

The Central European University & Institut Barcelona d'Estudis Internacionals

2023-2025

AI Firms Enabling Hybrid Warfare Operations in Israel-Gaza and Russia-Ukraine: Assessing Transparency and Accountability

Dissertation Submitted By

SYDNEY MARTIN

In partial fulfillment of the requirements for the degree of ERASMUS MUNDUS MASTER'S
PROGRAMME IN PUBLIC POLICY (MMAPP)

SUPERVISORS: Josep Ibáñez (IBEI) & Marie-Pierre Granger (CEU)

The United States, 31 July 2025

AUTHOR'S DECLARATION

I, the undersigned, Sydney Martin, candidate for the MMAPP degree in Global Public Policy declare herewith that the present thesis titled "AI Firms Enabling Hybrid Warfare Operations in Israel-Gaza and Russia-Ukraine: Assessing Transparency and Accountability" is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

If you are a student of a joint program and you are required/have the option to submit the same thesis to CEU and the partner university as well, use this paragraph instead of the last paragraph above. Other CEU students should delete this section. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree, except as part of the co-tutelle agreement between Central European University Private University and [University B].

Vienna, 31 July 2025

Sydney Martin

Abstract

Synthesizing literature on global power relations, hybrid warfare, artificial intelligence, international law, corporate governance, and various transparency frameworks related to these concepts, this research paper explores the question: “*what AI tools and firms are enabling hybrid warfare operations in the ongoing Russia-Ukraine and Israel-Gaza wars, and how do these meet transparency standards?*” As a primarily exploratory and descriptive study, this paper produces a dataset of 65 AI tools and 25 AI companies involved in the two case study conflicts. Based on theoretical frameworks gathered from AI and corporate transparency literature, the (1) AI tools and (2) companies received two scores indicating (1) algorithmic and (2) corporate transparency. The overall findings suggest that the vast majority of AI included in this dataset is overwhelmingly opaque, with severe deficiencies in terms of meeting democratic oversight, transparency and accountability standards. The collected dataset and a breakdown of the subsequent transparency scores are compiled into a centralized document that may assist future researchers and policy-makers interested in exploring transparency and accountability incentives, mechanisms, and outcomes, specifically for dual-use AI enabling hybrid warfare.

Table of Contents

1.0 Introduction

2.0 Literature Review & Theoretical Framing

The Evolving Nature of Power and Modern Warfare

Hybrid Warfare and 21st Century Artificial Intelligence

The Inherent Risks of Mainstreaming Dual-use AI & Implications for Accountability

Challenges Posed by Private Tech Firms Involved in Hybrid Warfare

Transparency in AI-Ecosystems

3.0 Methodology

Overview

Case Studies

Selection Procedure

CLeAR and TAiL Scoring Frameworks

Limitations

4.0 Results & Analysis

CLeAR Scores

TAiL Scores

Case Highlight: Palantir

5.0 Conclusion & Implications

1.0 Introduction

In light of changing geopolitical issues, new transnational actors, and the mainstreaming of powerful information and communication technologies, the nature and scope of warfare and global conflict are expanding dramatically. Modern warfare increasingly takes place in spaces where the line between militant and civilian, war and peace, truth and disinformation are deliberately blurred. This new form of warfare – coined hybrid warfare – involves a variety of tactics that deploy both conventional military weapons, like bombs or soldiers, with non-military methods, such as cyberattacks, surveillance, disinformation campaigns, and cognitive manipulation. The strategic advantage of hybrid warfare tactics, as opposed to traditional tools of war, include plausible deniability, cost efficiency, and the capacity to achieve political or military objectives without triggering formal declarations of war or conventional deterrence mechanisms.

Hybrid warfare is currently being driven by rapid advancements in artificial intelligence (AI) and its integration into information, cyber, and cognitive warfare practices. Dual-use AI technologies are now enabling a variety of state and non-state actors to collect, manipulate, and weaponize data at an unprecedented speed and scale. By creating and amplifying mass amounts of disinformation or deepfakes, surveilling behavior for predictive analytics, or generating super-viruses to take down computer networks across entire regions, AI is facilitating hybrid operations that erode public trust, undermine democratic institutions, and exacerbate accountability gaps.

While the capabilities of AI in hybrid warfare are becoming more visible, the transparency of both the algorithms used and the companies that develop them remain alarmingly opaque. Due to the highly complex nature of AI decision-making processes, AI tools frequently obscure how they make choices, how they process data, and who bears responsibility for their errors or misuse. Moreover, the private technology companies responsible for manufacturing many of these dual-use AI tools frequently operate in obscurity. In the context of the global AI arms race and in a time when societal infrastructure is increasingly relying on AI algorithms, AI companies have the expertise, technology, and capital to influence political and security outcomes, all while discouraging democratic oversight. This ability to circumvent transparency while exerting

considerable influence over public goods and policies complicates attempts at legal attribution, regulation, or ethical evaluation, widening accountability gaps.

This thesis attempts to address some of these gaps by asking, “what artificial intelligence (AI) tools and firms are enabling hybrid warfare operations in the ongoing Russia-Ukraine and Israel-Gaza wars, and how do these meet transparency standards?” To answer this question, two complementary transparency assessment frameworks are created and applied:

1. The CLeAR framework (Comparability, Legibility, Actionability, Robustness) to evaluate algorithmic transparency of AI tools; and
2. The TAiL framework (Transparency, Accountability, Legality) to assess corporate transparency and governance of the firms that develop these tools.

By collecting data on 65 AI tools and 25 firms, a descriptive dataset is made that exposes the distribution of transparency across both algorithmic and corporate dimensions, highlights accountability gaps, and provides a starting-point for future policy and regulatory discussions on dual-use AI in hybrid warfare. In the context of international law and corporate governance, the findings of this thesis contribute to the underexplored intersections of security studies, AI ethics, and global governance; it offers insights for future research, policymakers, and civil society that are seeking to conceptualize and address the risks of increasingly autonomous, opaque, and influential AI systems in conflict.

2.0 Literature Review & Theoretical Framing

The Evolving Nature of Power and Modern Warfare

In the state-centric world of the past, it was primarily military conquest and other forms of hard power – war capability, geography, population, and resources – that determined international power hierarchies (Nye 1990, 154; Garon 2022, 183). This state-centric system functioned within the limits of strict borders and legal frameworks, it held states as the primary actors in global politics, and conflict was clearly signaled with formal declarations of war. Today, however, we live in a notably different global context: economic interdependence, new transnational actors, changing political issues, rising nationalism, and the spread of technology

have diffused authority away from states and fundamentally reshaped what constitutes power in the global arena (Nye 1990, 167-168 & 160). Tools like capital, technology, and information control are used to achieve the same political or economic goals as traditional warfare, at lower financial and social costs (ibid, 162; Garon 2022, 181 & 186). Private companies, technology tycoons, proxies, and other non-state actors like terrorist groups are able to leverage significant capital, gatekeep access to resources and information, and influence public perception in ways that alter foreign policy outcomes (Jokinen et al. 2022, 7-10; Qureshi 2020, 903). This is a marked shift away from a period of international relations wherein state military strength and other hard powers determined international hegemony.

As a consequence of the changing nature of power and the emergence of new actors in the global arena, experts use the budding framework of "hybrid warfare" to understand modern transnational conflicts. Hybrid warfare is a broad term encompassing operations carried out by state and/or non-state actors, wherein both military and non-military tactics are used to achieve strategic or political goals (Wigell 2019, 256-257, 259; Virtosu & Goian 2024, 198). Hybrid warfare distinguishes itself from traditional warfare through, "...the conscious blurring of the line between war and peace, providing the aggressor with 'plausible deniability', and thereby reducing the risk of military blowback while also minimizing the costs associated with launching a full-scale military attack" (Wigell 2019, 257). This means that hybrid warfare operations may begin prior to official declarations of war and continue after peace agreements. The inherent ambiguity of hybrid operations creates a shield for agents to subvert blame for any harms that are outside of the traditional parameters of open warfare (ibid).

In practice, scholars agree that hybrid warfare includes instances when, "... special operations, espionage, special forces, and asymmetric operations are combined with non-military methods, such as propaganda, disinformation, manipulation of public opinion, cyber-attacks, and subversion," all deployed at varying intensities depending on the strategic objectives (Virtosu & Goian 2024, 198). Empirical examples illustrate the diversity and obscurity inherent in hybrid warfare: Russia's meddling in the 2016 U.S. presidential election through coordinated disinformation campaigns and cyber attacks; China's maritime militias' territorial claims in the South China Sea; the NSO groups' spyware, used for espionage and surveillance; and Hezbollah's mix of military force, political influence, and social services in its operations against

Israel (Wigell 2019, 255; Yoon & Wonhee 2023; Amnesty International 2022; Azani 2013). Notably, while scholars generally agree that these are all examples of hybrid warfare, few would assert, for instance, that the US was "at war" with Russia during the 2016 presidential election, or that Israel was "at war" with Spain when NSO Group – a company that is widely considered a de-facto arm of the Israeli state – conducted espionage operations on Spanish civilians (ibid). This is because hybrid warfare operates below traditional thresholds for open armed conflict, thereby complicating frameworks for international law, deterrence, and response.

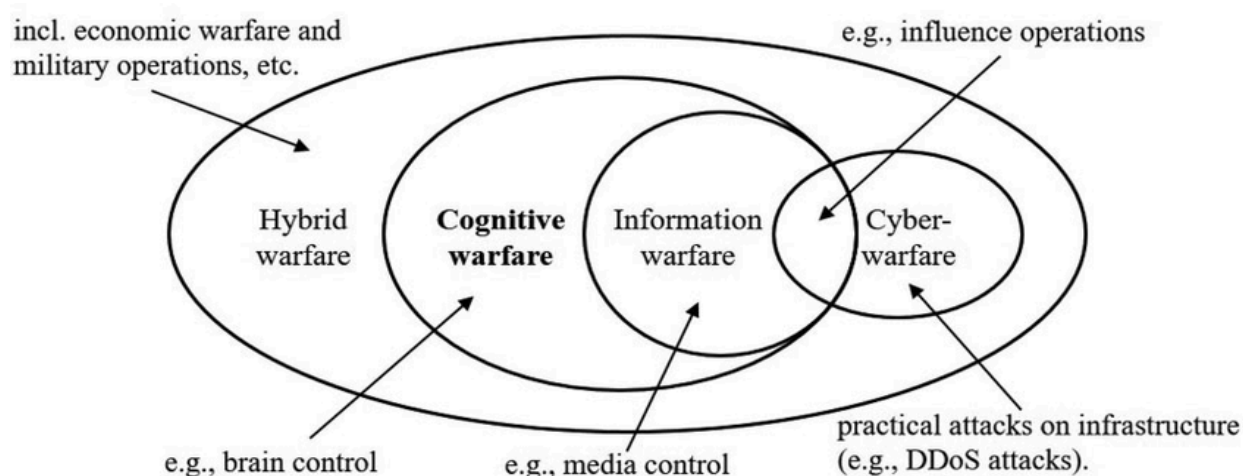
The covert nature of hybrid warfare has led some scholars to reject the term altogether. These scholars argue that,

"...branding non-military tactics... as 'war' may spur a potentially dangerous process of securitization... [and] doing so not only blurs the distinction between foreign policy and war, but also lowers thresholds of threat perception, even the threshold of war itself, and may thus unleash unnecessary escalation" (Wigell 2019, 258).

