The Socio-Technical Battlespace: Deception, Complexity, and Self-Sabotage in Modern Information Warfare

Peter Bednar[1,2,*]

[1] *Department of Informatics, Lund University, Sweden*
[2] *School of Computing, University of Portsmouth, United Kingdom*

Abstract

Modern Information Warfare (IW) has evolved far beyond a simple engineering or cybersecurity problem. It is now an adaptive, socio-technical conflict that directly targets human cognition and institutional trust. In this paper, we evaluate these systemic vulnerabilities by looking through the lens of the data-to-knowledge pipeline. We draw on the foundational principles of information taxonomy, systems theory, and geopolitical defense architectures to interrogate current UK and NATO defense postures. Specifically, we examine how the strategic intent of the National Security Act and the Online Safety Act manifests in practice. Through this analysis, we demonstrate how adversaries leverage targeted advertising technology (AdTech) and artificial intelligence (AI) to orchestrate coordinated inauthentic behavior at scale. However, external threats are only half the problem. We argue that these vulnerabilities are severely compounded from within by the uncritical, hype-driven deployment of Generative AI (GenAI) and Large Language Models (LLMs). When organizations blindly pursue automated productivity, they mistake structural efficiency for strategic effectiveness. The result is a form of cognitive self-sabotage. To counter this decay, we propose a pragmatic Superinformation Governance Architecture. This framework repositions language models as restricted processing tools subordinated to grounded data and human-in-the-loop validation, ultimately safeguarding decision-making structures from both external subversion and internal systemic decay.

Keywords

Socio-technical systems, Information warfare, Information Security, Generative AI governance, Superinformation architecture

1. Introduction

The nature of conflict is undergoing a fundamental shift. State apparatuses can no longer relegate the information battlespace to a secondary front. As physical defenses merge deeply with digital networks, Information Warfare (IW) has become a primary, non-kinetic force designed to destabilize how organizations make sense of reality. Hostile actors routinely manipulate digital environments, for example through commercial advertising technologies (AdTech) and algorithmic micro-targeting. Their goal is not just to spread propaganda, but to

STPIS2026: Proceedings of the 12th International Conference on Socio-Technical Perspectives in IS, September 16-17th L'Aquila, Italy
* Corresponding author.
peter.bednar@ics.lu.se (Peter. Bednar), peter.bednar@port.ac.uk (Peter. Bednar)
0000-0002-3631-2626 (Peter. Bednar)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-4334/paper15.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-19

achieve strategic surprise, drain resources, and trick target institutions into wasting cognitive and financial capital on phantom threats. Pervasive digital connectivity forces a transition toward mass operational participation [80]. Modern maneuvers face pervasive informational and human dimensions, meaning non-kinetic capabilities must integrate as coequal operational forces rather than supporting enablers [77]. Over the past decade, these digital maneuvers have moved from the fringes to become core components of modern state conflict [25].

Despite this fluid reality, contemporary military and policy frameworks remain trapped in a narrow, techno-centric mindset. Traditional defence paradigms consistently isolate information security as a pure engineering or cryptographic problem, operating under the false assumption that stronger firewalls and automated tracking algorithms can guarantee absolute system integrity. Historical assessments show this structural gap, where technology-driven systems routinely outpace operational doctrine and organizational design [76]. Rigid engineering metrics fail to evaluate complex human environments, and system platformisation forces a shift toward critical, societal-scale analysis [47]. Modern digital environments demand updated frameworks because traditional, isolated subsystem boundaries have completely dissolved [59]. Mechanistic reductionism fundamentally undermines how systems evaluate human organizational reality [74], and reducing information operations to raw software tools breeds complacency when modern deception demands a transdisciplinary, whole-of-government response [78]. This traditional perspective ignores the reality that any information system is an emergent property of complex socio-technical interactions. Technical architectures can never be isolated from human cognition, everyday work practices, and organizational culture [11, 12, 33, 35].

This paper presents a critical, multi-perspective inquiry into what happens to information systems when they are exposed to weaponised deception. Security analysis must integrate the core socio-technical space, acknowledging that technical networks remain inseparable from organizational and human structures [54]. Critically, this vulnerability is not confined to national apparatuses or state-level conflict; it directly threatens any and all forms of professional and organizational excellence by systematically eroding the capacity of human practitioners to make informed decisions.

Drawing on systemic strategic thinking and socio-technical frameworks [12, 13], we look past conventional, binary definitions of cybersecurity. Instead, we examine how information spaces are contextually undermined from within. By deconstructing the intent-harm matrix, we illustrate why security cannot rely on machine data processing alone, but must instead address the messy contextual dependencies that govern human interpretation [11].

We highlight the sharp mismatch that occurs when rigid, mechanical solutions are misapplied to dynamic, complex socio-cultural systems [9, 12]. Strategic threats operate as complex systems of changing problems, requiring defenders to manage these organizational "messes" using holistic, non-linear strategies [65]. Today, this vulnerability is heavily visible in Western democratic defence architectures. To show this, we evaluate current geopolitical postures within the United Kingdom and NATO, demonstrating how automated deception exploits technical blind spots to threaten institutional trust. Finally, by untangling the distinct layers separating raw data from internalized knowledge, we expose the risk of cognitive self-sabotage built into the current Generative AI hype cycle. We then outline a pragmatic, human-centred governance framework that subordinates technical speed to authentic semantic comprehension. To explore this, the paper first establishes the intent-harm matrix of modern

deception, evaluates how these vulnerabilities manifest within current UK and NATO defense architectures, and finally deconstructs the data-to-knowledge pipeline to propose a human-centered governance framework.

2. The Intent-Harm Matrix: Problematizing Context and the Illusion of Structural Neutrality

To understand how modern information systems are compromised, we must move past simple definitions of data correctness. We need a taxonomy rooted in human intent and contextual integrity. As established in foundational literature on organizational conflict, information warfare must not be viewed as a narrow technical subversion, but rather as a direct, structural attack explicitly engineered to degrade the robustness, quality, and timeliness of institutional decision-making loops [87]. In hybrid conflict, adversaries exploit the loose connections between objective data and subjective human perception, interpretation and sensemaking. This subversion is typically categorized into misinformation, disinformation, and malinformation, collectively known as the MDM triad, based on intent and truthfulness [28, 29, 39]. This triad forms a continuum based on two main variables: how false the information is, and whether there is an explicit intent to cause harm.

2.1. Misinformation

From a socio-technical perspective, misinformation is a structural failure where inaccurate data is shared without malicious intent. This happens when local operational blind spots or isolated departments cause someone to make an understandable mistake [29, 33]. For example, if a military operator unknowingly forwards an outdated intelligence report, it is not just an individual human error. Instead, it exposes a deeper system flaw: the technical architecture failed to flag that the data's origin or status had changed [12, 14].

2.2. Disinformation

Disinformation sits at the high-harm end of the matrix. It is an intentional, system-wide attack rather than a simple data entry mistake. It relies on deliberately creating deceptive material to exploit human psychological vulnerabilities and the neutral way communication networks handle traffic. Within an organization, disinformation does more than just feed false data (or metadata) into the system; it actively destroys the trust that holds human teams together. Left unchecked, it polarizes organizations, corrupts how staff interpret events, and disrupts critical decision-making cycles [24, 29]. Fundamentally this is undermining the baseline of informed judgment required for professional and organizational excellence.

