genuine disinformation campaigns. It makes 'disinformation' a moving target. The same utterance may oscillate between 'information' and 'disinformation' as background conditions change, even when its role in communication practices remains the same.

The second operational problem concerns contradictory outcomes. The context-sensitivity of ignorance generation means that the same content may count as disinformation relative to one audience but not another. This depends on their prior attitudes and cognitive dispositions. Sometimes this is harmless. For instance, a sophisticated audience may resist a misleading message that deceives less informed recipients. In other cases, we get contradictory classifications across similar audiences. Small differences in trust, emotion, or framing may produce opposing epistemic outcomes. For one group, the content reduces ignorance. For another, it increases it.

DIGC can represent this formally. It can say that the content is disinformation relative to context C_1 but not to context C_2. However, this is operationally unhelpful for institutions that must detect, regulate, or sanction disinformation. If classification turns on fine-grained and opaque features of individual uptake, large-scale classification becomes unmanageable. FC-DIGC addresses these problems by anchoring classification in a more stable production-side constraint. Ignorance generation (DIGC) remains necessary but is no longer sufficient. Content must also be embedded in a misleading function, realised in design intentions, organisational strategies, or systemic patterns of benefit.

In retroactive cases, FC-DIGC distinguishes between the original responsible communication and later re-use of the same content. The initial statements lack a misleading function and are not disinformation. Later campaigns that reuse the same statements to exploit new

misunderstandings may satisfy the FC condition. The classification then attaches to the later practice, not to the original testimony. In contradictory-outcome cases, FC-DIGC allows epistemic effects to vary across audiences while keeping the disinformation verdict stable wherever a misleading function is present. If a state campaign, a troll farm, or a platform-level strategy is organised around exploiting such effects, the resulting content is disinformation even if some audiences resist it. Where ignorance generation is purely accidental, the content is contextually harmful misinformation rather than disinformation. These operational failures also suggest that the notion of 'normal conditions' in DIGC is under severe strain in highly polarised or rapidly changing environments. FC-DIGC sidesteps this strain by treating ignorance generation as necessary but by anchoring disinformation status in more stable features of communicative practices.

5.3. Adaptive and true disinformation

Adaptive disinformation (AD) involves non-intentional but systematic deception. This includes the spread of conspiracy theories by sincere believers and some deceptive animal signals (Fallis, 2015; Skyrms, 2010). The content generates ignorance (DIGC+) and is sustained because this ignorance benefits producers or their groups (FC+). Fallis's model and DIGC both treat AD as disinformation. FC-DIGC agrees but emphasises that the classification rests on both epistemic consequence and non-accidental function.

True disinformation (TD) involves factually correct statements that produce false beliefs through implicature. The Athanasius 'not far away' case is the standard example (Adler, 1997; Fallis, 2014). The speaker exploits pragmatic mechanisms (DIGC+) and intends to mislead (FC+). Fallis's model, DIGC, and FC-DIGC all classify TD as disinformation. FC-DIGC's contribution is structural. It encourages detection systems to identify the misleading implicature (DIGC) and then establish the teleological basis (FC).

5.4. Black-box AI, honest journalism, and failed lies

Black-box AI. In black-box AI scenarios, autonomous systems generate false or misleading content without human intention or clear systemic benefit (Simion, 2024a). DIGC classifies such content as disinformation because it generates ignorance. Fallis's model excludes it, and FC-DIGC also excludes it. The ignorance-generating disposition is present (DIGC+), but there is no misleading function with a human agent or system as its bearer (FC−). FC-DIGC treats such content as epistemic pollution and directs policy attention to platform accountability and environmental clean-up (Simion, 2024b: 178-191).

The well-intentioned journalist. A journalist publishes an accurate report on scientific disagreement. In a polarised context, the report generates unwarranted scepticism and public harm (Simion, 2024a).

DIGC classifies the report as disinformation (DIGC+). Fallis's model and FC-DIGC both classify it as not disinformation (FC−). The journalist's function is to inform honestly (Ausili, 2025). FC-DIGC thus avoids blaming honest error and preserves a category of contextually harmful misinformation.

Malicious lies that fail to mislead. Consider intentional attempts at deception that are so implausible or poorly executed that no reasonable audience is misled. DIGC classifies these as not disinformation (DIGC−). Fallis's model also excludes them (FC−), since they are not misleading in normal conditions (Fallis, 2015). FC-DIGC agrees. FC-DIGC agrees. In these cases, both the function and consequence conditions fail (FC− and DIGC−). There is malicious intent at the level of the agent, but no misleading function and no ignorance generation. This confirms the intuitive idea that failed deception is not disinformation.