While these are reasonable concerns, dismissing non-military tactics during conflict makes it impossible to grasp an accurate understanding of how modern power struggles unfold, and why they may unfold in favor of one actor over another (ibid; Nye 2004, 167-168; Garon 2022, 181 & 186). Whether it spurs securitization or not, modern conflicts rely heavily on a fusion of military and non-military power. Consider Ronald Reagan's anti-communist information campaigns during the Cold War that, while not directly lethal, played a significant role in shaping global ideological positionings (Qureshi 2020, 911-912). Other examples include the 2003 U.S. invasion of Iraq, which was preceded by extensive psychological operations, intelligence manipulation, and public diplomacy aimed at justifying regime change; or consider Russia's 2014 annexation of Crimea, which was the result of a culmination of cyber attacks, media propaganda, political destabilization, and covert operations (ibid, 918; Miron 2023). In all of these examples – and in most modern conflicts – information manipulation, cyber operations, economic coercion, and political subversion are critical for determining geopolitical outcomes. Therefore, even though hybrid warfare may entail some risks of securitization, it is an ever-evolving concept that is currently necessary for understanding modern power struggles.

It is generally agreed that within hybrid warfare, there are a number of subtypes that are characterized by the specific non-military tools used to assert power, including: information warfare, cyber warfare, and cognitive warfare. While these three subtypes can be distinct, there is considerable overlap between them. In a study focusing on the inner workings of China's cognitive warfare and Taiwan's anti-disinformation campaign, Tzu-Chieh Hung and Tzu-Wei Hung (2022) conceptualize the intersections of hybrid, information, cyber, and cognitive warfare using the following visual:

Figure 1.



While these subtypes and tactics are not necessarily new concepts in traditional warfare frameworks, the rise and mainstreaming of the internet, social media, and other communication technologies has drastically increased the relevance of non-physical spaces and digital tools when it comes to determining the outcomes of global power struggles; information can be gathered, manipulated, and disseminated at a scope and pace previously unimaginable, and in a way that fundamentally shapes public perceptions of reality and thereby geopolitical outcomes (Taddeo 2016, 215). The importance of technological advancements cannot be understated in the context of hybrid warfare, and will be explored in the following section. For now, a general outline of these three hybrid subtypes will be provided.

Beginning with the outermost circle of Hung & Hung's (2022) typology, hybrid warfare involves activities which combine military and non-military tactics to achieve political, economic, or strategic goals. Cognitive warfare (CogW), the second-broadest circle, involves "... activities

designed to control others' mental states and behaviors," making the human mind the primary battlefield (ibid, 2). CogW depends in no small part on information warfare (IW), the main difference being that the latter designates the information environment as the battlefield as opposed to human minds. To wage IW, actors typically gather and/or withhold truthful information; they may also leverage disinformation. Disinformation is the intentional spreading of false or manipulative information with the goal of influencing public opinion, generating confusion, and undermining overall trust in legitimate institutions (Virtosu & Goian 2024, 198). Disinformation as part of IW can thus manifest in the proliferation of false, selected, or manufactured sources that support the agendas or narratives of the promoter; these narratives often target key audiences and socio-demographic categories that are most relevant to achieving a particular geostrategic goal (ibid, 199). Additionally, disinformation can be used to distract or divert public scrutiny away from more severe issues, and to manipulate the emotions of the public by pushing shocking or inflammatory imagery, speeches, and conspiracy theories that spur people into action in support of one actor or another (ibid).

Importantly, IW and CogW together are an ongoing and snowballing process, typically started with just a few pieces of disinformation. Almost naturally, disinformation eventually turns into misinformation, which is the unintentional spread of false or misleading information (Virtosu & Goian 2024, 206). This happens when disinformation is shared from source to source, acquiring credibility through online reposts and likes – a concept known as layering – leading to the spread of disinformation by people who do not realize the information's falsity. According to one study looking at how mis/disinformation is spread online, the authors found that, "falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information..." (ibid). If the IW campaign succeeds, it ultimately leads to integration, where mis/disinformation is adopted by credible news outlets and institutions, generating political and social influence that favors a particular actor in a power struggle (ibid, 205-206). This process is particularly hard to contain in the context of liberal democracies, because, "...whereas authoritarian states can suppress media sources that are undesirable for the regime, in a liberal democracy the right to alternative sources of information for all citizens is protected by law and cannot be easily censored" (Wigell 2019, 270).

Finally, cyber warfare (CyW) typically involves the attack or defense of a computer network system or within the cyberspace environment. In the case of an offensive attack, the intent is usually to disrupt, destroy, or gain access to the information and communication systems of an adversary (Robinson et al. 2015, 72-74). This can range from denial-of-service (DoS) attacks that overwhelm systems, to the deployment of malware that corrupts data, to sophisticated intrusions aimed at stealing sensitive information or controlling critical infrastructure. Beyond these direct assaults, CyW also involves activities like cyber espionage, where sensitive data is collected for intelligence gathering, or data manipulation, where information is subtly altered to sow confusion or undermine trust. The impact of successful CyW can be far-reaching, leading to significant economic damage, loss of public trust in institutions, and even physical consequences if critical infrastructure is targeted (Weiner 2022). CyW operations are often characterized by their stealth and the inherent difficulty in attributing responsibility, making them a potent tool in hybrid conflict by offering a low-cost, high-impact means of achieving strategic objectives with plausible deniability.

As mentioned, these non-military hybrid tactics – like espionage or propaganda, for example – are not new concepts to global actors. The increasing political weight of soft-power, partnered with rapid technological advancements, has significantly altered the scope and scale of hybrid operations happening in the 21st century (Taddeo 2016, 215; Burton 2023; Crootof 2022, 1-3). These evolutions have expanded the number of actors involved in both domestic and transnational political issues, confused traditional thresholds for war and peace, and fundamentally altered the nature of state and non-state activity. The following section will explore the centrality of different technologies – particularly digital tools and artificial intelligence – in redefining how hybrid warfare is planned, executed, and perceived on the global stage.

Hybrid Warfare and 21st Century Artificial Intelligence

Innovations surrounding information and communication technologies (ICTs) – like the internet, social media, smartphones, cloud computing, and more – combined with recent breakthroughs in generative artificial intelligence (AI) – like machine learning (ML), large language models (LLMs), natural language processing (NLP), neural networks, etc. – have significantly expanded

the scope, scale, and stealth of hybrid warfare operations (Monohan 2023; Burton 2023, 2; Garon 2022, 181 & 183). While many ICTs, AI programs, and digital technologies are not originally developed for hybrid warfare purposes, these technologies are considered dual-use, meaning that they can be leveraged by civilians, states, and non-state groups for both personal and military applications (Crootoof 2022, 1). State and non-state actors alike can now exploit digital infrastructure, AI, and big data to influence, disrupt, and surveil across borders and achieve their geopolitical objectives – all without engaging in kinetic conflict or triggering legal frameworks that would otherwise govern their behavior (Virtosu & Goian 2024, 206; ICRC 2021, 465, 467-468; Wigell 2019, 257; Qureshi 2020, 924). Consequently, these developments allow for more efficient and scalable forms of IW, CogW, and CyW that smudge the boundaries between war and peace, combatant and civilian, truth and disinformation. In order to understand exactly how AI has revolutionized hybrid warfare, it is first important to understand how AI operates and what it can be used for in the context of conflict and transnational power struggles.

Most broadly, AI is the simulation of human intelligence by machines and digital technologies, and it can be used for a wide variety of purposes. AI can automate data collection and analysis, problem-solve, and make decisions or predictions by using deep learning techniques that mimic human brain activity – i.e. artificial neural networks – which analyze, classify, and learn from massive amounts of data and information (Monahan 2023; Garon 2022, 192-193 & 213-215). Moreover, it can leverage newly developed machine learning (ML) algorithms to learn and adapt from data without ever being prompted by a human to do so (Barredo Arrieta et al. 2020). A common example of AI using ML is facial recognition software, which is able to distinguish between faces after learning patterns and facial features from a given dataset of images and videos (Sharma et al. 2020, 1162-1164). AI with ML can also be embedded within ICTs – like social media and smartphones – to enhance adaptability, efficiency, and accuracy in completing tasks. For instance, when integrated within online platforms, AI leveraging ML algorithms can predict user interests and needs based on demographic and use-behavior data, and then recommend content and products that will maximize user engagement and profit (Bernstein et al. 2023, 1).

Another common manifestation of AI comes in the form of large language models (LLMs), which are computer programs that are fed an enormous amount of text-based data sources –

ranging from books, to website content, to social media posts, to scholarly articles – so they can learn to comprehend language and create new text that is human-like (Monahan 2023). LLMs use natural language processing (NLP) techniques "...to learn and recognize patterns and identify relationships between words" (ibid). Popular AI using LLMs and NLPs include general-use chatbots like ChatGPT, writing assistants like Grammarly, or fact-checking tools like Get Clarity AI. LLMs can also be combined with ML and other deep learning techniques for an AI product that can, for instance, generate deepfakes – hyper-realistic visual and audio content – in response to a text-based prompt (Garon 2022, 199; Virtosu & Goian 2024, 214).