2.3. Malinformation

Malinformation changes the threat by weaponizing real, authentic facts rather than fabricating false ones. This occurs when completely true data is deliberately stripped of its context and released to cause harm. In socio-technical terms, malinformation does not corrupt system inputs. Instead, it uses genuine data, such as leaking classified operational

details, to cause cognitive overload and organizational panic. This tricks institutions into wasting time and money reacting to a threat that is real but contextually distorted.

2.4. The MDM Triad

Mapping these threats within an intent-harm matrix requires us to identify the specific methods used to alter content and exploit human cognitive vulnerabilities. Fabricated material involves entirely false data built from inception to deceive, whereas manipulated media alters genuine imagery or telemetry to distort evidence while retaining a veneer of authenticity. Imposter content directly hijacks institutional trust metrics by spoofing branding or security credentials to validate false narratives. Meanwhile, misleading messaging involves the structural manipulation of real data, often achieved by selectively deploying statistics or framing a subjective opinion as an objective fact. This distortion is frequently compounded by a false context, where legitimate information is weaponised by stripping away its original timeline or geography. Similarly, false connections use headlines and captions that contradict the underlying substance to manipulate system traffic. Finally, satire and parody, though containing no intrinsic intent to cause harm, are frequently weaponised when automated filtering systems strip away their creative framework.

2.5. Sociotechnical Aspects

Standard computer science models fail to stop these tactics because they treat information as a simple technical problem of data transmission rather than contextual truth [14]. Rigid, traditional risk analysis fails in complex environments; security instead requires tools that can evaluate human and organizational settings [43]. A socio-technical approach reveals how narrow technical designs create systemic vulnerabilities [9, 12].

Adversaries heavily exploit the bias of omission, intentionally leaving out critical history to make an information environment deeply misleading [11]. Because static system designs enforce rigid worldviews, managing complex information spaces requires us to challenge these simplistic, binary ways of thinking [69]. This challenge is severely compounded by what sociotechnical research identifies as path-dependency and architectural lock-in; once an automated, context-blind pipeline is embedded within an enterprise infrastructure, the organization becomes structurally resistant to change, transforming a temporary design flaw into a permanent strategic vulnerability [88].

This blind spot is made worse by distorted framing [11]. When modern defensive dashboards try to capture messy geopolitical threats, they force complex human realities into oversimplified operational metrics [43, 74]. Forcing complex anomalies into rigid, binary choices, like labeling a nuanced narrative simply as "*true/false*" or "*secure/compromised*", leaves defenders open to tactical manipulation [11, 69]. This vulnerability is mirrored in private-sector operations, where uncritical trust in machine data over commercial experience systematically obscures underlying process errors [89]. This dynamic is also visible across broader public sector, where empirical studies shows that forcing complex human environments into rigid, pre-programmed entry fields strips away the interpretive flexibility needed to capture qualitative context, forcing practitioners

into a dangerous illusion of technological infallibility [88]. The system design prioritizes easy automated processing over actual human situational awareness [47, 56].

This failure stems from a fundamental misunderstanding: the reductionist assumption that the ethics, validity, and context of a communication network can be treated as static, pre-programmed technical features. The systemic integrity and moral reality of an information system cannot be isolated inside technical components; they emerge dynamically from the design of the entire human-and-machine loop (the continuous human process of questioning, interpreting, and validating data) [82].

Ultimately, these factors lead to the paradox of neutrality bias. When a technical defense network tries to maintain strict algorithmic neutrality, it fails to account for human cognitive biases and asymmetric threats [11]. By treating all incoming data equally, the system commits what Bateson [5, 6] calls a catastrophic contextual error: it separates the message from its context, making verification impossible. As a result, the system accidentally hosts, validates, and speeds up hostile campaigns, turning a passive platform into an active asset for information warfare. In other words this pursuit of neutrality removes the very filters needed to resist manipulation. In management cybernetics, any organization trying to survive must have active ways to manage and filter complexity [68]. Algorithmic neutrality violates this rule by flattening the information landscape and blinding the system to the threat's qualitative context. This triggers the exact error Bateson [5, 6] identified: separating a message from its ecological framework, leaving it completely stripped of relevance and judgment.

3. The Epistemological Mismatch: Complicated Paradigms vs. Complex Socio-Cultural Reality

This cybernetic error, separating a message from its ecological framework, is directly visible in how modern military infrastructure approaches information warfare. Military and organizational infrastructure often treats information warfare (IW) incorrectly by viewing conflict as a complicated engineering problem rather than a complex socio-technical phenomenon. It becomes a systemic mismatch which erodes defensive readiness because it misjudges the nature of the environments involved. Complicated architectures are fundamentally mechanistic. Designing a ballistic missile or configuring an encrypted enterprise network relies on linear cause-and-effect relationships where outcomes are predictable, repeatable, and controllable. However, mainstream software and AI design consistently overlooks social environments and contextual relevance. Massive-scale engineering failures trace back to social context anomalies, and focusing purely on code layer interactions degrades system performance [55, 56]. Experts can easily disassemble mechanistic problems by applying deterministic rules because the solutions are permanent.

Conversely, Strategic Systemic Thinking (SST) demonstrates that information environments are dynamically organic, defined by fluid, non-linear interactions where outcomes cannot be fully modeled beforehand [9, 13]. Risk and crisis management require socio-technical approaches because mechanistic control structures fail during complex organizational disruptions [42]. Drawing on second-order cybernetics and open systems theory [5, 67], we recognize that IW spaces are human activity systems. They defy absolute,

centralized control because they are constantly reshaped by the contextually dependent interpretations of individual human actors [8, 9].

Rules-based, engineered workflows cannot anticipate the emergent properties of human-centric conflict. Disinformation targets human cognition, and because adversarial behaviour adapts constantly, software patches cannot resolve the threat. Recent strategic analysis confirms that while LLMs optimize the 'science of control,' they are structurally incapable of abductive logic—the creative, non-linear reasoning needed to navigate entirely novel threats [83]. Applying Ashby's [3] cybernetic principle of Requisite Variety, a defensive architecture can only successfully contain a threat if its internal capacity for variation matches the complexity of the environment it seeks to control. Organisational viability demands balanced cybernetic control structures, meaning systems must filter environmental complexity to survive ongoing threats [68]. In the digital age, internet channels have supercharged traditional deception, shifting information operations from isolated psychological campaigns into continuous, automated subversion [12]. This operational vulnerability is directly evidenced by recent wargaming experiments at the US Army War College, where the integration of generative AI into planning environments consistently triggered severe cognitive anchoring and automation bias, leading planners to unthinkingly accept structurally flawed algorithmic models [83].

Hostile (state and business) actors exploit these structural vulnerabilities through two primary, interconnected vectors.

First, they use algorithmic social subversion by leveraging automated bot networks and coordinated sock-puppetry to systematically manipulate the perceived popularity of specific narratives [28]. By actively sharing technical infrastructure to automate and amplify content, hostile networks overwhelm targeted defensive filters with sheer informational volume [79]. This manufactured artificial consensus exploits human cognitive shortcuts and confirmation biases, flooding the digital ecosystem to distort both public and military perception at an unprecedented scale [29].

Second, hostile actors deploy psychometric micro-targeting by hijacking commercial advertising technology (AdTech) and big data analytics to map individual psychological vulnerabilities [28]. Moving past generic, broadcast propaganda, they deliver tailored disinformation directly to specific subsets of a population or military command structure cognitively primed to accept it [29]. This weaponization of localized context transforms open, customer-centric communication networks into precision weapon delivery platforms that operate entirely below the defensive threshold of standard perimeter firewalls.