5.5. LLMs and epistemic authority

The debate over the epistemic status of Large Language Models can be understood as a dispute about epistemic authority. A source has epistemic authority when others may rationally defer to it when forming beliefs. Hauswald (2025) argues that we should revise the standard notion of epistemic authority. On his view, highly reliable 'truth indicators' such as LLMs can count as authorities in virtue of their functional role. Mizrahi (2025) disagrees. He claims that LLMs are 'bullshit machines' (in Frankfurt's sense) and do not deserve epistemic respect (see also Goodwin, 2011). Frankfurt's analysis clarifies the dispute. Lying involves deliberate opposition to truth. Bullshit involves indifference to truth (Frankfurt, 1988). LLMs are optimised to produce fluent continuations of text and not to track truth. They sometimes output truths and sometimes falsehoods. The underlying generative process is not guided by veracity.

In this respect, their outputs resemble bullshit. They may be accurate or inaccurate, but truth is not the guiding norm. FC-DIGC allows a more precise classification of LLM outputs. Consider autonomous LLM output, detached from explicit human campaigns. In many contexts, such output satisfies DIGC. It can generate or increase ignorance (DIGC+). Yet, as such, it lacks a misleading function. There is no stable design or etiological history that ties the behaviour of the system to human malice or to systematic human advantage beyond general aims such as efficiency or engagement. The FC condition is not met (FC−). The content is best described as epistemic pollution, not disinformation. This diagnosis helps to adjudicate the Hauswald–Mizrahi dispute. Full epistemic authority requires more than reliable performance. It requires a responsible subject that can answer for its assertions and whose attitudes toward truth are normatively assessable (Goodwin, 2011; Mizrahi, 2025). LLMs, considered as stand-alone agents, do not satisfy

this requirement. FC-DIGC therefore supports Mizrahi's conclusion that LLMs do not deserve epistemic respect as authorities. At the same time, the framework preserves Hauswald's insight that LLMs can serve as useful truth indicators in restricted and supervised contexts. They can assist epistemic agents without themselves being authorities.

FC-DIGC also clarifies where policy responsibility lies. If LLM outputs count as epistemic pollution rather than disinformation, regulation should not treat the systems as if they were malicious agents. The focus should be on the human designers and deployers who use these systems in ways that systematically generate ignorance. When LLMs are integrated into human campaigns that exploit ignorance for advantage, the FC condition is satisfied at the level of the socio-technical system. FC-DIGC will then classify the resulting content as disinformation. The framework reconciles Frankfurt's concept of bullshit with contemporary debates on AI and epistemic authority. This system-level perspective on LLMs as components of broader socio-technical practices mirrors the treatment of algorithmic amplification in Section 6.2. The next subsection summarises these classifications in Table 1, which sets out how the three accounts treat each of the central cases.

5.6. Comparative summary

The classifications in Sections 5.1–5.5 are collected in Table 1. The table contrasts how Fallis's functional model, Simion's DIGC model, and the FC-DIGC hybrid treat Backfire, adaptive and true disinformation, black-box AI and autonomous LLM output, the well-intentioned journalist, failed lies, and weaponised bullshit. As the table shows, FC-DIGC consistently avoids the Over-generation Problem, classifies Backfire and honest journalism as contextually harmful misinformation rather than disinformation, treats autonomous AI and LLM output as epistemic pollution rather than malicious deception, and reserves the label "disinformation" for cases that are both ignorance-generating and embedded in a misleading function. In this way, the hybrid teleological model preserves normative distinctions between malice and honest error, and offers policy clarity on AI, LLMs, and structural forms of epistemic pollution.

6. Policy implications and generalisation

6.1. Tracking, regulation, and epistemic pollution

The FC-DIGC framework has direct implications for detection and policy. It suggests that effective counter-disinformation efforts must go beyond simple fact-checking (Kumar & Geethakumari, 2014; Simion, 2024a). Detection systems should integrate both pragmatic analysis and functional analysis. Pragmatic analysis can detect misleading implicatures and other phenomena highlighted by DIGC (Simion, 2024a).