These technologies are incredibly useful for the purposes of IW, CogW, and CyW operations. First, AI-based digital tools can be used by hybrid agents to generate and spread massive amounts of mis/disinformation, propaganda, and deepfakes across online websites, social media, and other ICTs, reaching billions of people in a short timeframe (Hung & Hung 2022, 3, & 12-13; Virtosu & Goian 2024, 202; Garon 2022, 199). Furthermore, since AI is capable of harvesting and exploiting tremendous amounts of data, it is able to personalize mis/disinformation by appealing to user demographics, values, and cognitive biases. This makes AI generated propaganda incredibly convincing (ibid). Hybrid agents can also use AI to create millions of fake online profiles that use fabricated identities and credentials – these are known as bots – to produce propaganda that mimics the style and format of credible news articles, scholarly publications, or even the likenesses of political figures (Virtosu & Goian 2024, 183 & 201-202). Then, more AI-driven bots can be used alongside human users to amplify this content by automatically liking, sharing, and commenting on it, which creates the illusion of popularity and reliability (ibid; Garon 2022, 187). For example,

"A research from January 2023, realised by the analytical centre NewsGuard discovered that the popular chatbot ChatGPT is capable of generating texts that extend current conspiracy theories and integrate genuine events in their context. This programme has the ability to automate the delivery of numerous messages (via bot farms), the topic and tone of which will be selected by a person and the direct text created by AI" (Virtosu & Goian 2024, 212)

In this way, ML, deep learning, LLMs, and NLPs make AI-generated disinformation, deep fakes, and other forms of propaganda incredibly effective at facilitating hybrid warfare operations

aimed at influencing public opinion, polarizing populations, and manipulating perceptions of reality (Burton 2023, 5).

Beyond these applications for IW, AI can also be embedded within broader IT infrastructures that are used by state militaries and non-state actors for daily operations and hybrid activities like CyW. IT infrastructure involves technology that organizes, "...holds, routes, assembles, and shares information," to maximize an organization's efficiency and reduce operation costs; examples include an actor's servers, storage systems, computers, operating systems, applications, and cloud services (Nyrhinen 2008, 4). When integrated with AI, these systems can be used to automate logistics and decision-making, optimize battlefield or other communications, manage and analyze real-time surveillance data, or conduct predictive analytics to anticipate behavior (Burton 2023, 7; ICRC 2021, 468-469). In the context of hybrid CyW, AI powered IT infrastructures become a strategic advantage that allow actors to coordinate information, cyber, and cognitive operations, quickly and accurately, across decentralized networks. For example, militaries may use AI-integrated infrastructure to simultaneously manage drone fleets, analyze intercepted communications, and launch tailored disinformation campaigns online (ibid).

While not the direct focus of this paper, it is relevant to briefly mention that AI can also be integrated directly within combat weapons known as Lethal Autonomous Weapons (LAWs) – including drones, sentry guns, and missile defense systems. These systems, once activated, can select and engage targets without further human intervention, raising profound ethical and legal questions. Since this paper is focused on AI enabling hybrid operations that manipulate soft powers – e.g., information, ideologies, and other intangible resources – LAWs remain outside of its scope. However, for clarity, AI-empowered drones that collect data and surveillance, or serve some other strategic purpose for IW or CyW operations, without *directly* causing tangible harm, remain within scope; modern hybrid activities could not be pursued if not for the AI and machine technologies that harvest and synthesize massive amounts of data, whether through a drone on a battlefield, a satellite in space, or an algorithm on social media (Burton 2023, 4 & 8; Wigell 2019, 266).

The Inherent Risks of Mainstreaming Dual-use AI & Implications for Accountability

There is considerable disagreement among experts about the exact risks and challenges resulting from the mainstreaming of AI tools; the opportunities for exploitation and the potential for accidents are limitless, but simultaneously difficult to track, study, and regulate (Ho et al. 2023, 8). Experts warn that the increasing reliance on this technology creates new vulnerabilities, particularly in high-stakes and ambiguous environments like hybrid warfare scenarios (ibid). AI is not infallible, and empirical evidence of its imperfections and the scope of damage they can cause is rapidly making itself known as states, companies, and individuals are integrating AI systems into their official operations and personal activities. Understanding these multifaceted dangers is crucial for grasping the true implications of AI's role in contemporary conflict.

A primary concern arises from unintentional errors and inherent biases embedded within AI systems. Since AI learns from vast datasets often created by humans, it frequently internalizes and reproduces societal biases, including those related to race, gender, or nationality (Köbis et al. 2022, 3; Crootof 2022, 2). The American Civil Liberties Union cites multiple peer reviewed studies finding that AI facial recognition softwares used by police, government administrative agencies, and private companies consistently misclassified people of color, women, children and the elderly; these softwares are most accurate for middle-aged white men, while black women are misclassified almost 35% of the time (Crockford, 2020). This is largely because the datasets used to train facial recognition software, LLM programs, and other AI systems have quantitatively more information on middle-aged white men than, say, black people, who are systemically underrepresented in media, news, and research (Yang 2025). This has led to false arrests and detainments that disproportionately affect marginalized communities, contributing to their systematic political disenfranchisement (ibid).

Beyond biased datasets, any code-based system is susceptible to minor programming errors that result in unintended outcomes (Crootof 2022, 2). Because complex systems come with higher risks of mistakes per the basic tenets of normal accident theory, the extremely complicated AI tools used in hybrid warfare operations come with an increased likelihood of errors (ibid). While these bugs may initially seem minor, they can snowball into severe, even deadly consequences.

A recent example includes a US drone strike intended to kill an ISIS-K member in Afghanistan that, instead, killed 10 civilians due to automation bias and faulty AI facial recognition software (Crootof 2022, 2; Lopez 2021). This makes AI system errors in high-stakes military scenarios a serious cause for humanitarian concern.

While accidents have always been a grim reality in the "fog of war," new military technologies integrated with AI introduce new manifestations of old errors while also significantly magnifying their potential for harm; these tools facilitate action at "superhuman speed and unprecedented scale" (Crootof 2022, 2-3). The sheer velocity that AI systems operate at, coupled with their capacity to process and act upon vast amounts of data, means that errors can distribute with alarming speed across networked systems. Simultaneously, this rapid dissemination of flawed information or faulty decisions reduces opportunities for human oversight and intervention; the pace of AI operations often exceeds human cognitive and reactive capabilities (ibid, 4). This dynamic creates a "precision paradox," wherein the very ability to conduct more precise hybrid attacks – a seemingly positive development – may paradoxically lead to a greater overall number of attacks. Additionally, the "ironies of automation" suggest that as hybrid systems become more automated, the human role becomes – again, paradoxically – more critical and difficult; operators are prone to automation bias (over-reliance on machine-made decisions), deteriorated situational awareness, and skill fade (ibid, 2-3). A tragic example of this sort of interaction error is when the US Patriot system misclassified friendly planes as missiles, despite contradictory human information, resulting in "painfully avoidable fratricides" (ibid, 3).

Exacerbating these technical flaws is the pervasive risk of intentional manipulation and malicious interference; even well-designed AI systems using accurate data (which, in itself is a difficult thing to determine) can be exploited or tricked (Burton 2023, 7). Multiple studies have been published in recent years that explore some of the ways AI can be deceived:

"Researchers have already demonstrated how to fool an AI system into misreading a stop sign, by carefully positioning stickers on it. They have deceived facial-recognition systems by sticking a printed pattern on glasses or hats. And they have tricked speech recognition systems into hearing phantom phrases by inserting patterns of white noise in the audio" (Heaven 2019).

Beyond these tricks, hybrid agents can deliberately program and tweak AI algorithms to perpetuate biases or promote their agendas (Burton 2023, 6-7). This can manifest as "algorithmic capture," where AI fact-checking tools, for example, are designed to favor certain viewpoints, distorting users' perceptions of truth (Köbis et al. 2022, 8; Binns 2021, 4). Moreover, hybrid agents can introduce "insider threats" during AI development, creating backdoors to bypass security protocols and deploy AI-generated polymorphic malware, capable of disabling, destroying, or extracting data from IT infrastructure, all while evading traditional antivirus defenses (Ho et al. 2023, 5; Burton 2023, 7). This sort of interference has already occurred within critical institutions like hospitals, rendering their IT infrastructure useless and keeping thousands of people from receiving critical medical care (Weiner 2022).

Furthermore, the inherent complexities of AI systems – particularly their self-learning capabilities – create a significant accountability gap within various legal frameworks; it is often extremely technically difficult, if not impossible, to trace AI's decision-making processes or distinguish between autonomous AI actions versus human guidance or tampering (Köbis et al. 2022, 10-11; ICRC 2021, 476). This is known as "black box" AI, and more than just widening accountability gaps, it also makes predicting behavior and preventing mistakes from repeating difficult. In hybrid warfare scenarios wherein international or domestic legal frameworks only prohibit *intentional* civilian casualties and *tangible* harms like death, black box AI endows the hybrid agent with plausible deniability; AI can easily be blamed – whether true or not – for *accidentally* misidentifying a person as a target or *unwittingly* spreading disinformation that causes no direct, tangible harm (Virtosu & Goian 2024, 206; Köbis et al. 2022, 11; Crootof 2016, 1386-1390; Qureshi 2020, 901). In this way, current legal structures that solely regulate intentional and tangible harms incentivize "...the development and use of weapons that exploit the grey zones of permitted incidental harm" (Crootof 2022, 1).

The rapid pace of AI development and deployment further exceeds the capacity of legal and regulatory frameworks to adapt. The global nature of AI development and its dual-use potential mean that a unified international approach to governance is difficult to achieve, leaving significant gaps where malicious actors can operate with immunity (Ho et al. 2023, 8, 10, 12). For instance, existing International Humanitarian Law struggles to apply concepts like "direct participation in hostilities" to autonomous AI systems, and national laws often lack specific

provisions for AI-driven harms, creating a liability gap (ibid; ICRC 1977, Article 51). This regulatory vacuum, combined with the inherent opacity of advanced AI systems, creates fertile ground for the proliferation of unaccountable hybrid warfare tactics, making it increasingly difficult to deter, detect, and prosecute those responsible for harm.

Challenges Posed by Private Tech Firms Involved in Hybrid Warfare

As innovations like AI have changed the way war is waged, private technology companies have become indispensable actors in the execution, facilitation, and expansion of hybrid operations. Whereas traditional warfare prioritized militaries and hierarchically organized state actors, hybrid operations empower those who control the flow and structure of data, develop algorithmic systems, and influence the online environments society operates within (Nye 1990, 163-164). The rapid digitization of conflict, especially through IW, CogW, and CyW, has allowed private tech firms to shift from passive infrastructure-providers to active intermediaries in global conflict, operating in what Garon (2022, 186) calls a domain of "shadow sovereignty," in which private ICT and dual-use AI providers wield sovereign-like power over state military activity, political narratives, societal cohesion, and even the legitimacy of democratic institutions. Hence, on top of the technical risks that come with integrating dual-use AI technology into both broader society and hybrid operations, there are still more concerns to consider regarding private interests who develop and control access to these critical technologies.