While this fluid environment creates severe vulnerabilities, it also provides defenders with opportunities for active, non-linear security strategies. Rather than relying purely on passive perimeter firewalls, modern information security can incorporate active cyber deception. Defenders introduce controlled complexity by deploying decoys and honeypots where synthetic data streams confuse attackers. AI-driven forecasting automates scenario generation lifecycle tasks, allowing tactical threat intelligence engines to deploy synthetic content dynamically [60]. Traditional frameworks rely on passive, reactive containment strategies, which represents a restrictive Safety-I security approach. Conversely, modern resilience requires an active Safety-II mindset where systems must dynamically adapt to survive ongoing operational threats [53]. This approach lures hostile actors away to

isolated environments to safeguard real infrastructure, allowing defenders to monitor cognitive intent in real-time.

Crucially, active cyber deception must not be misconstrued as a contradiction of our critique against absolute technical neutrality. While strict, passive neutrality leaves networks exposed, active deception introduces deliberate, non-linear system imbalances to misdirect an attacker's predictive capabilities. However, within a socio-technical framework, this strategy introduces a severe risk of backfiring. If active deception is deployed without rigorous, cross-layer coordination, the unmanaged proliferation of internal synthetic noise can inadvertently pollute the defender's own information environment. This risk demands strict architectural isolation: tactical synthetic telemetry must be tightly contained at the machine infrastructure layer so that it never leaks upward into the human data-to-knowledge pipeline, precipitating a catastrophic slide toward cognitive self-sabotage.

4. Strategic Geopolitical Realities: Contextual Disintermediation in the UK and NATO Defense Posture

The practical reality of system complexity and the MDM triad is clearly visible in modern geopolitical conflicts. The United Kingdom now explicitly classifies Information Warfare (IW) as a top-tier, systemic threat to national security. Instead of an isolated problem, IW is treated as a continuous form of hybrid warfare used by both state and non-state actors. On this digital battlefield, adversaries embed information operations directly into their national strategies to undermine democratic institutions and polarize civil society [29]. This threat has evolved far beyond basic propaganda; it now heavily exploits Artificial Intelligence, automated advertising networks (AdTech), and large-scale, coordinated deceptive behaviours to manipulate how organizations and societies perceive reality [29].

To protect its socio-technical infrastructure, the UK uses a dual-layered defence posture combining national legislation with specialized operational capabilities. Analyzing these legal instruments allows us to map exactly how a state tries to hardcode trust and regulatory filters into a rapidly shifting digital landscape. Because system resilience requires flexible, pluralistic frameworks, state regulation must adapt to protect media diversity during active information campaigns [81]. On the legislative front, *the UK Online Safety Act 2023* [30, 37] legally forces major digital platforms to identify and stop state-sponsored disinformation under the Foreign Interference Offence. To standardize internal resilience across departments, the *Government Functional Standard GovS 007* establishes management metrics that link technical cyber security directly with personnel and physical asset protection [41]. This provides a stable baseline for institutional risk management.

This legal framework was heavily updated by the *Data (Use and Access) Act 2025* [21] and subsequent 2026 regulations [19, 20, 21]. Under this modernized legal landscape, including *the Crime and Policing Act* 2026 [17], the state criminalized the creation of non-consensual AI deepfakes. This designated malicious synthetic media as a statutory "*Priority Offence*" that digital platforms must proactively scan and block. Alongside this, *the National Security Act 2023* [36] introduced a Foreign Influence Registration Scheme to disrupt state-backed sabotage, espionage, and covert psychological operations. Because dangerous gaps

still exist regarding hybrid proxies operating just below the threshold of open conflict, the legislature fast-tracked the *National Security (State Threats) Bill* [26]. This allows the state to ban subversive networks and criminalize proxy actors. Simultaneously, *the Cyber Security and Resilience Bill 2026* [20] enforces real-time breach reporting metrics on critical infrastructure operators facing systemic, AI-driven compromises.

Operationally, the Department for Science, Innovation and Technology (DSIT) directs the National Security and Online Information Team (NSOIT). To match the shifting technological landscape, NSOIT's open-source monitoring now specifically targets the spread of generative AI and deepfakes designed to undermine elections and public safety [22]. Rather than acting as a passive technical filter, this unit functions as a socio-technical sensor. It analyzes how automated, synthetic content interacts with human psychological biases to alter public cognition during active campaigns.

This hostile behaviour directly impacts wider international defense. Allied nations recognize that adversaries routinely use hybrid tactics to test state resilience, exploit rapid digitalisation, and target domestic institutional trust [29]. To fix fragmented domestic capabilities, modern military doctrines advocate for an interoperable approach designed to identify, prevent, contain, mitigate, and recover from cross-domain informational attacks [29]. As established in socio-organizational change theory, technology can never be separated from the cognitive, national, and cultural environments where it is deployed [4]. For defenders, this means data-driven capabilities must be combined with clear diplomatic statements, public attributions, and media literacy campaigns to raise the cost for hostile actors [29]. A key tool here is NATO's Information Environment Assessment (IEA) infrastructure, which uses open-source intelligence (OSINT) to audit narratives and map adversarial movements in real time.

Synthesizing these strategic positions reveals the ultimate goal of modern information warfare: the structural disintermediation of the human actor from reality. When evaluated through Gregory Bateson's [5] cybernetic lens, these public policies expose a fundamental systemic tension. While the UK and NATO intend to secure the ecosystem, their frameworks often commit an ecological error by treating fluid, organic human networks as deterministic, linear technical circuits. The unrestricted use of algorithmic AdTech and generative AI does not just spread isolated lies; it systematically cuts away the intermediate layers of institutional trust, objective journalism, and reliable intelligence. This disintermediation creates a broken operational environment where cognitive systems are continuously fed simulated inputs, separating individuals from an objective baseline of reality. Consequently, decision-making environments collapse, leaving military commands and democratic societies unable to establish a reliable Common Operational Picture, rendering them highly vulnerable to systemic paralysis and defeat.

5. Architectural Deconstruction: The Data-to-Knowledge Pipeline in Cyberspace

To build an information security framework capable of resisting disintermediation, defensive strategies must establish clear architectural boundaries between the technical and cognitive spheres of conflict. This requires isolating the operational differences between cyberwarfare

and information warfare. Cyberwarfare is strictly technical and structural. Cyberwarfare disrupts physical computer systems. It degrades software architectures and database storage. Malware and hacking damage network infrastructure. Its target is the machine. Information Warfare, by contrast, is cognitive and behavioral. It utilizes ICT infrastructure to manipulate or weaponize the information environment itself to alter human cognition, undermine institutional trust, and disrupt an opponent's decision-making loops [2, 29].

Understanding how deception penetrates these layers requires adopting rigorous breakdown of the information pipeline. In socio-technical systems analysis, data is not an objective, standalone monolith. Instead, it evolves through four distinct, vulnerable phases: Data, Capta, Information, and Knowledge. Data consists of raw symbols, text, and uninterpreted content. Knowledge acquisition must adapt to continuously changing realities, meaning true competence requires observers to actively bound system observations [71]. Capta represents selected data that observers or algorithms bound, which sensor networks filter based on machine perspectives [11, 18, 23]. Capta is simply the small subset of raw data that a sensor or observer actively selects to look at.

