HOZINT
HORIZON INTELLIGENCE

12 AUG 2025

Decision-Making Paralysis & Trust Erosion: From Psychological Warfare to an AI-Led "Information War"

By Antonella Calò

Information Warfare is no longer confined to the battlefield; it permeates every layer of society. From institutional decision-making and intelligence analysis to everyday users seeking trustworthy news, we are all operating in an environment shaped by increasingly sophisticated disinformation tactics. What was once a secondary concern in OSINT operations is now a primary threat: the credibility of sources and content is under continuous attack.

While not a new phenomenon, Information Warfare has evolved into one of the most complex hybrid threats, especially with the rise of Generative AI. As highlighted by organisations like NATO, EDMO, and the Joint Research Centre, AI-powered deepfakes, doctored footage, and manipulated texts are now weaponised to distort perception, paralyse decision-making, and undermine public trust.

At Hozint, ensuring source reliability and content credibility has always been at the core of our methodology. But in light of recent developments, we are enhancing our approach, already built upon NATO's AJP-2 standards, by leveraging AI to support analysts in identifying and navigating the growing "credibility trap."

This article examines the evolving strategies of disinformation and AI-driven manipulation, and their implications for situational awareness in an era of digital deception. We also briefly discuss potential approaches to enhance the detection and mitigation of disinformation.

Challenges in Source Reliability and Content Credibility

Malicious actors employ a wide range of strategies to manipulate information. These are compounded by unintentional actions, such as the widespread resharing of unverified content, which also contribute significantly to the spread of disinformation. As previously highlighted, these hybrid threats are multimodal and increasingly realistic, often triggering what is known as the Liar's Dividend: a phenomenon where the sheer volume of manipulated content leads people to distrust even accurate information.

When highly convincing visual content is combined with sophisticated textual or audio manipulation, it becomes increasingly difficult to escape the credibility trap they create. Moreover, disinformation is no longer limited to emotional language or overtly biased narratives. Today's manipulation strategies are far more subtle and complex, making even text-based distortions harder to detect and more dangerous in their ability to shape perceptions.

To begin, it's essential to distinguish between the different tactics used in the spread of misinformation and disinformation:

Framing: This is a subtle yet powerful manipulation technique in which factual information is presented in a way that steers interpretation. By selectively emphasising or omitting details, using metaphors, repetition, or specific visual cues, actors can guide audiences toward a desired narrative. Framing is particularly dangerous because it does not rely on false information per se, but on how the information is packaged. Its sophistication allows it to not only reinforce existing biases within echo chambers but also mislead well-intentioned individuals seeking accurate information. Disinformation about migration in Europe often relies on **framing migrants as threats** to society. Recent studies by the EU have found that false narratives frequently portray migrants as dangers to Europeans' health, wealth, or cultural identity. For example, in August 2023, Greece's Evros region (a key entry point for asylum seekers from Turkey) was engulfed in the largest wildfires ever recorded in the EU, killing at least 20 people; many of them migrants who were hiding in the forests. Instead of uniting the country in mourning, these fires became the backdrop for a disinformation campaign that falsely blamed migrants for the disaster. Almost immediately after the fires broke out, **conspiracy-laden stories** began spreading. "600 more [migrants are] ready to burn you... they will burn the city down", urging citizens to form groups to "protect the homeland" or "We are at war – illegal immigrants... have set more than 10 fires". Other examples of framing can also be found in Ukrain-Russia related contents, with even more sophisticated strategies.

Translation and Transcreation Manipulation: Disinformation is often tailored, or localised, to specific audiences, whether national populations or niche online communities. As a result, translation becomes a critical factor in both the detection and dissemination of manipulated narratives. Simply translating a piece of content into English (or any other language) may not be sufficient to uncover the underlying framing tactics or narrative distortions. In some cases, translation itself is weaponised: content is deliberately mistranslated to alter meaning, inject bias, or provoke division. During the war in Ukraine, for example, Chinese-language disinformation campaigns paired genuine video footage with fabricated subtitles, completely distorting the statements of Ukrainian officials. This tactic, known as transcreation, goes beyond translation or localisation; it actively reshapes the message to manipulate perceptions across linguistic and cultural lines.

However, the propagation of disinformation goes far beyond textual manipulation; it also involves psychological strategies and technically sophisticated tactics designed to exploit both individual cognition and digital ecosystems.

Artificial Amplification and the Illusion of Consensus

Disinformation campaigns frequently rely on coordinated bot networks, troll farms, and fake accounts to artificially amplify false or misleading narratives. This manufactured amplification creates a false sense of consensus and perceived credibility. The more often a piece of content is seen, the more likely it is to be accepted as true by the average user.

During recent conflicts, such as the Iran-Israel escalation, armies of fake and automated social media accounts disseminated AI-generated videos, like viral deepfakes of missile strikes and downed jets, to millions. Similarly, Russian influence operations have employed bot-driven campaigns across platforms to reinforce preferred narratives while drowning out dissenting or fact-based information. These efforts entrench users within filter bubbles and echo chambers, where repeated exposure reinforces belief, even in demonstrably false claims.

Search Engine Optimisation (SEO) Exploitation

Closely linked to amplification, SEO manipulation is another key tactic. Disinformation actors leverage marketing strategies, such as keyword stuffing, traffic hijacking, and exploiting data voids, to dominate search results.

Bot and troll networks can fill gaps in online discourse by introducing new narratives and repeatedly sharing them until they become trending topics. Conversely, they can hijack already trending keywords to attach visibility to completely unrelated or misleading content.

Moreover, these networks often coordinate to drive traffic toward a specific "anchor source," which may appear authoritative while being entirely fabricated. With the rise of AI agents capable of automating and scaling these efforts, this tactic is only set to grow in sophistication and reach.