A fundamental concern expressed by scholars who explore this diffusion of power away from states and towards companies is the issue of private interests being prioritized over public-goods without opportunities for democratic oversight. Corporate actors that manage social media, for example, are incentivized to embed AI algorithms into their platforms that will maximize user engagement and thereby increase their profits. Guided by robust evidence that humans are more likely to engage with shocking, exaggerated, or otherwise sensational social media content that may not be completely accurate, these companies are incentivised to ultimately encourage radicalization and conflict on their platforms (Virtosu & Goian 2024, 200; Garon 2022, 187). For example, Facebook (owned by Meta) and its AI algorithms have faced accusations – supported by ample evidence – of directly enabling the 2017 genocide of Rohingya Muslims in Myanmar by, "...helping the military and Buddhist majority population spread false propaganda against the

Rohingya and by systematically removing factual content posted by the victims of the ethnic cleansing" (Garon 2022, 190). The propaganda spread on Facebook by the Buddhist majority was disturbing – including fabricated reports of families setting fire to their own homes – and consequently prompted greater user engagement, which in turn caused Facebook's AI algorithm to amplify the mis/disinformation to even more viewers (ibid, 189). This shifted the political landscape and altered public opinion to such an extent that studies indicate Facebook played a "decisive" role in perpetuating the genocide against Rohingya Muslims (ibid).

Despite the considerable evidence, Facebook has not been held accountable in any official capacity for its role in the Rohingya genocide as a hybrid agent. Facing informal criticism, Facebook has since updated its AI algorithms to attempt to prevent the spread of mis/disinformation – not by automatically removing this content, but rather by labeling inaccurate posts as false or misleading (Virtosu & Goian 2024, 210). Nonetheless, a recent report by the Center for Countering Digital Hate – focused on tracking the spread of online conspiracy theories about the US funding bioweapons in the ongoing Russia-Ukraine conflict – found that Facebook failed to label 80% of these posts as dis/misinformation (ibid). Facebook's ability to continuously circumvent formal legal accountability despite its involvement in intense hybrid warfare operations has raised important questions regarding the applicability of different accountability mechanisms – especially legal frameworks – to private companies.

Arguably, existing legal frameworks are unable to hold technology companies sufficiently accountable for enabling hybrid warfare operations. Only potentially applicable – in instances of serious allegations like perpetuating genocide or other tangible and transnational harms – is international humanitarian law (IHL). However, IHL is not clearly befitting for holding private companies accountable due to its narrow view of who can be involved in global conflicts; according to most interpretations, IHL is written through a statist lens and intended to apply to either states or individuals – not companies (Hughes 2021, 48-49, 66). Beyond IHL, other frameworks like international human rights law (IHRL), international criminal law (ICL), and domestic legal frameworks are often assumed to be applicable to private firms in hybrid warfare; however, in practice these frameworks are often not tailored to the nuances of company structures or operations, nor do they adequately identify and regulate the kind of hybrid warfare activities that these firms may partake in (Hughes 2021, 53). IHRL is arguably the most readily

applicable to companies, as it claims to govern “‘individuals and every organ of society’” and covers a range of topics related to preserving public goods threatened by hybrid warfare (ibid, 57-58). ICL does criminalize the most egregious forms of corporate misconduct, such as trafficking, vast environmental derogation, and massive transnational corruption, but its applicability to the more subtle hybrid activities – like Facebook’s involvement in the Rohingya genocide – is less straightforward, according to legal experts (ibid).

Furthermore, all of these international legal frameworks typically require proof of *actus reus* (i.e. guilty act or omission) and *mens rea* (i.e. intentionality) in order to hold an actor accountable for a violation; as previously discussed, it is inherently difficult to trace or understand an AI system’s decision-making process, but what’s more is that it is also increasingly difficult to trace human-decision makers and their intentions within modern companies (Hughes 2021, 70-71; Crootof 2016, 1386). Attempts to attribute individual responsibility to corporate actors are often eluded due to the nature of collective decision-making within private corporations: “Complex management structures, the role and operations of subsidiaries, and the bifurcated nature of decision-making between the board and management create degrees of separation” (Hughes 2021, 71). Companies frequently subvert *mens rea* because they tend to separate those who have the information about an act’s implications and those that actually do the act (ibid). The entire point of the LLC corporate structure – common amongst many tech companies – is to maximize risk-taking but disperse liability away from individual owners of the firm, ensuring they are not held accountable for the company’s debts, legal violations, or harmful actions, even when they involve negligence, misconduct, or complicity in hybrid activities with serious consequences (ibid, 75). Authors also note that in these kinds of companies, subsidiaries can become undetectable hybrid agents, “... formed to undertake a singular transaction and distant the parent’s business interest from legal consequences or negative connotations” (ibid, 76).

At the domestic level, it is often unrealistic or undesirable for a state to hold private technology companies accountable to international legal frameworks — because of the unavoidably opaque nature of black box AI that makes attributing fault difficult, and for fear of stifling innovation in their country (Ho et al. 2023, 5; Poeson 2022, 11-12). US companies are responsible for creating and managing the vast majority of AI technologies currently on the market, and US legislatures have consistently opted to prioritize innovation over caution to maintain a competitive edge in

the global AI arms race for themselves and their allies (Garon 2022, 194). This combination of legal gaps, corporate opacity, and geopolitical incentives is exploitable by hybrid agents. Private technology firms operating behind complex management structures and shielded by jurisdictional fragmentation can facilitate – or even directly participate in – hybrid operations without fear of consequence. As long as domestic regulators prioritize innovation and economic advantage over human rights and oversight, and international frameworks remain largely state-centric and unfit to govern collective corporate conduct, these risks will expand.

Transparency in AI-Ecosystems

Without mechanisms in place to ensure democratic oversight, AI can be used to help topple governments, to enable gross war crimes like genocide, and to exert psychological control over masses of people, distorting their sense of reality and exacerbating radicalization. Given the expansive implications of AI-enabled hybrid warfare for democratic institutions, human rights principles, and global security, it is of the utmost importance that experts explore potential frameworks for transparency and accountability at all levels. It is near impossible to assess regulatory needs and nuances if we cannot see and understand what is happening within AI algorithms, or know who is involved in their creation, management, and deployment. This is especially true for such a complex and rapidly-evolving technology within the context of hybrid warfare. Consequently, there is considerable scholarship investigating what transparent AI looks like, why it is important, and how it can be achieved given the challenges inherent to AI traceability and modern corporate governance.

Transparency is most often understood as the disclosure of information to resolve knowledge asymmetries, and it is intended to facilitate trust, accountability, and autonomy by utilizing principles like traceability and verifiability (Felzmann et al. 2020, 3337-3338). However, scholars argue that transparency is a more multifaceted term that can be conceptualized as (1) a virtue, (2) a relation, or (3) a system depending on the context; in other words, transparency can be understood as having intrinsic value, as necessitating that recipients/users understand – not just see – the agents' behavior, and as existing within an institutional context where there are regulatory and organizational mechanisms that impact practical implementation (ibid, 3336). Authors interested in AI transparency claim that integrating all three of these conceptualizations

is necessary when crafting a framework for transparency in AI ecosystems (ibid). As Felzmann et al. (2020, 3341-3342) and Bhatt et al. (2020, 7-8) argue, transparency for AI cannot be reduced to merely the availability of technical documentation or open-source code. Transparency must be tailored to the specific needs, knowledge levels, and contexts of different stakeholder groups – ranging from average citizens to military personnel. This implies that meaningful transparency frameworks must incorporate context-dependent disclosures, traceability mechanisms, and participatory design principles to ensure that oversight is actually achievable in high-risk settings like hybrid warfare.

Applying these principles to AI, algorithmic transparency involves making the algorithms – their logic, the data they use, and known flaws or uncertainties – accessible and understandable to users and stakeholders. This allows for a better understanding of how AI systems make decisions and empowers users to assess the reliability or potential biases of those systems (Chmielinski & Newman et al. 2024; Barredo Arrieta et al. 2020; Bhatt et al. 2020, 1). According to experts focused on ethical AI use, documenting and publicizing AI algorithms serves several purposes on top of the broader benefits of transparency; it is crucial for:

“(1) Supporting responsible development and use, as well as mitigation of downstream harms, by providing transparency into the design, attributes, intended use, and shortcomings of datasets, models, and AI systems; (2) Motivating dataset, model, or AI system creators and curators to reflect on the choices they make; and (3) Facilitating dataset, model, and AI system evaluation and auditing” (Chmielinski & Newman et al. 2024).

In many countries, it is currently optional for tech companies to disclose and explain the algorithmic functioning and training processes of their AI systems. In cases where such disclosures are required, such as with the GDPR in the EU, there is substantial disagreement amongst legal scholars regarding what exactly constitutes requirements for “meaningful” disclosures of information, algorithms, and processes (Felzmann et al. 2020). This lack of specificity leaves a dangerous gap in regulatory landscapes that can be exploited by both state and non-state actors in hybrid warfare operations (Crootof 2022; Virtosu & Goian 2024). Without enforced standards for transparency and interpretability, companies developing powerful AI tools may continue to prioritize profit, secrecy, or national security alignment over public interest, accountability, or ethical concerns – all without thought for consequence due to legal

ambiguity (Burton 2023; Garon 2022; Barredo Arrieta et al. 2020). This opacity obstructs meaningful oversight and erodes democratic norms, leaving decision-makers, citizens, and civil society in the dark about technologies that increasingly influence political processes, social dynamics, and human rights (ibid).

Beyond transparency for the technical and algorithmic aspects of AI, public access to information on AI companies is equally important. The ability to assess a company's governance structure, financial data, contractual agreements, policies, and other basic organizational information is foundational for empowering democratic oversight and pressuring actors into compliance with ethical, safety, and legal norms regarding their AI products (Král & Schnackenberg 2024, 310; Albu & Flyverbom 2016; Brandes & Darai 2017). Research shows that organizational transparency enhances corporate performance, profitability, and competitive advantage; it also strengthens public trust, reduces the risk of misconduct, and improves stakeholder engagement across employees, regulators, and investors (ibid; Ahmed 2021, 100). In this sense, transparency at the organizational level is both a segue to accountability for watchdogs and a strategic asset for companies.