Traditional computer science paradigms consistently fail to mitigate these delivery mechanisms because they analyze information through a technical lens that prioritizes binary data transmission over contextual truth (or relevance). Applying socio-technical theory reveals how these narrow technical design choices introduce systemic vulnerabilities. This failure stems from a fundamental epistemological error: the assumption that dynamic human qualities, such as trust, systemic validity, and contextual integrity, can be pre-programmed as static, deterministic properties within a software architecture. Because these qualities are not mechanical components but are entirely emergent properties of human activity systems [5], they cannot be hard-coded into database structures or algorithmic filters. Static system architectures that attempt to automate these interpretive dimensions enforce rigid underlying worldviews, coercing subjective interpretation into rigid, bivalent cognitive frameworks. This type of algorithmic compression completely challenges traditional doctrinal frameworks where professionals build deep understanding through continuous dialogue and refinement; instead, humans are forced to choose from options that have already been completely pre-filtered and pre-shaped by the machine [84]. When defenders assume these human dynamics can be outsourced to automated filters, they actively blindfold their own inquiring systems, initiating the exact process of organizational self-sabotage explored below.

Because the human mind operates as an autopoietic, self-referential system, cognitive structures are active creators of meaning rather than passive processors of external electronic telemetry [27]. Authentic command loops must therefore function as purposeful, open systems possessing the independent capacity to determine strategic ends. This requirement directly mirrors the foundational systems principle that an informational infrastructure cannot be divorced from the moral and purposeful framework of its human users; systemic validity resides entirely within the active, human-led inquiring system [82]. True knowledge, consequently, is the internalized understanding of a potential for action, representing a human cognitive capacity to execute decisions within a unique, complex organizational setting [64]. Machines process syntax; they cannot manage authentic semantic knowledge.

5.1. Superinformation

To help human operators achieve genuine meaning, technical systems must be wrapped in what Bednar, Welch, and Graziano [14] define as "*superinformation*." Superinformation is not a faster database or a larger data pipeline. Instead, it is a dynamic layer of context that explicitly tracks where information came from, how it has changed, and the perspective of the user [14]. It functions as a diagnostic shield against Generative AI architectures, which inherently operate in a context vacuum. While Large Language Models (LLMs) output slick, ungrounded text that strips away the source material's lineage, a Superinformation Architecture forces the system to display its background assumptions, systemic boundaries, and operational limitations. By giving users the tools to interrogate the machine, this layer prevents humans from blindly accepting or rejecting automated data due to automation bias.

According to this definition [14], superinformation relies on three core principles:

- **Anti-Reductionism:** While standard metadata is purely mechanical (like file sizes or timestamps), superinformation focuses on the relationships and contextual dependencies between technical and organizational fields.
- **Symmetry of Inquiry:** It ensures the human user can question and critique the machine's generation process, rather than being forced into a passive role.
- **A Bridge to Knowledge:** Machine outputs cannot become a true "potential for action" without passing through this human-curated context, which turns raw text back into a tool for genuine understanding.

To see this architectural distinction in practice, we can contrast standard metadata with active superinformation across three critical information warfare scenarios. An effective infrastructure must reject artificial automation shortcuts in favor of systemic 'complexification' [14]. Rather than using algorithms to hide backend messiness, a Superinformation Architecture explicitly reveals environmental complexity to empower the user.

In the first scenario*, preserving complexity stands in direct opposition to artificial simplification.* When an LLM summarizes an adversarial theater report, a standard interface hides contradictions to provide a clean reading experience. Conversely, a Superinformation layer forces the system to preserve and display conflicting raw data streams, map divergent timelines, and reveal background ambiguities. This ensures the user is confronted with the true, messy reality of the data rather than an artificially smoothed algorithmic illusion.

The second scenario *focuses on enabling a symmetrical control of inquiry*. In a standard automated threat-detection dashboard, the workflow forces the operator into a passive role of either blindly accepting or rejecting a machine-generated alert. Superinformation infrastructure breaks this dynamic by providing explicit contextual hooks. These allow the human to interrogate the algorithm’s underlying assumptions, challenge its training limitations, and query alternative hypotheses, matching the variety of the human mind against the machine.

The third scenario *exposes the contextual field of view.* When a strategic unit uses natural language processing to evaluate open-source psychological campaigns, standard

technical formats isolate the text from its ecosystem. Superinformation allows the human to map the entire structural field surrounding the output by displaying the source's operational history, tracking content drift, and exposing the system's own constraints. This keeps the human actively aware of the system's operational boundaries and limitations, preventing single-loop automation complacency.

Without this active supporting layer of superinformation, human operators lack the diagnostic environment required to make responsible and informed judgments. System architectures must deploy active, user-led evaluation environments, using reflective prototyping methods to protect independent human critical validation [73]. Genuine knowledge is the internalized understanding of a potential for action, representing the cognitive capacity to execute decisions within a unique, complex organizational setting [12]. Machines process text, but they cannot manage authentic knowledge [50].

This distinction is critical for information security and defence. A cybersecurity network may successfully protect the technical integrity of the machine layer, yet still fail completely if the human operator misinterprets the manipulated context, leading to corrupted cognitive execution at the human layer [11, 44].

5.2. Cognitive Defense

Superinformation rejects the artificial simplification of modern natural language interfaces, which strip away context to deliver a clean consumer experience. Instead, it embraces systemic *'complexification'*. This architecture recognizes that information warfare operates by separating a message from its ecological framework. Therefore, a Superinformation Architecture is designed not merely to display content, but to empower human users to understand a deeply developed description of its context, creation, and origin.

We can see this definition operationalized across three critical dimensions of modern cognitive defense.

1. **Context:** First, it establishes the context of message creation by reconstructing the ecological framework in which the message was born. While standard language models present text in a historical vacuum, a superinformation layer exposes the underlying political, cultural, or institutional pressures surrounding the source material. It explicitly maps out what was altered, omitted, or framed during its initial generation. By rendering these background assumptions visible, it prevents the user from committing a catastrophic contextual error.
2. **Origin**: Second, it maps the trajectory of origin to reveal provenance and systemic lineage. Deception thrives when the true lineage of data is masked. Superinformation acts as an immutable structural trace, exposing the complete history of a narrative's journey across digital spaces. It reveals to the human analyst who orchestrated the data pipeline, how the text mutated over time, and where adversarial algorithms or commercial AdTech platforms accelerated its velocity. Crucially, the construction of this superinformation layer cannot be trusted to Generative AI or automated language models. Because GenAI architectures inherently operate as epistemic black boxes, generating text through statistical averaging that actively erases data lineages,

they are structurally incapable of verifying provenance. Entrusting an LLM to generate superinformation merely introduces a secondary layer of ungrounded hallucinations, compounding the risk of systemic self-deception. This limitation is heavily emphasized by contemporary information systems governance literature, which notes that Generative AI severely intensifies organizational risks because it produces massive volumes of derivative text while completely obscuring data provenance [85]. Instead, this diagnostic shield must rely on deterministic code layers, immutable forensic tracking, and human-in-the-loop verification to ensure the integrity of the origin trace. This structural independence is the only mechanism that allows the user to judge the validity of a claim based on its authentic systemic origin rather than an artificially generated algorithmic illusion.

3. **Control:** Third, it enables the symmetrical control of inquiry, allowing the human to actively interrogate the machine. True support occurs when an information system expands, rather than restricts, a user's capability to investigate. Superinformation equips the human actor with the tools required for a symmetrical control of inquiry.

The human is no longer a passive recipient of a machine-generated output; instead, they are given explicit structural levers to challenge the machine's boundaries, interrogate its processing constraints, and project alternative contextual interpretations. Without this active layer of superinformation, human operators lack the diagnostic environment required to execute double-loop learning, leaving them vulnerable to single-loop automation complacency [17] and external adversarial subversion.