Doppelgänger Operations: Cyber Deception Meets Disinformation

One of the most insidious hybrid threats is the Doppelgänger operation, a blend of cyber tactics and information warfare. At the basic level, this may involve spoofing trusted domains by slightly altering top-level domain names (e.g., nationalgeographic.**com.co** instead of nationalgeographic.**com**). More complex operations involve cloning entire news websites to create fake but convincing replicas that publish fabricated or heavily distorted content.

In 2024, German investigative outlet CORRECTIV exposed how Russian networks conducted such operations, cloning reputable domains like spiegel.de into spiegel.ltd. These campaigns used rotating domains, traffic obfuscation, and cybercriminal infrastructure to bypass detection by social media platforms and content moderation tools.

Doppelgänger campaigns represent a full-spectrum hybrid threat. They combine psychological manipulation, linguistic deception, and technical obfuscation to exploit trust, deceiving not only readers but also platforms, search engines, and even regulators. Their "credibility trap" is not merely narrative but structural, making them especially difficult to counter.

From AI-weapons to AI for Good: how AI can support OSINT analysts

As AI-War is always closer, for us at Hozint it is always increasingly important to make sure to provide trusted information and situational awareness solutions, while supporting our Analysts against Information Overload and disinformation and misinformation strategies. In a world where AI can be used as a weapon, we want to leverage it in an ethical way.

Firstly, we want to work on AI as a support for Analysts in assessing sources' reliability and content credibility. Once again, AI will not be a replacement for humans. Hybrid Tactics need Hybrid Responses. A Multi-Layer Solution is essential. Furthermore, our AI-based solution could have a dual usability: supporting analysts from one side, and revealing disinformation and cognitive strategies patterns and clusters on the other, making our platform always closer to investigating new and more sophisticated Hybrid Threats.

But how can we achieve this? How can AI support OSINT Analysts in filtering among reliable and trusted sources and content?

Among the various experiments and solutions to test, here are some ideas on how AI can support analysts both in assessing reliability and in investigating new hybrid threat patterns.

Following the order of the disinformation strategies presented previously, we will start with linguistic and narrative-related strategies.

Both **Framing and Multilingualism** can be supported by the use of Agentic AI and Large Language Models, which are increasingly improving their multilingual skills, supporting OSINT companies and analysts in detecting mistranslations, and/or performing preliminary automatic analysis of contents in several languages, which, of course, need to be checked by a human native speaker. Several solutions are being tested lately, just like LlamaLens, which can already support the analysts in checking coherence between titles and the news of texts (thus also coherence with SEO keywords), perform fact-checking and linguistic analysis. Framing analysis is still complex, and this is where we want to put our efforts in research.

Moving to metadata and more technical aspects, some of the viable solutions may be related to increasing the extraction of metadata. Our AI crawlers can be enhanced to not only extract relevant articles and reports, but also to comprehensively collect and evaluate key metadata, such as author, publisher, publication date, domain registration (WHOIS checks), and even location or topic clustering. This supports analysts in quickly identifying suspicious sources, such as doppelganger websites or domains with unusual propagation patterns, while maintaining a continuously updated reliability database.

AI can flag anomalous propagation of news, such as rapid distribution across obscure domains or clusters of articles that link back to suspicious anchors, by automatically analysing temporal data and SEO keywords. This identifies signs of bot/troll amplification and SEO abuse, which are increasingly common tactics in hybrid threats. Analysts can then focus their expertise on these red flags, instead of focusing on content from the whole Internet.

It is important to note that we are aware of the fact that metadata richness should be intended as a positive sign, but it could itself be part of a malicious strategy. What if the names of authors are simply pseudonyms? Or they can even be real, and still want to disseminate low-quality content. By leveraging AI for preliminary analysis, to be then confirmed or corrected by analysts, our databases will avoid static scoring, and will also keep track and monitor scores per source and authors (if any). While this information will be kept private, it will support the continuous building of a credibility trap table per content, depending on the several analyses performed, while also highlighting and revealing possible clusters of threats.

Data and Information Findability, accessibility, transparency, interoperability, and shareability are all important in OSINT analysis, and now more than ever, it is important for us to leverage them for ethical AI use in mitigating disinformation and psychological warfare.

We are still at the very beginning of our research and development journey, aiming not only to improve the real-time discovery of information but also to equip analysts with new, more powerful tools to defend themselves from today's rapidly evolving information warfare.

About The Author:

Antonella Calò, former Intern at HOZINT - Horizon Intelligence, is a **National Ph.D. Candidate in "Regulation, Management, and Law of Public Sector Organisations**" (38th Cycle) at the University of Salento. Her academic research intersects with applied projects in cooperation with the **European Commission's Joint Research Centre (JRC),** focusing on Critical Entities Resilience, Hybrid Threats detection, and computational linguistics applications in defence, security, and societal preparedness.

In 2022, leveraging a background in linguistics, Antonella joined the **Datalab at the Department of Engineering for Innovation** (University of Salento), where she expanded her expertise in Data and Information Sciences. This led to her specialisation in Computational Linguistics, including Natural Language Processing (NLP), Large Language Models (LLMs), and Ontology Engineering.

Her work also integrates principles of **Cyber and Open-Source Intelligence (OSINT)**, particularly in the context of threat detection and information environments. Her main research focus revolves around the availability, quality, lack of harmonisation, and/or manipulation of information, and AI-based methods to mitigate their negative impacts on societal resilience. Her work at HOZINT, also consisted of doing research and sharing her perspectives on source reliability and content credibility methodology assessment in the era of AI-led Information and Cognitive Warfare.