For stakeholders trying to determine whether an AI firm is prioritizing long-term ethical responsibility over short-term investor returns, insights into ownership structures, revenue streams, and investor influences are critical; this information displays the incentives, risks, and power structures that shape a company's behavior far more than its marketing or public-facing messages (Tallarita 2025; Perrigo 2024). This is particularly important in AI ecosystems, wherein market and investor pressures – especially from “superstakeholders” like Big Tech partners (e.g. Google, Amazon) who combine equity influence with operational indispensability – drives rapid deployment of technologies that potentially circumvent adequate risk mitigation tests (ibid). Recognizing this, some AI companies have specifically created their governance model to shield the company's board from conventional investor pressures, prioritizing AI safety and the “long-term benefit of humanity” over profit maximization (Perrigo 2024). Ultimately, transparency regarding venture capital funding, distributions of equity, and public-private partnerships allows journalists, researchers, and regulators to track the influence of powerful backers, including those with geopolitical, military, or surveillance interests.

That said, not all corporate-level transparency is substantively meaningful. Aligned with Felzmann et al.'s (2020) argument that meaningful AI transparency goes beyond basic disclosure, Král & Schnackenberg (2024, 309-312) show that many companies engage in “hollow” or “deceptive” transparency, disclosing company information that is opaque, selectively framed, or strategically performative. In other words, firms may only meet superficial disclosure standards without providing the necessary clarity, explanation, or figures for substantive accountability (Felzmann et al. 2020, 3341-3342; Bhatt et al. 2020, 7-8). The issue of performative transparency is particularly relevant in AI ecosystems, wherein companies release ethical statements, governance diagrams, or model cards that create the illusion of responsibility while concealing critical details about investor control, board decision-making, or risk trade-offs; they may rely on overly-simplistic disclosures, or alternatively, overly-dense and difficult to understand ones. This misleads stakeholders into believing that ethical commitments are embedded in corporate practice when, actually, market pressures and power asymmetries continue to drive short-term opportunistic behavior (Král & Schnackenberg 2024, 310-312; Tallarita 2025).

3.0 Methodology

Overview

Using two case studies, this exploratory research begins to answer the question, “what AI tools and firms are enabling hybrid warfare operations, and how do they meet transparency standards?” Focusing on hybrid operations within the ongoing Israel-Gaza and Russia-Ukraine wars, a total of 65 AI tools and 25 AI companies were identified for analysis, serving as a starting point for wider efforts to document AI involved in hybrid warfare. Importantly, every AI company included in the dataset is responsible for at least one of the 65 AI tools.

To assess the transparency of these AI tools/companies, two scoring systems were created and applied: (1) the CLeAR scoring framework for assessing AI algorithmic transparency, and (2) a newly developed framework tailored to AI tech companies focusing on their transparency, accountability, and legality (i.e. the TAIL framework). Supplementary information – including investor lists, owner names, owner affiliations, and relevant contracts – were also taken note of

in some cases. In doing so, a descriptive document is created that compiles relevant governance, financial, policy, and algorithmic data on each AI tool/company that can be used to identify accountability gaps, expose political or financial conflicts of interest, pinpoint misuse potential, and guide users towards making educated decisions regarding the AI they use.

Companies and their AI systems are given a score to indicate their compliance with the CLeAR and TAIL frameworks, supporting the ability of a range of stakeholders to understand the AI that is shaping transnational conflicts. For clarity, a three-tier system is used: AI tools scoring between 0 and 4 are labeled as having low transparency, scores between 4.5 and 8 indicate that the AI is moderately transparent, and those with scores between 8.5 and 12 are considered highly transparent; this is the case for both the CLeAR and TAIL scoring systems.

While this study is intended to be primarily exploratory and descriptive, one hypothesis is tested using a Pearson correlation coefficient (r). The hypothesis is:

H1: “Higher company transparency is positively correlated to higher algorithmic transparency.”

This hypothesis draws on literature that investigates the driving forces and conditions behind meaningful transparency, such as Felzmann et al. (2020), Albu & Flyverbom (2016), Brandes & Darai (2017), and Král & Schnackenberg (2024). It is reasoned that firms embracing proactive corporate-level transparency could have stronger over-all cultures of accountability, thereby increasing the likelihood of AI product transparency. Testing this hypothesis will contribute to scholars' understanding of the relationship between product-level and corporate-level transparency as two potentially distinct or related dimensions.

Case Studies

The Israel-Gaza and Russia-Ukraine wars serve as particularly relevant case studies for this research focus for at least four reasons: (1) both of these wars are receiving extensive media coverage, (2) the primary actors (Russia, Ukraine, Israel, and Hamas) are all accused of using hybrid warfare methods, (3) these wars involve a myriad of actors and interests across different regions, and (4) they are taking place in the midst of the global AI arms race. In fact, Ukraine has been labeled an “AI War Lab” by multiple journalists due to its openness to receiving private firms looking to test weaponized AI systems in real conditions, and Israel has received consistent

criticism from human rights organization regarding the ethics of the various AI systems they are using against Palestinians in Gaza (Bergengruen 2024; Human Rights Watch 2024). With AI tactics and IW, CogW, and CyW campaigns playing a central role in each conflict, Israel-Gaza and Russia-Ukraine serve as fitting case studies for exploring what private interests are involved in enabling hybrid warfare with AI.

Selection Procedure

This study uses a qualified convenience sampling method to collect 65 AI tools and the corresponding 25 companies for scoring. In order for an AI tool to be included in the dataset, at least two different online sources must have made claims regarding the AI's use or the AI company's involvement in either of the two wars, specifically regarding IW, CogW, or CyW operations. This qualification helps to ensure that the AI included in the dataset is relevant to AI-enabled hybrid warfare occurring in the case studies, and within the scope of the selected hybrid tactics. Other than this qualification, a convenience method was used wherein Google was prompted – using search terms like “Russia-Ukraine / Israel-Gaza + AI + hybrid warfare” – to identify reports of AI use in either conflict, and then AI tools and/or companies were selected for inclusion on a first-come-first-serve basis. This collection process took place between January and May 2025, resulting in the selected 65 AI products and the 25 developers.

CLeAR and TAIL Scoring Frameworks

The CLeAR framework was recently developed by a team of experts involved in private AI development, research, and regulation, with the goal of guiding documentation efforts for private companies and providing regulators with the necessary context for creating effective and meaningful AI policy (Chmielinski & Newman et al. 2024). CLeAR is an acronym standing for Comparable, Legible, Actionable, and Robust. When applied to AI documentation, these principles are intended to create a foundation of transparency in the AI ecosystem that will, in turn, enable the realization of other values (ibid, 4). According to this framework, AI documentation should be:

“Comparable: Able to be compared; having similar components to documentation of other datasets, models, or systems to permit or suggest comparison; enabling comparison by following a discrete, well-defined format in process, content, and presentation.

Legible: Able to be read and understood; clear and accessible for the intended audience.

Actionable: Able to be acted on; having practical value, useful for the intended audience.

Robust: Able to be sustained over time; up to date” [emphasis added] (ibid, 5).

Building on the CLeAR Framework principles as provided by Chmielinski & Newman et al. (2024, 14-16), a scoring system for AI algorithmic transparency was created and applied to the 65 AI tools identified for this study. The CLeAR principles were divided into 12 subvariables that ask specific yes-or-no questions related to the four CLeAR dimensions. Then, a points system was created wherein Yes = 1, Some = 0.5, and No = 0. The higher the score (with the highest possible being 12), the more compliant an AI tool is with the CLeAR principles. Each CLeAR subvariable question was created following the guidance of AI experts focused on transparency and accountability, including Chmielinski & Newman et al. (2024), Gebru et al. (2018), Mitchell et al. (2018), Bhatt et al. (2020), and BIROĞUL et al. (2025), and Felzmann et al. (2020). The CLeAR subvariables and scoring questions are as follows, with footnotes providing specific details on each subvariable as needed:

1. Comparability (C)

C1 – Documentation Format Consistency: Does the AI system follow a standard documentation format that allows easy comparison with other tools?¹

C2 – Availability of Benchmarking Data: Does the AI system include performance benchmarks that allow for external comparison?²

¹ Model Cards, Data Sheets for Datasets, System Cards, Algorithmic Impact Assessments (AIAs), and ISO/IEC standards (e.g., ISO/IEC 42001 for AI Management Systems) are all considered standard documentation formats for the purposes of this paper.

² Benchmark data refers to performance metrics, descriptions of tested datasets and environments, comparisons against industry-standards (e.g. NIST tests), and transparency evaluation protocols.

C3 – Cross-Platform / Cross-Use Comparability: Is the AI tool designed in a way that allows its performance to be compared across different contexts?³

2. Legibility (L)

L1 – Clarity of Intended Use: Is the AI system's intended use clearly stated in its documentation?⁴

L2 – Disclosure of Limitations & Uncertainties: Does the documentation acknowledge limitations, biases, or areas for improvement?⁵

L3 – Accessibility of Documentation: Is the documentation readable and understandable by relevant stakeholders?⁶

3. Actionability (A)

A1 – Availability of Auditing & Evaluation Tools: Does the AI system provide information or tools that allow stakeholders to audit it?⁷

A2 – Guidance for Mitigation of Risks: Does the documentation provide actionable guidance on mitigating identified risks?⁸

A3 – User Control & Override Features: Does the system include control mechanisms for human oversight?⁹

³ This is based on: (1) whether an AI tool can be assessed across different operational contexts (e.g. different countries, datasets, or clients), (2) whether it provides clear and consistent metrics for comparisons across platforms, and (3) whether it has transferability or adaptability of information (e.g. does performance vary depending on factors like camera types, data sources, cultural diversity of faces, etc).

⁴ Determined by whether the documentation publicly specifies who the users are, what the system should/shouldn't be used for, and any application boundaries.

⁵ This includes proactive statements about potential errors, demographic performance variability, or ethical risks. Importantly, L2 intends to evaluate what a company willingly discloses, not what a third party reports.

⁶ Based on whether the documentation avoids excessive technical jargon without compromising clarity, with information easily locatable without legal or technical barriers.

⁷ Auditing tools include Public APIs or sandboxes allowing stakeholders to test/evaluate the tool's behavior, clear instructions or tools to allow third-party audits (e.g., access logs, explainability features, auditing dashboards), and the use of recognized auditable frameworks or external independent audits that are publicly disclosed.