5.3. Generative AI Aspects

While deception remains an ancient component of warfare, modern digital networks have fundamentally altered how it operates across our information hierarchy. Generative AI allows adversaries to manufacture synthetic data and false contexts at an unprecedented scale. By manipulating metadata, AI masks the true origin of a message, tricking automated defensive filters into treating hostile capta as legitimate telemetry. System escalation risks stem directly from contested network pipelines, meaning AI reliability hinges entirely on the vulnerability of its underlying data architecture [75].

The most dangerous frontier of modern information warfare occurs when AI induces cognitive self-deception within defence commands. AI systems process massive data volumes using simplified user interfaces designed to induce human automation bias. This causes operators to trust algorithmic outputs unthinkingly while ignoring conflicting environmental telemetry [17, 31]. Recent studies validate this dynamic, demonstrating that, the speed and clean formatting of the interface routinely displace independent human validation, misleading decisionmakers to internalise corrupted data as true understanding [83]. If an adversary subtly poisons the underlying data stream, the AI will generate flawed models. The human operator, believing the system to be objective and neutral, internalizes this as true understanding. Ultimately, they deceive themselves, acting on a compromised potential for action that directly benefits the attacker. This danger was illustrated by a critical 2026 operational near-miss, where the U.S. military nearly initiated a direct conflict with China after field commands relied on an error-prone, hallucinated large language

model intelligence report [86]. The fact that an automated, simplified interface almost disarmed human critical judgment and triggered a global crisis proves that organizational self-sabotage is an immediate operational reality.

6. The Sovereign Threat of Generative AI: Cognitive Self-Sabotage and the Socio-Technical Governance Framework

6.1. The Proliferation of Automated Competence

The uncritical, market-driven deployment of Generative AI (GenAI) and Large Language Models (LLMs) severely worsens the vulnerabilities in human-machine interaction. Both commercial and defense organizations are rushing to integrate GenAI workflows, frequently ignoring deep structural flaws and unproven returns on investment. This rush has created an unstable security environment flooded with "AI washing", the marketing of fake competence to hide underlying software limitations. This is evident in commercial B2B environments, where over-reliance on the perceived objectivity of algorithmic models causes personnel to suppress their own professional intuition and critical judgment [89]. Beyond simple corporate risk, this unchecked rollout functions as an unregulated global experiment. It supercharges synthetic deepfakes and algorithmic tools that trigger structural power imbalances, replacing genuine institutional trust with superficial automated compliance [61, 62].

This substitution creates a system that displays superficial success metrics but remains fundamentally incapable of learning or adapting to changing operational conditions [68]. When organizations try to insulate themselves by embedding automated proxies for human critical judgment directly into software design, they trigger a dangerous feedback loop of complacency. Shifting evaluation from the human actor to an algorithmic protocol removes the need for independent oversight. Human operators, suffering from lack of resources and automation bias, unthinkingly defer to the machine's perceived neutrality, mistaking a mechanical checklist for genuine operational validity. This dynamic effectively warps the foundation of organizational trust, shifting it away from human collaboration and instead misplacing it onto the perceived reliability of an automated platform [84]. Professionals under time and resource constraints are therefore under cognitive pressure to systematically defer to automated recommendations even when those algorithmic outputs are profoundly incorrect [84]. As the Association for Information Systems (AIS) task force warns, relying on rigid, pre-programmed protocols traps organizations in a dangerous *'compliance trap*.' This creates a false sense of order and moral comfort that actively blinds professionals to deeper contextual threats [85]. Empirical analysis of automated triage systems in public healthcare demonstrates that rigid algorithmic amplification forces professionals into the very same *'compliance trap*,' where mechanical box-ticking entirely displaces professional judgment and causes severe healthcare issues [88]. Similar systemic failure is also visible in contemporary geopolitical diplomacy, where major global powers have found it impossible to establish enforceable treaties around military AI precisely because these algorithmic architectures lack any viable verification or compliance mechanisms [86]. This complacency has directly paralyzed global security governance; in late 2026, leading military powers actively stripped provisions from draft UN rules that required autonomous systems to operate *'predictably'* and *'reliably*,' explicitly removing the requirement for mandatory human verification of AI-identified targets before striking [86]. True resilience

cannot be hard-coded into software; it requires an active, human-led inquiry supported by a governance framework that explicitly protects the operator's capacity for deep reflection.

6.2. From Commercial Risk to Societal Isolation

This disruption is not isolated to corporate software. The prevalence of unverified, AI-washed technologies compromises defence supply chains by undercutting the validation metrics needed to accurately score vendor capabilities. Beyond these institutional risks, the unchecked rollout of these models functions as an unregulated global experiment. It supercharges synthetic deepfakes, targeted blackmail, and algorithmic companion applications that reinforce toxic behavioral patterns and societal isolation, ultimately leading to corrupted cognitive execution at the human layer [11, 44].

6.3. The Conflation of Efficiency and Effectiveness

Integrating this digital expansion with systems theory exposes a severe strategic vulnerability: the systemic confusion of efficiency with effectiveness [12]. Under pressure from data-driven markets, standard agile workflows consistently ignore humanistic values [48, 49]. Effective AI alignment demands careful management of strategy, capability, and work identity, yet faster machine outputs consistently mask broken processes [58]. This uncritical acceleration violates classic minimal critical specification rules, which state that technical designs must remain compatible with human learning and adaptability [63]. When defense structures rush to adopt GenAI to optimize productivity, they confuse the rapid generation of text (efficiency) with the contextual validity and strategic utility of that content (effectiveness). As operational evidence warns, automating these workflows effectively dislocates command from process, creating a critical vulnerability where the rapid velocity of machine output merely masks the systematic atrophy of independent human reflection [83]. This illusion is mirrored in public sector infrastructure research, which demonstrates that when rigid algorithmic pathways fail to align with organic human routines, they generate intense *'techno-stress'* across staff layers; rather than liberating human capacity, the uncritical introduction of these automated systems *increases overall cognitive workloads* and triggers a gradual atrophy of operational quality [88]. This mismatch results in severe internal cognitive self-sabotage. For example, commercial case studies confirm that outsourcing B2B decisions to automated analytical proxies causes a profound loss of strategic flexibility, leaving organizations unable to adapt to non-linear market disruptions [89].

6.4. The Pathologies of Linguistic Shorthand and Methodological Mismatch

Because language is a product of intelligence rather than its source, LLMs lack true semantic comprehension, relying instead on statistical pattern matching across scraped datasets. This substitution of automated shorthand for authentic human reasoning introduces severe pathologies into institutional memory [46, 50, 52]. Ultimately, this degradation threatens any and all kinds of professional excellence; by blinding practitioners to the qualitative context of data, it strips professionals of the exact critical capacity needed to make informed, responsible decisions. In Bateson’s framework of logical typing, this

confuses the manipulation of symbols (syntax) with the generation of meaning (semantics) [6]. By mistaking simulated text for intelligence, organizations degrade their capacity for high-order learning. Uncritical automation restricts commands to superficial, single-loop adjustments, completely eliminating the double-loop reflection required to evaluate underlying strategic assumptions [52].

Furthermore, replacing junior analytical positions with automated linguistic proxies disrupts the educational pipeline required for long-term professional adaptation and critical problem-solving [34]. Over-reliance on intelligent automation causes severe deskilling, where organizations risk losing control of core processes through skill deterioration [57]. Operational analysis warns that this heavy reliance directly erodes the intrinsic decision-making skills of leaders, permanently compromising their ability to manage complex, non-linear environments when automated architectures fail [84]. Exposure to these models degrades analytical confidence, fueling systemic anxieties centered on data security, institutional privacy, and the operational risks introduced by structural hallucinations [1, 38]. This methodological fragility is mirrored across other critical domains; in medical informatics, for instance, contemporary pipelines are heavily critiqued for relying on noisy data and uncalibrated metrics that generate overly optimistic performance data while failing real-world clinical validation [16].