Table 1. Comparative analysis of disinformation models

Case	Functional Model (FC)	Consequencial Model (DIGC)	FC-DIGC classification	FC-DIGC advantage
Backfire effect (accurate vaccine message that backfires)	Not disinformation (no misleading function)	Disinformation (ignorance-generating)	Not disinformation (contextually harmful misinformation)	Avoids over-generation and preserves normative asymmetry
Adaptive disinformation (sincere conspiracy believers)	Disinformation (systemic benefit)	Disinformation (ignorance-generating)	Disinformation (DIGC+ and FC+)	Confirms both systemic function and epistemic threat
True disinformation (Athanasius-style truthful deception)	Disinformation (pragmatic misleading function)	Disinformation (ignorance-generating via implicature)	Disinformation (pragmatic weaponisation)	High theoretical support
Black-Box AI / autonomous LLM output (Autonomous falsehood)	Not disinformation (no human misleading function)	Disinformation (ignorance-generating)	Not disinformation (epistemic pollution)	Distinguishes pollution from disinformation and avoids conflating malice with malfunction
Well-intentioned journalist (accurate but harmful reporting)	Not disinformation (aim to inform)	Disinformation (ignorance-generating)	Not disinformation (contextually harmful misinformation)	Avoids penalising honest error and keeps a space for non-malicious harm
Implausible lie (malicious intent, no uptake)	Not disinformation (not misleading in normal conditions)	Not disinformation (no ignorance generated)	Not disinformation (failed deception)	No divergence from existing models; baseline consistency checks on clear non-disinformation cases.
Weaponised bullshit (truth indifference in organised campaigns)	Disinformation (misleading function based on indifference to truth)	Disinformation (ignorance-generating)	Disinformation when embedded in a misleading function; otherwise, epistemic	Captures indifference to truth as epistemic malice while allowing non-weaponised bullshit to count as pollution

Note. FC= function condition (non-accidental misleading function); DIGC= disinformation as ignorance-generating content. “Contextually harmful misinformation” refers to cases that are ignorance-generating but lack a misleading function. “Epistemic pollution” refers to broader degradations of the information environment that do not qualify as disinformation under FC-DIGC.

Functional analysis can trace patterns of intent and systemic benefit that indicate misleading functions (Fallis, 2014; 2015). On the policy side, FC-DIGC's distinction between disinformation (FC+ and DIGC+) and contextually harmful misinformation (DIGC+ and FC−) supports a two-track response. Legal and technical sanctions (such as removal, demonetisation, or sanctions on state actors) should target disinformation. Social and educational measures (such as epistemic literacy, trust repair, and improved risk communication) should target harmful but non-malicious content (Ausili, 2025). This two-track approach helps to protect freedom of expression while still regulating acts of deception and indifference that exploit ignorance.

6.2. Beyond these cases: Where FC-DIGC applies

The FC-DIGC framework is not limited to the cases discussed above. It is designed as a general model for classifying epistemically harmful communication in complex information environments. Here we sketch applications in three further domains:

- **Electoral and democratic disinformation.** Electoral disinformation campaigns are central examples of FC-DIGC disinformation. At the DIGC level, such campaigns aim to increase ignorance or systematic error concerning procedures, candidate positions, vote counts, or institutional integrity. They exploit cognitive and social mechanisms (for instance source confusion, motivated reasoning, and in-group trust) to produce resilient false beliefs. At the FC level, they display stable teleological profiles. Their production and dissemination are organised to secure partisan, financial, or geopolitical advantage by undermining warranted confidence in democratic processes. FC-DIGC thus classifies them as disinformation in the strict sense. In contrast, grassroots sharing of misinterpreted information may be DIGC+ but FC−. That activity is better classified as contextually harmful misinformation and addressed through education and deliberation rather than criminalisation.
- **Health and climate disinformation.** Vaccine refusal and climate denial are often sustained by content that neutralises expert testimony and generates ignorance or unwarranted doubt about well-established science. At the DIGC level, such content destabilises trust in epistemic authorities and reframes uncertainty in ways that obscure the evidential landscape. At the FC level, many campaigns are funded or coordinated by actors who benefit from public confusion (for example through delayed regulation or preserved consumption patterns). FC-DIGC again distinguishes malignant campaigns (FC+ and DIGC+) from non-malicious misunderstanding (DIGC+ and FC−). This supports differentiated responses in health and environmental policy.
- **Algorithmic amplification and epistemic pollution.**

Recommendation engines and engagement-driven feeds can create large-scale epistemic pollution when they systematically promote content that increases ignorance, polarisation, or unwarranted scepticism. In many cases, this pollution satisfies DIGC. It predictably generates epistemic distortions even when individual items are not flatly false. Whether such systems satisfy the FC condition depends on their objectives and governance. If design and maintenance choices privilege engagement or profit in ways that produce stable patterns of ignorance that benefit specific actors, the resulting socio-technical configuration may have a misleading function. In that case FC-DIGC classifies the amplified content as disinformation at the system level. If pollution arises as an unintended side-effect of otherwise neutral or benevolent goals, the FC condition is not satisfied. FC-DIGC then calls for redesign and regulation of algorithms and governance structures, rather than for treating individual users as disinformers.