⁸ This includes explicit outlines of potential risks (bias, misuse, misidentification, privacy violations) and clear, actionable guidance on how users should mitigate these risks (e.g., policies for verification, double-checking outputs, safe deployment practices). These limitations should be included in user materials, not just fine print legal disclaimers.

⁹ Control mechanisms include user control settings, review/verification steps, override options, and explicit policies on human oversight responsibilities (e.g., manual confirmation before actioning system outputs).

4. Robustness (R)

R1 – Update Frequency & Versioning: Is documentation consistently updated to reflect AI modifications?¹⁰

R2 – Long-Term Maintenance Plan: Does the AI provider outline a clear maintenance and update strategy?¹¹

R3 – Data & Model Lifecycle Transparency): Does the documentation track how data and models evolve over time?¹²

Whereas the CLeAR Framework is primarily focused on assessing algorithmic transparency of AI tools, the TAIL framework is proposed to assess the transparency of the company responsible for the AI. TAIL is an acronym standing for Transparency, Accountability, and Legality, with each of these principles broken down into 12 subvariables with yes-or-no questions for scoring. The TAIL Framework also uses a 0-1 scale wherein 0 = No, 0.5 = Some, and 1 = Yes,^{*13} with a higher score indicating greater compliance with TAIL principles and thereby higher levels of transparency. The following TAIL subvariable questions were created based on the works of scholars focused on corporate transparency like Král & Schnackenberg (2024), Albu & Flyverbom (2016), Shen (2024), Ahmed et al. (2024), and Brandes & Darai (2017):

1. Transparency

T1 - Disclosure of Contracts: Are contracts with government agencies and/or militaries willingly disclosed on the developers' website?

¹⁰ This manifests in the form of clear, public version history showing algorithm updates, new feature rollouts, bug fixes, when updates occurred, what changes were made, and why the changes matter (impact on performance, risks).

¹¹ This is determined by the presence of explicit commitments to ongoing maintenance, support timelines (e.g., version support periods, end-of-life policies), clear governance on responsibility for system upkeep, and the public disclosure of planned improvements, bug reporting mechanisms, and model improvements schedules.

¹² This includes data sourcing and updates (where/how new data is acquired), model retraining schedules, policies for dataset curation, pruning, or retirement, and transparency on how datasets evolve over time and what criteria govern changes.

¹³ *For subvariable L1, which asks, “has the developer faced lawsuits for noncompliance with national law, international law, or international treaties?” the scoring is reversed. 1 = No, 0.5 = Partial, and Yes = 0. This exception is to ensure that, overall, higher scores reflect better compliance with the TAIL principles. Since being involved in a noncompliance lawsuit is a negative trait for a company, Yes must = 0, and No must = 1.

T2 - Disclosure of Partnerships: Are partnerships with other private entities, organizations, or governments, willingly disclosed on the developers' website?

T3 - Update Forum: Is there a blog or news section on the website where the company publishes relevant company updates for the public?

T4 - Publication of Transparency/Responsibility Reports: Does the developer publish annual transparency, responsibility, or impact reports (similar to ESG reports) disclosing practices, commitments, and operations? (Grounded in: Corporate Social Responsibility (CSR), ESG reporting literature)

T5 - Public Ethical AI Principles/Policies: Does the developer publicly disclose ethical AI principles, guidelines, or responsible AI policies? (Grounded in: IEEE Ethically Aligned Design, OECD AI Principles)

T6 - Publication of Financials / SEC Filings

2. Accountability

A1 - Governance and Oversight Structures Disclosed: Does the company disclose board members, advisory boards, or oversight committees relevant to ethics, compliance, or governance?

A2 - Open Investors: Are the lead investors published willingly on the developers' website?

A3 - Mechanisms for Public Grievance or Whistleblowing: Does the developer have clear, publicly accessible channels for whistleblowing, grievance submission, or ethical concerns? (Grounded in: OECD Guidelines, UNGP Reporting Framework)

3. Legal Compliance

L1 - Lawsuits: Has the developer faced lawsuits for noncompliance with national / international law or international treaties?

L2 - Public Disclosure of Regulatory Filings/Certifications: Does the company disclose certifications or compliance reports with key national/international standards (e.g., ISO, GDPR compliance statements)?

L3 - Membership in Regulatory or Ethical AI Initiatives: Is the developer a signatory/member of recognized regulatory, human rights, or ethical AI initiatives (e.g., Partnership on AI, UN Global Compact)?

In order to score each AI tool and its responsible company, three source types were used to gather data: (1) the AI tool's official website, (2) secondary sources including media reports and investigative journalism (e.g. from TechCrunch, ProPublica, Bellingcat) or NGO and think-tank reports (e.g from EU DisinfoLab, Human Rights Watch, RAND, etc.), and (3) CrunchBase, a third-party platform that provides financial and ownership information on various companies for a fee and is generally considered reliable. Using these three sources, the information gathered on the AI and its company can be cross referenced, allowing for any discrepancies to be taken note of.

Limitations

Regarding the sample, this is a study focused on 65 AI tools managed by 25 AI companies that were not randomly selected, and the results are therefore not generalizable. This small-scale study is not meant to serve as an exhaustive list of AI involved in hybrid warfare or in the two case studies; rather, it is intended to serve as a starting point for broader documentation efforts of AI-enabling hybrid operations, providing analytical findings that are relevant to these specific AI tools/companies. Additionally, there were practical difficulties differentiating between distinct AI algorithms as they were listed on the AI company's websites. Some websites listed specific and patented AI products that were clearly distinct from each other, making it easy to differentiate between the AI tools they offer. However, others only provided a list of vague services "enabled by AI," making it difficult to know if these were dependent on multiple distinct algorithms or a single one. The researcher used their best judgement, supported by their knowledge of AI algorithms, to determine which AI services were distinct in cases of ambiguity.

Another caveat to this study is that the transparency scores for each AI tool and company are defined by the availability of information at the time of data collection, and are constrained by the limited time and resources of the researcher; a comprehensive investigation of every publicly available online source that may contain relevant information for scoring each AI tool/company was not practical. Rather, data collection primarily relied on publicly available information from the AI company's official website, Crunchbase, and some secondary sources such as news articles or NGO reports. Thus, when labeling these AI tools and companies as highly transparent, somewhat transparent, or not transparent, it is essential to remember these labels are made solely within the context of this paper's specific definitions, scope, and limitations.

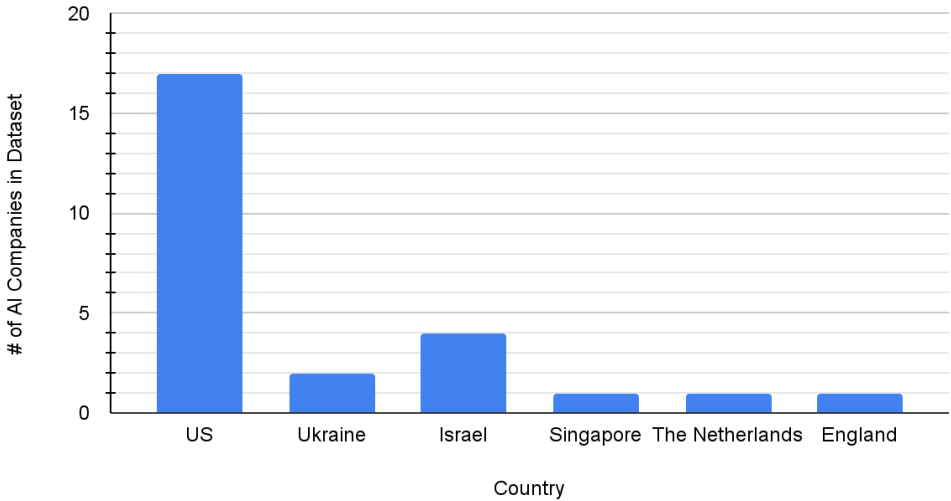
It is also important to note that while the scoring system and subvariable questions used in this study were created based on broader theoretical principles and scholarly literature, they have not been empirically tested or scrutinized outside of this paper. The implications of this limitation include, for example, that while the two scoring frameworks currently assume that every subvariable is equally important for transparency, it could be the case that some are actually more important than others, and therefore they should be given greater weight. Future research should investigate and test the efficacy of the proposed subvariables from the CLeAR and TAIL frameworks; they should also explore how alterations to the language or weighting schemes may change the scoring outcomes.

4.0 Results & Analysis

The majority of AI tools and companies examined in this study are primarily involved in the Russia-Ukraine war. Specifically, 18 of the 25 AI companies have tools that are connected to Russia-Ukraine, 11 are associated with Israel-Gaza, and six are involved with both conflicts. Shown in Figure 1, 65% (n=17) of the 25 AI developers have their headquarter location(s) within the US, 8% (n=2) in Ukraine, 16% (n=4) in Israel, and 4% (n=1) in Singapore, the Netherlands, England, and Mexico (some have multiple headquarter locations). Taken together, these insights may support previously observed trends that identify vast technological divides between more affluent nations (Russia, Ukraine, Israel) and poorer, non-Western ones (Palestine).