6.5. The Slide into Functional Stupidity

Functional Stupidity is the (often unintended) organizational pressure to suppress critical thinking to maintain a superficial appearance of smooth operations. When military, public sector, or private enterprise commands rely on these ungrounded engines to draft assessments, they introduce hallucinated data into the command loop. Uncritical document summarisation introduces severe text distortions, as systems mistake superficial layouts for validity [10]. Automated text generation creates dangerous cargo-cult processing environments, where slick formatting simulates scientific competence while removing grounded structural verification [72]. Authentic semantic verification requires independent oversight. By sacrificing human insight for algorithmic shortcuts, organizations disarm their decision-makers, decouple structural claims from operational realities, and create scenarios where talk and decisions conflict with meaningful action [70].

Organizational systems default to a dangerous functional stupidity where personnel suppress critical reflection to maintain an artificial, ungrounded simulation of competence [66]. This systemic danger is directly supported by recent Association for Information Systems (AIS) frameworks, which warn that uncritical digital acceleration creates a dangerous 'compliance trap' that drives organizations toward functional stupidity by replacing active professional judgment with rigid, mechanical checklists [85].

This self-deception functions as an internal form of information warfare, blinding institutions to the strategic damage done to their internal knowledge structures and pushing culture away from critical reflection toward an artificial simulation of competence [7, 15, 40].

Countering these threats does not require a total rejection of machine learning, which is both impractical and impossible. Pattern-recognition indices remain highly valuable tools for parsing complex, structured data arrays. Instead, security architectures must actively

dismantle the LLM-centric fallacy by establishing a Superinformation Governance Architecture [14]. Rather than treating LLMs as standalone intelligent agents, they must be structurally restricted to narrow natural language processing components wrapped in an active *superinformation* layer. Within this framework, algorithmic outputs are never permitted to operate in a vacuum. Instead, they are strictly subordinated to an active, systemic layer of dynamic contextual dependencies, immutable provenance tracking, and historical system trajectories. Machine data cannot trigger action automatically; it must pass through a human-curated diagnostic shield. This architecture insulates the command loop, simultaneously thwarting external adversarial subversion and arresting internal algorithmic decay [12].

7. Conclusion: Contextual Inquiry and Information Governance as a Defensive Imperative

Modern Information Warfare is a profound socio-technical challenge that technical engineering alone cannot solve. The convergence of the MDM triad with targeted exploitation and systemic disintermediation from reality turns cyberspace into a highly volatile, non-linear battlespace.

A socio-technical systems perspective exposes a key vulnerability: defense institutions consistently treat organic cognitive threats mechanically [12]. This mistake confuses linguistic efficiency with operational effectiveness and causes accidental self-sabotage. Because Generative AI models lack semantic meaning, they operate completely detached from grounded telemetry. Adopting them uncritically disarms human critical thinking. This leaves command loops dangerously exposed to both external deception and internal algorithmic hallucinations. Ultimately, this vulnerability ripples far beyond the military sphere, directly threatening any and all kinds of professional and organizational excellence. When the integrity of the data-to-knowledge pipeline is compromised, it directly impacts the ability of professionals to make informed decisions, leaving them unable to execute responsible and effective operational judgments.

To mitigate these threats, national and organizational defense postures must move past passive compliance or reactive cyber defenses. Security frameworks must span the socio-technical divide because data integrity is meaningless without cognitive validity. This requires a double-pronged defense strategy.

First, legislative frameworks must connect with decentralized societal resilience. Statutory initiatives, like *the UK Online Safety Act 2023* [37] and *the National Security Act 2023* [36], provide a vital baseline for state deterrence. However, top-down state actions are incomplete without a foundation of decentralized media literacy and institutional trust to serve as active defensive infrastructure [29, 32]. This behavioral resilience directly implements classic management cybernetics: a viable infrastructure must match environmental variety dynamically rather than relying on rigid, hard-coded safety metrics [68]. This shift aligns with the latest frameworks in digitalization governance, which argue that because complex outcomes cannot be predicted beforehand, institutions must move away from fixed rule-sets and instead cultivate a virtue-ethics orientation that prioritizes independent professional judgment and critical reflection [85]. Cultivating media literacy and psychological safety frameworks [51] within defense institutions creates a dynamic behavioral architecture. This armor enables

personnel to use double-loop reflection, empowering them to challenge flawed operational assumptions without interpersonal risk. This shields collective decision-making environments from adversarial manipulation.

Second, robust Information Governance frameworks must be deployed as an active strategic countermeasure. Defense commands can break internal feedback loops of self-deception by restricting large language models to narrow processing tools and wrapping automated architectures in a layer of *superinformation* [14, 24]. This architecture reinforces a foundational premise: the systemic validity and moral integrity of an information ecosystem can never be programmed into isolated software components. Instead, they emerge entirely from the design of the broader human-and-machine inquiring loop [10, 82]. Implementing strict, cross-functional processes to validate data provenance ensures that strategic context is preserved across both technical and organizational fields. Socio-technical frameworks counter covert exploitation effectively because long-term security depends on protecting these integrated environments rather than patching isolated hardware perimeters [45]. This imperative is reinforced by empirical lessons in infrastructure change, which warn that replacing screen-mediated, transactional interfaces for human dialogue systematically strips away the *'relational infrastructure'*, the informal networks of communication, trust, and shared values, that holds teams together and keeps an institutional decision making capability resilient [88].

Ultimately, safeguarding the data-to-knowledge pipeline requires protecting the human element within the machine loop. Only by preserving independent human insight, conceptual critique, and rigorous verification can modern institutions defend their decision-making frameworks from subversion and maintain a clear, uncorrupted grasp on reality.

Declaration on Generative AI

During the preparation of this manuscript, the author utilized Google Gemini (Gemini 1.5 Pro) to improve the readability, narrative structure, and thematic cohesion of the text. The author reviewed, edited, and verified all outputs generated by the AI assistant and assume full accountability for the accuracy, originality, and integrity of the final text and its references.

References

[1] AI Now Institute. (2025). Artificial power: 2025 landscape report. https://ainowinstitute.org

[2] Andress, J., & Winterfeld, S. (2013). *Cyber warfare: Techniques, tactics and tools for security practitioners* (2nd ed.). Syngress.

[3] Ashby, W. R. (1956). *An introduction to cybernetics*. Chapman & Hall.

[4] Avgerou, C. (2001). The significance of context in information systems and organizational change. *Information Systems Journal*, 11(1), 43–63.

[5] Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chandler Publishing.

[6] Bateson, G. (1979). *Mind and nature: A necessary unity*. E. P. Dutton.

[7] Bayer, K. (2025). We were worried about the wrong dystopia. We were all so busy worrying about Big Brother we forgot to worry about Soma. 2025, November 24th. *Medium*. https://medium.com

[8] Bednar, P. (2009). Contextual analysis – a multiperspective inquiry into emergence of complex socio-cultural systems. In G. Minati, M. Abram, & E. Pessa (Eds.*), Processes of*

emergence of systems and systemic properties: Towards a general theory of emergence (pp. 299–312). World Scientific Publishing.

[9] Bednar, P. (2016). *Complex methods of inquiry: Structuring uncertainty.* Lund University Press.