Across these domains, FC-DIGC offers a unified way to separate organised epistemic threat from wider and more heterogeneous forms of informational harm.

7. Conclusion and Future directions

7.1. Summary of critique and model

We have argued that disinformation is best understood as weaponised ignorance: content or practices that systematically undermine knowledge in ways that are not accidental but functionally entrenched. Existing accounts tend to privilege either functional malice (Fallis) or epistemic consequence (Simion). The FC-DIGC framework brings these strands together. On FC-DIGC, disinformation is present only when two conditions are satisfied. First, the content must be ignorance-generating in the DIGC sense: its uptake, in normal conditions, tends to decrease warranted confidence in truths or to strip existing knowledge. Second, this ignorance-generating disposition must be part of a misleading function, realised through design intention, foreseen intent, or systemic and etiological advantage. This dual constraint captures the idea of weaponisation. Ignorance is not merely produced; it is exploited.

This hybrid account resolves the main weaknesses of purely functional or purely epistemic approaches. It answers the over-generation problem by refusing to classify accurate but contextually harmful communication (such as Backfire-style vaccine messaging or honest reporting that misfires) as disinformation. These remain serious forms of contextually harmful misinformation, but they are not cases of weaponised ignorance. The model also introduces the category of epistemic pollution for structural degradations of the infosphere (such as autonomous AI or LLM hallucinations) that are ignorance-

generating but lack a misleading function. FC-DIGC unifies Fallis's insights on teleology with Simion's focus on epistemic consequence and integrates Frankfurt's analysis of lying and bullshit. It treats both opposition to truth (lying) and functional indifference to truth (weaponised bullshit) as routes to disinformation, provided the content plays a misleading function. At the same time, it offers policy clarity on LLMs and other automated systems by classifying most autonomous output as epistemic pollution and locating responsibility in the human designers, deployers, and platforms that decide whether and how to weaponise these tools.

In short, FC-DIGC preserves the normative urgency of the topic while tightening its conceptual core. Disinformation, on this view, is not just bad information. It is ignorance deliberately or systematically put to work.

7.2. Suggestions for future research

The idea of disinformation as weaponised ignorance opens several lines for further inquiry:

- More work is needed on the epistemic status of LLMs and other AI systems. FC-DIGC provides reasons to deny full epistemic authority to these systems while allowing them a limited role as supervised truth indicators. Future research should refine criteria for responsible reliance on such tools, especially in high-stakes domains, without collapsing the distinction between artificial indicators and genuine epistemic agents.
- FC-DIGC highlights the need for a richer social epistemology of normal conditions. In polarised environments, echo chambers, structural distrust, and affective dynamics all shape how content impacts belief. A more detailed account of normal conditions would improve the application of both DIGC and FC-DIGC, and sharpen the boundary between disinformation, contextually harmful misinformation, and mere epistemic noise.
- The mechanisms through which bullshit generates epistemic anxiety and misleading defeat invite formal modelling. Probabilistic and dynamic models of evidential dilution, source confusion, and anxiety-driven doubt could support the design of detection tools and help distinguish weaponised bullshit from non-malicious noise in large datasets.
- The category of epistemic pollution raises questions about AI governance and platform responsibility. If much of what large-scale systems produce is pollution rather than disinformation, regulatory frameworks should focus on the duties of designers, deployers, and intermediaries to mitigate contamination of the infosphere, rather than on attributing malicious intent to non-agentic systems.

Taken together, these directions develop the central claim of this paper: *disinformation is not simply falsehood or error, but ignorance systematically mobilised as a tool*. Understanding disinformation as weaponised ignorance allows us to draw principled lines between malice and mistake, between pollution and attack, and between technological malfunction and genuinely teleological uses of deception in the contemporary infosphere.

Conflict of interest

The authors declared no conflicts of interest.

Ethical considerations

The authors have completely considered ethical issues, including informed consent, plagiarism, data fabrication, misconduct, and/or falsification, double publication and/or redundancy, submission, etc. This article was not authored by artificial intelligence. AI Agents Used for literature survey, Grammar and Stylish writing revisions.