Figure 1

AI Company Headquarter Locations Distribution



While it would be helpful to visualize and sort the AI from the dataset based on how it is being used and in what contexts, given the blended and overlapping nature of IW, CogW, and CyW tactics, it is difficult to definitively separate the AI tools into one category and not another. Additionally, given that many of these AI tools are both vaguely described and dual-use, the developers often only advertise the basic intended use of the AI, not all of its potential uses; this makes it difficult to determine how the AI could be used in hybrid contexts, further hindering specific categorization. That said, the following word cloud highlights some of the more common functions, features, and intended uses of the AI tools in the dataset:

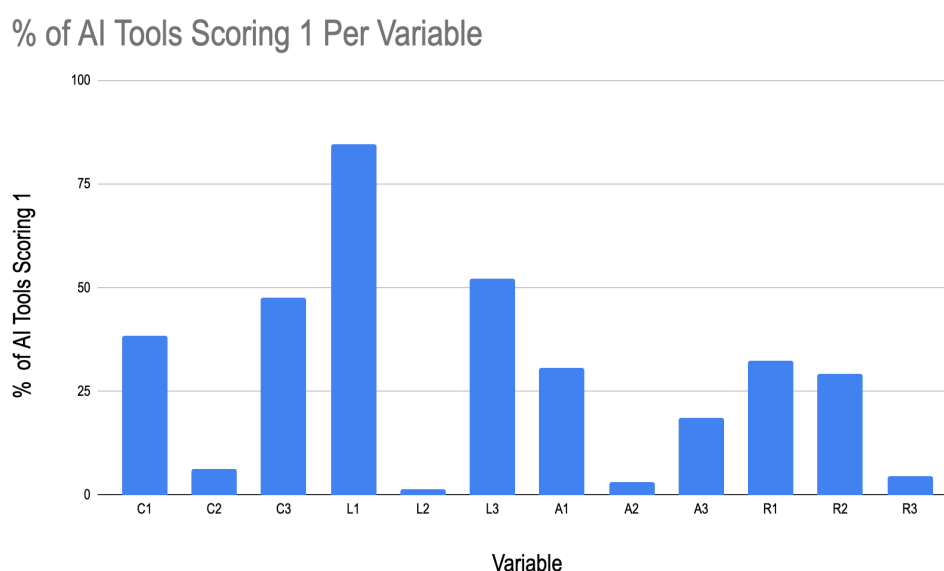


CLeAR Scores

Nearly half (49.2%) of the 65 AI tools included in the dataset scored between a 0 and 4, indicating low algorithmic transparency and a lack of verifiable compliance with the CLeAR framework's principles; 29.2% are considered moderately transparent, and the remaining 21.5% are considered highly transparent. Importantly, of the 14 AI tools scoring highly transparent, all are owned by just four different companies, and six of the tools are owned by one company. Figure 2 visualizes the distribution of overall CLeAR scores for each AI tool:

and buying incentives: clearly stating intended use helps market the product, frames it as a solution to a specific threat, and reduces ambiguity in customer perception (Schnackenberg & Tomlinson 2016; Albu & Flyverbom 2024). However, this means it is possible that the overwhelming compliance in the dataset with the L1 principle is more of a branding exercise than it is an ethical safeguard.

Figure 3



Also shown in Figure 3, AI tools were least likely to fully comply with the L2 variable, which asks “Does the documentation acknowledge limitations, biases, or areas for improvement?” Only one AI tool, Primer Technologies’ “Primer Command,” scored a 1 (yes) for L2. This gap is profoundly important. Transparency about algorithmic limitations is both a widely accepted best practice in ethical AI governance, and it is critical for realizing the following benefits:

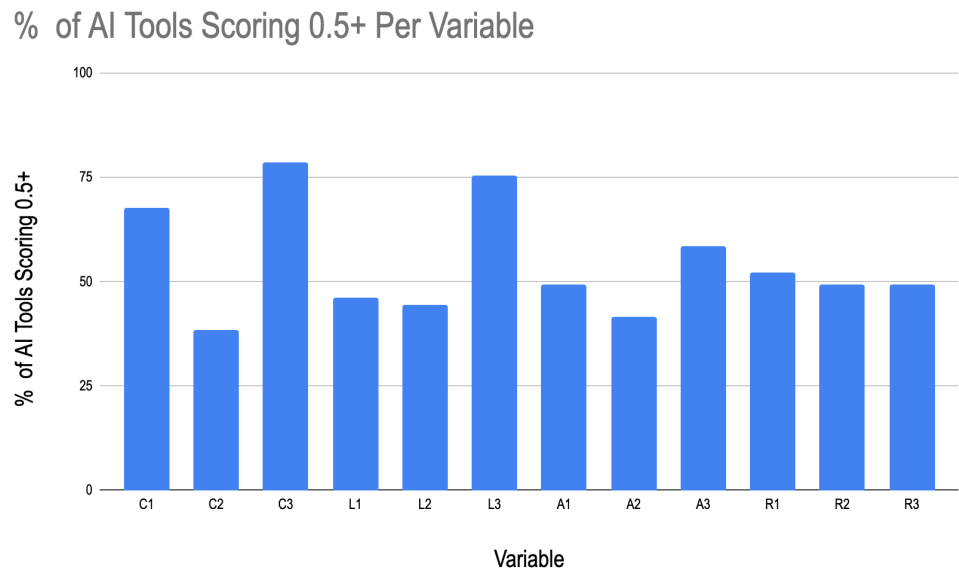
- Operational integrity, by explaining to human users how and when to trust the system outputs;
- Legal accountability, by identifying where errors or biases emerge in targeting, surveillance, or classification processes; and
- Human rights compliance, especially when enabling those hybrid warfare tactics involving things like attack prediction, fact-checking genocide claims, or facial

recognition for target identification (Chmielinski et al. 2024; ICRC 2021, 486-469; Barredo Arrieta et al. 2020; Bhatt et al. 2020, 1).

The near-complete absence of L2 disclosures suggests a potentially widespread aversion to acknowledging uncertainty. This could be due to concerns over reputational risk, legal liability, or competitive advantage (Ho et al. 2023, 5). This avoidance is particularly dangerous in the context of hybrid warfare, where AI systems are used in environments of high uncertainty, ambiguous legality, and civilian proximity (Virtosu & Goian 2024, 198; Felzmann et al. 2020, 3337-38; Crootof 2022, 1). Slightly better than L2, only two AI tools scored a 1 (yes) for the A2 variable, which asks “Does the AI system provide information or tools that allow stakeholders to audit it?” As with the alarmingly low L2 scores, this unwillingness to facilitate third-party audits creates significant risk. Without the means to audit an AI tool, it is near impossible to assess the AI’s operational integrity, human rights compliance, or regulatory compliance in hybrid warfare contexts.

Shown in Figure 4 below, the variables most likely to receive at least partial compliance (0.5) include C3, L3, and C1. The tendency to partially comply with C3 (comparable design) and L3 (readability for stakeholders) likely are a result of initial partial compliance with C1 (use of standardized documentation format). While many of the AI tools did have some form of product description in their website, these were often surface-level and full of buzz-words, missing proper explanations. Their compliance with these subvariables was likely not intended for meaningful transparency, but rather for marketing. Without comprehensive, standardized documentation (C1), the likelihood of scoring over 0.5 on C3 or L3 is near impossible. The tendency of AI tools to partially comply with these three subvariables prevents external scrutiny and reduces the ability of users, researchers, or regulators to meaningfully compare one system to another (Chmielinski & Newman et al. 2024).

Figure 4



Interestingly, while AI tools were unlikely to receive a score of 1 (yes) for variables C2 (benchmarking), L2 (limitations disclosure), A2 (auditability), A3 (user control), and R3 (lifecycle transparency), Figure 4 indicates that a higher percentage of AI tools were at least partially complying (0.5) with these CLeAR subvariables. This suggests that developers are often initiating transparency measures in these categories, but falling short of meeting the full standards required for meaningful compliance. This could suggest a form of nascent or performative engagement, wherein AI firms acknowledge the importance of these transparency practices but still only incorporate the bare-minimum. Alternatively, partial compliance could reflect resource constraints or a lack of standardized infrastructure for documentation, especially for small or mid-size companies. Future research should explore what internal or external factors are most influential for encouraging or discouraging full compliance with these CLeAR principles.

TAiL Scores

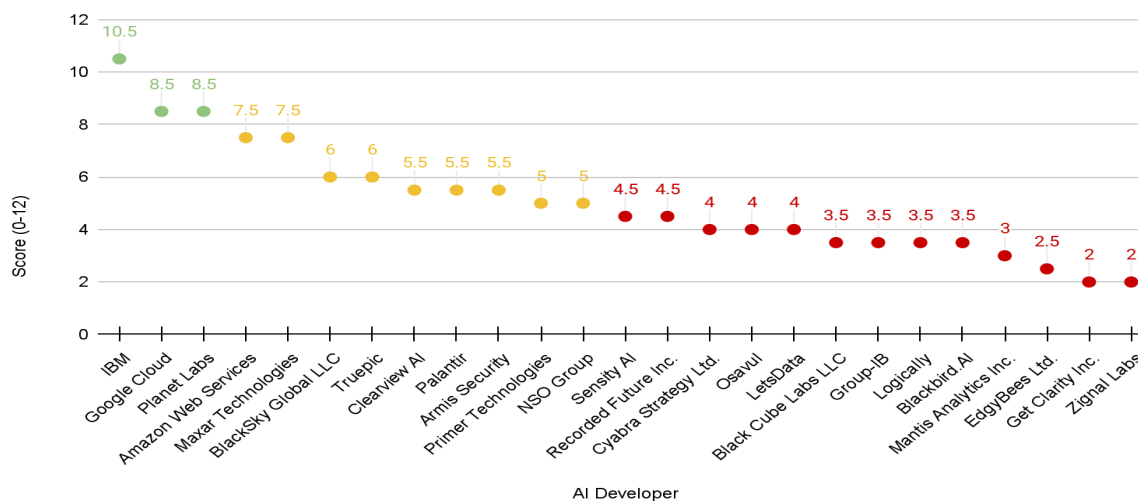
As depicted in Figure 5, over half (52%) of the 25 AI developers received an overall TAiL score between 0 and 4, indicating low transparency and inadequate compliance with TAiL principles. 36% of developers received a score indicating moderate levels of transparency, whereas only

12% earned scores indicating high levels of transparency and compliance. Similarly to the CLeAR scores, the TAIL scores indicating high corporate transparency are concentrated amongst a few companies; opacity is a more common practice for these tech firms involved in hybrid warfare. A concerning implication of this trend is that civil society, international organizations, and the general public are left unable to assess:

- Whether these tools comply with international humanitarian law (IHL), International Criminal Law (ICL), Private International Law (PrIL), or any other potentially relevant regulations related to data privacy, human rights, or warfare;
- How risks like algorithmic bias, collateral damage, or misuse are (or are not) being mitigated; and
- Who is responsible when AI-enabled tools contribute to human rights abuses, distortions of reality, or faulty target identification.

Figure 5

Overall TAIL Scores Per AI Developer



Moreover, the large number of AI tools owned by some companies shows the potential for an AI firm to flood the operational field or markets with tools – garnering significant influence over policy and norms related to their operations – without adhering to principles of transparency, accountability, or legality. For instance, Osavul – a Ukrainian firm with the most AI tool products (n=8) of any company included in this study – scored a 4 out of 12 for the TAIL framework,

indicating low corporate transparency. Compared to a company offering only one or two AI tools for hybrid warfare, Osavul could be considered more involved in hybrid operations, making their lack of transparency more impactful in hybrid warfare scenarios.