[10] Bednar, P. (2025). GenAI summaries of articles: Effective and useful? *Proceedings of Socio-Technical Perspectives in Information Systems (STPIS 2025)*, Skopje, North Macedonia.

[11] Bednar, P. M., & Welch, C. (2008). Bias, misinformation and the paradox of neutrality. *Informing Science: The International Journal of an Emerging Transdiscipline,* 11, 85–106. https://doi.org/10.28945/441

[12] Bednar, P. M., & Welch, C. (2020). Socio-technical perspectives on smart working: Creating meaningful and sustainable systems. *Information Systems Frontiers*, 22(2), 281–298.

[13] Bednar, P. M. (2000). A contextual integration of individual and organizational learning perspectives as part of IS analysis. *Informing Science: The International Journal of an Emerging Transdiscipline*, 3(3), 145–156. https://doi.org/10.28945/590

[14] Bednar, P. M., Welch, C., & Graziano, A. (2007). Learning objects and their implications on learning: A case of developing the foundation for a new knowledge infrastructure. In K. Harman & A. Koohang (Eds.*), Learning objects: Applications, implications & future directions* (pp. 157–185). Informing Science Press.

[15] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM FAccT Conference,* 610–623.

[16] Cabitza, F., Jurman, G., Molinari, F., & Bellazzi, R. (2026). Why almost all ML models for medicine are wrong—and what we need for evidence-based medical AI. International *Journal of Medical Informatics,* 219, Article 106538. https://doi.org/10.1016/j.ijmedinf.2026.106538

[17] Center for Security and Emerging Technology [CSET]. (2024). *AI safety and automation bias.* Georgetown University.

[18] Checkland, P., & Holwell, S. (1998). *Information, systems and information systems: Making sense of the field.* John Wiley & Sons.

[19] Crime and Policing Act 2026, c. 20. United Kingdom Parliament. https://legislation.gov.uk

[20] Cyber Security and Resilience (Network and Information Systems) Bill (2026). Parliament: House of Commons. Bill 10442. https://commonslibrary.parliament.uk/research-briefings/cbp-10442/

[21] Data (Use and Access) Act 2026, c. 18. United Kingdom Parliament. https://legislation.gov.uk

[22] Department for Science, Innovation and Technology. (2024). National Security Online Information Team (NSOIT): Privacy notice. GOV.UK. https://www.gov.uk

[23] Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1).

[24] Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality.* Oxford University Press.

[25] Goldman, E. O., & Arquilla, J. (2007). Cyber strategies and information insecurity. *Journal of Strategic Studies*, 30(3), 381–403.

[26] House of Lords Library. (2026, June 19). National Security (State Threats) Bill: HL Bill 35 of 2026–27. UK Parliament. https://lordslibrary.parliament.uk/research-briefings/lln-2026-0030/
[27] Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel Publishing
[28] NATO Strategic Communications Centre of Excellence. (2023). Research reports. NATO StratCom COE. https://stratcomcoe.org
[29] North Atlantic Treaty Organization. (2024). NATO's approach to counter information threats: Public summary. NATO Official Texts.
[30] Osborne Clarke. (2026). Digital regulation: UK regulatory outlook February 26th. 2026. Osborne Clarke Insights. https://www.osborneclarke.com/insights/regulatory-outlook-february-2026-digital-regulation
[31] Coco, A. (2023). Exploring the Impact of Automation Bias and Complacency on Individual Criminal Responsibility for War Crimes, *Journal of International Criminal Justice*, Volume 21, Issue 5, November 2023, Pages 1077–1096, https://doi.org/10.1093/jicj/mqad034
[32] Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, 3(1), 2, 1–9. https://doi.org/10.5334/joc.91
[33] Sadok, M., Alter, S., & Bednar, P. (2020). "It is not my job": Exploring the disconnect between corporate security policies and actual security practices in SMEs. *Information & Computer Security*, 28(3), 467–483. https://doi.org/10.1108/ICS-01-2019-0010
[34] Sharwood, S. (2025, August 21). AWS CEO says AI replacing junior staff is 'dumbest idea'. *The Register*. https://www.theregister.com/software/2025/08/21/aws-ceo-says-ai-replacing-junior-staff-is-dumbest-idea/
[35] Trist, E. (1981). *The evolution of socio-technical systems*. Tavistock Institute.
[36] United Kingdom Parliament. (2023a). National Security Act 2023. The Stationery Office.
[37] United Kingdom Parliament. (2023b). Online Safety Act 2023. The Stationery Office.
[38] Vigliarolo, B. (2025, October 8). AI gets more 'meh' as you get to know it better, researchers discover. *The Register*. https://www.theregister.com/software/2025/10/08/ai-gets-more-meh-the-more-you-use-it-researchers-find/
[39] Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. https://rm.coe.int/information-disorder-report-november-2017/1680764666
[40] Widder, D. G., West, S., & Meredith, W. (2023). Open (Source) AI is corporate AI. AI Now Institute. https://ainowinstitute.org
[41] Cabinet Office. (2021). Government Functional Standard GovS 007: Security (Version 2.0). UK Government Central Functions. https://www.gov.uk/government/publications/government-functional-standard-govs-007-security.
[42] Borodzicz, E. P. (2005). *Risk, crisis and security management*. Wiley.
[43] Baskerville, R. (1991). Risk analysis: an interpretive feasibility tool in justifying information systems security. *European Journal of Information Systems*, 1(2), 121-130. https://doi.org/10.1057/ejis.1991.20

[44] Bednar, P., & Katos, V. (2009). Addressing the human factor in information systems security. In *MCIS2009: 4th Mediterranean Conference on Information Systems* (pp. 900-912). Athens University of Economics and Business.
[45] Sadok, M., Welch, C., & Bednar, P. (2019). A socio-technical perspective to counter cyber-enabled industrial espionage. *Security Journal*, 33, 27-42 https://doi.org/10.1057/s41284-018-00198-2
[46] Mumford, E. (1999). *Dangerous decisions: Problem solving in tomorrow's world*. Kluwer Academic/Plenum Publishers.
[47] Wynn E, Vallo Hult H (2019), "Qualitative and Critical Research in Information Systems and Human-Computer Interaction". *Foundations and Trends in Information Systems*, Vol. 3 No. 1-2 pp. 1–233, doi: https://doi.org/10.1561/2900000014
[48] Alter, S. (2019). Applying socio-technical thinking in the competitive, agile, lean, data-driven world of knowledge work and smart, service-oriented, customer-centric value creation ecosystems. *Complex Systems Informatics and Modeling Quarterly*, (18), 1–21. RTU Journals.
[49] Alter, S. (2015). Sociotechnical systems through a work system lens: A possible path for reconciling system conceptualizations, business realities, and humanist values in IS development. In *Proceedings of the 1st International Workshop on Socio-Technical Perspective in IS Development (STPIS 2015*) (Vol. 1374, pp. 32–40). CEUR-WS.org
[50] Nissen, H.-E. (2012). What or whom do knowledge-managers manage? In J. G. Cegarra (Ed.), *Proceedings of the 13th European Conference on Knowledge Management* (pp. 839–847). Academic Conferences and Publishing International Limited.
[51] Edmondson, A. (1999). 'Psychological safety and learning behavior in work teams', Admin. Sci. Q., vol. 44, no. 2, pp. 350–383, Jun. 1999, doi: 10.2307/2666999.
[52] Argyris, C. and Schön, D. A. (1978) *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.
[53] Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Farnham, UK: Ashgate Publishing, 2014.
[54] Sarker, S., Chatterjee, S., Xiao, X., Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance, *MIS Quarterly* 43 (2019) 695–720.
[55] Sommerville, I. (2015). Designing software for a million users: It's not about the technology (Keynote). In *Proceedings of the 1st International Workshop on Socio-Technical Perspective in IS Development (STPIS 2015)* (CEUR-WS.org, Vol. 1374, pp. 1–4)
[56] Baxter, G., Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering, *Interacting with Computers*, Volume 23, Issue 1, January 2011, Pages 4–17, https://doi.org/10.1016/j.intcom.2010.07.003
[57] Rinta-Kahila, T., Penttinen, E., & Lyytinen, K. (2021). Organizational transformation with intelligent automation: Case Nokia Software. *Journal of Information Technology* Teaching Cases, 11(2), 101-109. https://doi.org/10.1177/2043886920954874
[58] Baiyere, A., Constantiou, I., Indulska, M., & Tuunainen, V. K. (2026). Beyond adoption : aligning strategy, capability, identity, and work in digital transformation. *Journal of Strategic Information Systems*, 35(3), 1-3. Article 101974. Advance online publication. https://doi.org/10.1016/j.jsis.2026.101974