Data availability

The dataset generated and analyzed during the current study is available from the author on reasonable request.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

References

Adler, J.E. (1997). "Lying, deceiving, or falsely implicating". *Journal of Philosophy*. 94(9): 435-452. https://doi.org/10.2307/2564617.

Ausili, L. (2025). "The inseparable link between disinformation and attitudes". *Inquiry*. 1-8. https://doi.org/10.1080/0020174X.2025.2542393.

Chisholm, R.M. & Feehan, T.D. (1977). "The intent to deceive". *Journal of Philosophy*. 74(3): 143-159. https://doi.org/10.2307/2025605.

Dretske, F.I. (1983). "Précis of *Knowledge and the Flow of Information"*. *Behavioral and Brain Sciences*. 6(1): 55-90. https://doi.org/10.1017/S0140525X00014631.

Fallis, D. (2015). "What is disinformation?". *Library Trends*. 63(3): 401-426. https://doi.org/10.1353/lib.2015.0014.

----------. (2014). "The varieties of disinformation". In L. Floridi & P. Illari (Eds.). *The Philosophy of Information Quality* (pp. 135–161). Springer. https://doi.org/10.1007/978-3-319-07121-3_8.

----------. (2009). "A conceptual analysis of disinformation". *iConference Proceedings*. http://hdl.handle.net/2142/15205.

Fetzer, J.H. (2004). "Disinformation: The use of false information". *Minds and Machines*. 14(2): 231-240. https://doi.org/10.1023/B:MIND.0000021683.28604.5b.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199232383.001.0001.

Frankfurt, H.G. (1988). "On bullshit". In *The Importance of What We Care About: Philosophical Essays* (pp. 117–133). Cambridge University Press. https://doi.org/10.1017/CBO9780511818172.011

Goodwin, J. (2011). "Accounting for the appeal to the authority of experts". *Argumentation.* 25(3): 285-296. https://doi.org/10.1007/s10503-011-9219-6.

Grice, H.P. (1989). *Studies in the Way of Words*. Harvard University Press.

Hájek, A. (2007). "The reference class problem is your problem too". *Synthese*. 156: 563-585. https://doi.org/10.1007/s11229-006-9138-5.

Hauswald, R. (2025). "Artificial epistemic authorities". *Social Epistemology*. 39: 716-725. https://doi.org/10.1080/02691728.2025.2449602.

Kelp, C. & Simion, M. (2025). "What is information?". *Aristotelian Society Supplementary Volume*. 99(1): 189-208. https://doi.org/10.1093/arisup/akaf008.

Kingsbury, J. & McKeown-Green, J. (2009). "Definitions: Does disjunction mean dysfunction?". *Journal of Philosophy*. 106(10): 568-585. https://doi.org/10.5840/jphil20091061034.

Kumar, K.P.K. & Geethakumari, G. (2014). "Detecting misinformation in online social networks using cognitive psychology". *Human-centric Computing and Information Sciences*. 4(1): 22. https://doi.org/10.1186/s13673-014-0014-x.

Mahon, J.E. (2016). "The Definition of lying and deception". *The Stanford Encyclopedia of Philosophy*. Winter. E.N. Zalta (ed.). https://plato.stanford.edu/archives/win2016/entries/lying-definition.

Mizrahi, M. (2025). "No epistemic respect for bullshit machines or LLMs". *Social Epistemology Review and Reply Collective*. 14(9): 138-146. https://wp.me/p1Bfg0-amy.

Nyhan, B.; Reifler, J.; Richey, S. & Freed, G.L. (2014). "Effective messages in vaccine promotion: A randomized trial". *Pediatrics*. 133(4): e835-e842. https://doi.org/10.1542/peds.2013-2365.

Pluviano, S.; Watt, C. & Della Sala, S. (2017). "Misinformation lingers in memory: Failure of three pro-vaccination strategies". *PLoS One*. 12(7): e0181640. https://doi.org/10.1371/journal.pone.0181640.

Simion, M. (2024a). "Knowledge and disinformation". *Episteme*. 21(4): 1208-1219. https://doi.org/10.1017/epi.2023.25.

----------. (2024b). *Resistance to Evidence*. Cambridge University Press. https://doi.org/10.1017/9781009298537.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199580828.001.0001.

Williamson, T. (2007). *The Philosophy of Philosophy*. Wiley-Blackwell. https://doi.org/10.1002/9780470696675.