[59] Fischer, L. H. & Baskerville, R. (2020). Revising the Socio-Technical Perspective for the 21st Century: New Mechanisms at Work. *Journal of Database Management* (JDM), 31(4), 69-87. https://doi.org/10.4018/JDM.2020100104
[60] Zacharis, A., Katos, V., Patsakis, C., 2024. Integrating AI-driven threat intelligence and forecasting in the cyber security exercise content generation lifecycle. *International Journal of Information Security*, 23 (4), 2691-2710.
[61] Monod, E. (2025). Robots as Managers? Organizational Justice, Power Asymmetries, and the Future of AI Supervision. *Management Research Quarterly*, 2(1), 41-49. https://doi.org/10.63029/rqg5z883
[62] Monod, E., Lissillour, R., Köster, A. & Jiayin, Q. (2023) Does ai control or support? power shifts after ai system implementation in customer relationship management. *Journal of Decision Systems*, 32(3), 542–565. doi:10.1080/12460125.2022.2066051
[63] Cherns, A. (1976). 'The principles of socio-technical design', *Human Relations*, vol. 29, no. 8, pp. 783–792, Aug. 1976, doi: 10.1177/001872677602900806.
[64] Ackoff, R, and Emery, F E (1972). *On Purposeful Systems*. Chicago: Aldine-Atherton.
[65] Ackoff, R (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: Wiley.
[66] Alvesson, M and Spicer, A (2012) A Stupidity-Based Theory of Organizations. *Journal of Management Studies*, November 2012. 49(7), 1194–1220.
[67] Bertalanffy, L von (1968). *General Systems Theory: Foundations, Development, Applications*, New York: George Braziller.
[68] Beer, S (1985). *Diagnosing the System for Organizations*, Chichester: Wiley.
[69] Bergvall-Kåreborn, B. (2006). Diversity is in the Air – It is time to Challenge our Weltanschauung. *Systemist*, 28(2), 13-25.
[70] Brunsson, N (2002). *The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations*, Copenhagen: Copenhagen Business School Press.
[71] De Zeeuw, G (1997). Knowledge Acquisition in Changing Realities, *Kybernetes*, 26 (6-7), 837-847.
[72] Feynman, R.P. (1974). CargoCult Science: Some remarks on science, pseudoscience and learning how to not fool yourself. Caltech's 1974 Commencement address. *Engineering & Science*, 37(7), 10-13.
[73] Friis, S (1995). The PROTEVS Approach: A Short Presentation of Background, Principles and Methods. *AI & Society*, 9(2): 193-207.
[74] Nissen, H-E (1984). Acquiring knowledge of Information Systems – Research in a Methodological Quag-mire, in Mumford et al (1984). Research Methods in Information Systems. *Proceedings of the IFIP WG 8.2 Colloquium*, Manchester Business School, 1984, pp 39-52. Amsterdam: North Holland.
[75] Yuan, S. (2026). World and war in the age of AI: cyber risk, deterrence, and decision-making in U.S.-China competition. *Defence Studies*, 1–28. https://doi.org/10.1080/14702436.2026.2718345
[76] Poole, R. M. (2000). A UK perspective on information warfare. *Air Power Review*, 3(2), 43–57. https://www.raf.mod.uk/what-we-do/centre-for-air-and-space-power-studies/aspr/apr-vol3-iss2-5-pdf/

[77] Hall, S.C. (2026) 'The Silent Battlefield: Information Warfare Maneuver in the Era of Strategic Competition', *Military Review*, Online Exclusive (July), available at: https://www.armyupress.army.mil/Journals/Military-Review/Online-Exclusive/2026-OLE/The-Silent-Battlefield/ (Accessed: 24 August 2026).
[78] Vladu, L., Bârgăoanu, A., & Nastasiu, C. (2025). Information Warfare: Adapting to the Ever-Changing Nature of War. *International Journal of Intelligence and CounterIntelligence*, 38(2), 348–368. https://doi.org/10.1080/08850607.2024.2408713
[79] Wallner, C., Copeland, S. and Giustozzi, A. (2025) *Russia, AI and the Future of Disinformation Warfare*. London: Royal United Services Institute (RUSI). Available at: https://www.rusi.org/explore-our-research/publications/emerging-insights/russia-ai-and-future-disinformation-warfare (Accessed: 24 August 2026)
[80] Ford, M. (2024). From innovation to participation: Connectivity and the conduct of contemporary warfare. *International Affairs*, 100(4), 1531–1549. https://doi.org/10.1093/ia/iiae061.
[81] Tambini, D. (2021). Algorithmic pluralism: Media regulation and system resilience in the age of information warfare. In T. Clack & R. Johnson (Eds.), *The world information war: Western resilience, campaigning, and cognitive effects* (pp. 231–254). Routledge.
[82] Churchman, C. W. (1968). The systems approach. Delacorte Press.
[83] Wilcox, A. B. and Metcalf, C. (2026). "AI Command and Staff—Operational Evidence and Insights from Wargaming," *Military Strategy Magazine*, Volume 10, Issue 4, winter 2026, pages 4-10. https://doi.org/10.64148/msm.v10i4.1
[84] Mishra, P. (2026). Command without control: Mission command in the age of artificial intelligence. *Military Review*, (September-October). https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/September-October-2026/Command-Without-Control/
[85] Bardhan, I., Damsgaard, J., Lee, G., Sarker, S., Whitley, E. A., & Xu, H. (2026). Beyond codes of conduct: Shared values for digitalization. ACM Transactions on Management Information Systems, 17(2), Article 7. https://doi.org/10.1145/3810951
[86] The Economy Editorial Board. (2026). Military AI control and the verification gap. *The Economy AI Review*. 2026, September 16th. https://economy.ac/review/2026/09/202609295622
[87] Kott, A. (2006). *Information warfare and organizational decision-making*. Artech House
[88] Greenhalgh, T., & Shaw, J. (2025). Algorithmic amplification and operational friction: Empirical analysis of automated triage failure in public healthcare. *BMJ Health & Care Informatics*, 32(1), Article e101142. bmj.com
[89] Papagiannidis, E., Mikalef, P., Conboy, K., & Van de Wetering, R. (2023). Uncovering the dark side of AI-based decision-making: A case study in a B2B context. *Industrial Marketing Management*, 115, 253–265. https://doi.org/10.1016/j.indmarman.2023.10.003
[90]