Disaggregating the data, Figure 6 shows that AI companies were most likely to score a 1 (yes) on the T3 variable, which asks “Is there a blog or news section on the website where the company publishes relevant company updates for the public?” However, they were least likely to score a fully compliant score (1) on the T1 (disclosure of contracts) and A2 variables (disclosure of investors). This disparity between T3, T1 and A2 variables is interesting because it may support the observation that companies will often engage in symbolic transparency (T3, news blogs) rather than substantive transparency (T1, disclosure of contracts and A2, investors). This trend is perhaps especially true in sensitive or militarized sectors. Furthermore, comparing Figure 6 with Figure 7, while AI companies were not as likely to fully comply with the T1 or T2 variables, they were more likely to at least partially comply.

Figure 6

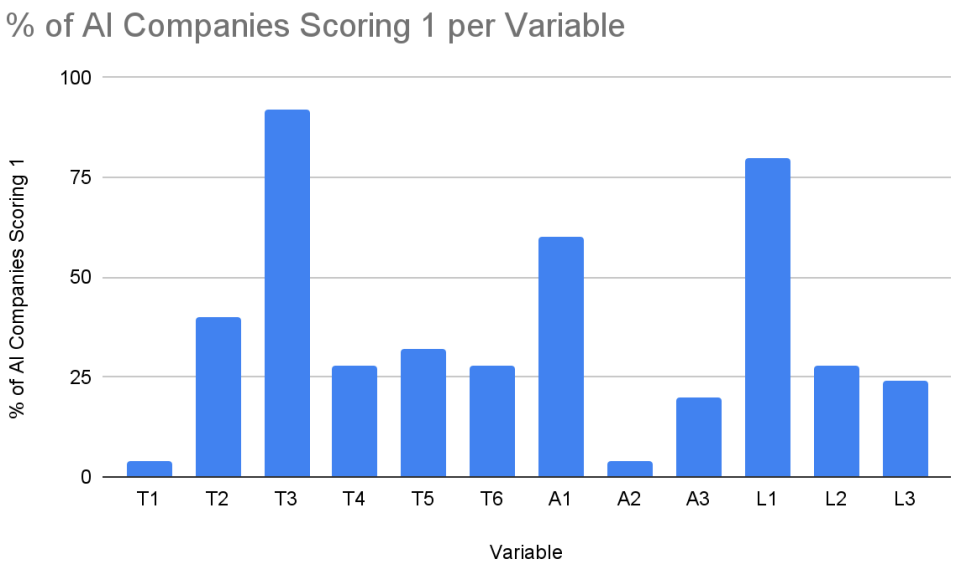
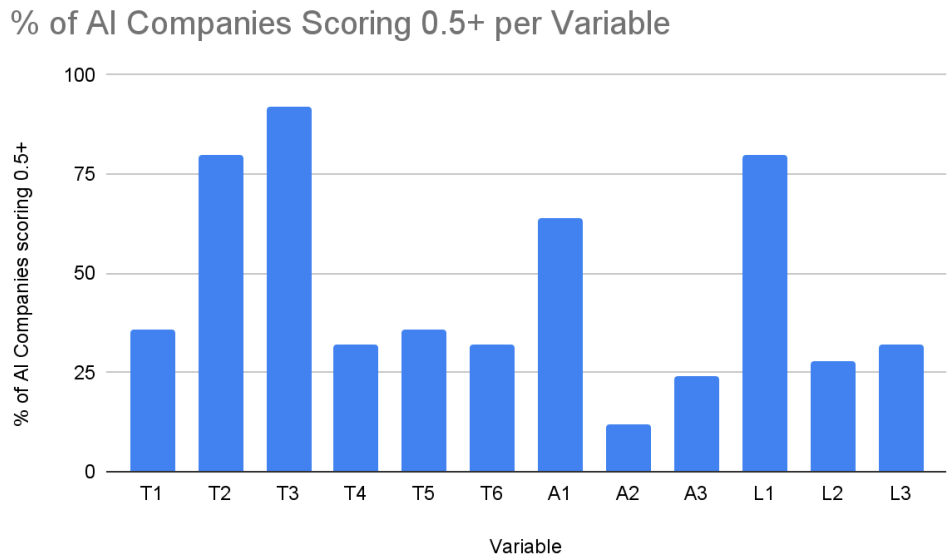


Figure 7



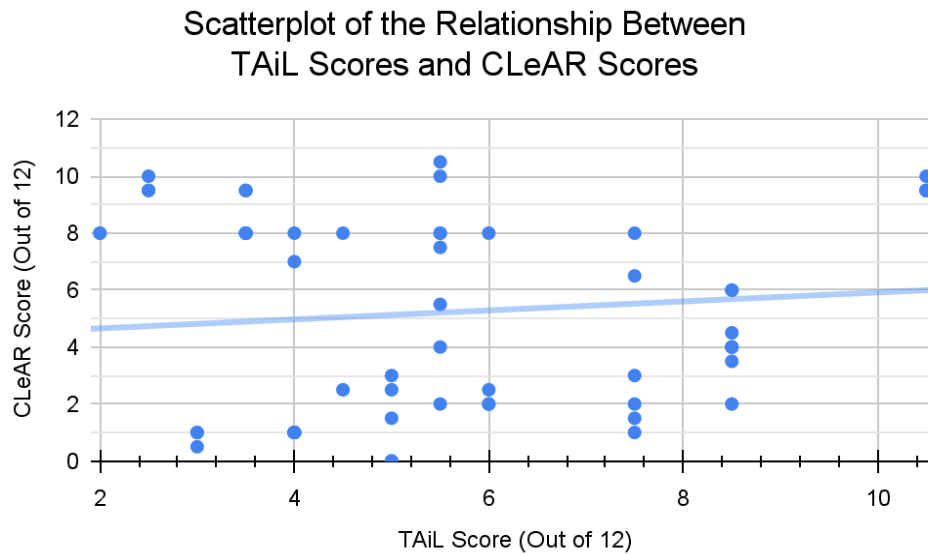
Importantly, the L1 variable asks, “Has the developer faced lawsuits for noncompliance with national/international law or international treaties?” and a score of 1 indicates “no” for this variable, unlike the rest – that being the case, 80% of the AI companies from the dataset were *not* found to be involved in relevant lawsuits at the time of data collection. Five companies, however, were identified as having been involved in lawsuits alleging potentially relevant violations of international or national regulations regarding their AI:

1. Google Cloud (Alphabet Inc.)
2. Amazon Web Services (Amazon.com Inc.)
3. Clearview AI
4. Palantir
5. NSO Group

While the outcomes of the cases were not considered in scoring, this is because most eventually settled outside of court, making it difficult to assess whether a violation actually occurred. However, lawsuits alleging issues like data privacy violations or unethical conduct nonetheless indicate a poor history of transparency, thus negatively impacting their score.

Synthesizing CLeAR and TAIL Scores

Figure 8



To assess whether there is a relationship between company-level transparency (TAIL) and algorithmic transparency (CLeAR), a Pearson correlation coefficient (r) was calculated as $r=0.109$, indicating that there is a very weak positive correlation between TAIL and CLeAR scores, but not enough for it to be considered significant. Therefore, H1 is not supported, and TAIL scores measuring company transparency are not *strongly* correlated with CLeAR algorithmic transparency scores. So, for example, an AI company in this dataset with greater company-level transparency is not necessarily more likely to have higher algorithmic transparency for their individual AI products too. Such discrepancies between TAIL and CLeAR scores suggest that organizational commitments to transparency and accountability do not necessarily translate into transparent design, documentation, or oversight in individual AI tools. This highlights the importance of assessing both the transparency within corporate practices and individual AI products, suggesting that policy-makers should consider both dimensions when attempting to address accountability gaps.

Case Highlight: Palantir

Palantir is an American AI company involved in both Russia-Ukraine and Israel-Gaza that specializes in facial recognition and predictive policing services, among other AI powered services. Regarding their role in Israel-Gaza, external sources – and even Palantir’s own employees – have alleged that the company’s work with Israel amounts to complicity in gross war crimes against Palestinians (Business & Human Rights Resource Centre, 2025). Regarding Palantir’s AI aiding Ukrainian forces, Palantir, “... has provided its AI software to analyze how the war has been unfolding, to understand troop movements and conduct battlefield damage assessments” (Fontes & Kamminga 2023). Palantir CEO Alex Karp suggests that Ukraine’s use of Palantir AI is enough to level the playing field between Russia and Ukraine, wherein Ukraine is perceived as the weaker force (Kosoy 2025). If true, this sentiment supports the notion that AI technologies have the power to alter outcomes of international conflicts, regardless of hard military powers – and if not true, it is still clear that this is the end-goal of some AI tech companies, to make AI so powerful that it can tip the scales of global conflict.

Palantir earned a 5.5 TAIL score (see Figure 9), indicating moderate levels of corporate transparency. Of their five AI products, four earned CLeAR scores within the 4.5 to 8 range indicating moderate algorithmic transparency, while one (Skykit) scored in the low transparency range (see Figure 10). Overall, Palantir exhibits some transparency – but often in the more artificial sense. For example, while Palantir does publish a news blog (T3) and advertise their AI’s intended use (L1), they do not disclose their board members (A1) or their investors (A2) on their website. They also only disclose select contracts, excluding key government and military arrangements (T1).

Figure 9. TAIL Subvariable Score Breakdown for Palantir

	TAIL Subvariables													Transparency Level
	T1	T2	T3	T4	T5	T6	A1	A2	A3	L1	L2	L3	Total	
Palantir	0.5	1	1	0	1	1	0	0	1	0	0	0	5.5	Medium

Figure 10. CLeAR Subvariable Score Breakdown of Palantir' AI Tools

		CLeAR Subvariables												Total	Transparency Level
		C1	C2	C3	L1	L2	L3	A1	A2	A3	R1	R2	R3		
Palantir AI Systems	Gotham	0.5	0	0.5	1	0.5	0.5	0.5	0.5	1	1	1	1	8	Medium
	Foundry	0.5	0	0.5	1	0	0.5	1	0	1	1	1	1	7.5	Medium
	AIP	0.5	0.5	0.5	1	0	0.5	1	0	1	1	1	1	8	Medium
	Skykit	0.5	0	0.5	1	0	0.5	0	0	1	0.5	0	0	4	Low
	MetaConstellation	0.5	0	0.5	1	0	0.5	0	0	1	1	0.5	0.5	5.5	Medium

According to CrunchBase, Palantir has received a total of \$3 billion in investments from a variety of sources. One such source is Founders Fund, an investment firm with roughly \$17 billion in total assets. Interestingly, Peter Thiel is both the cofounder of Palantir and a managing partner at Founders Fund. Founders Fund is also a listed investor in a second AI company from the dataset, Cyabra Strategy Ltd., which has a single AI tool associated with the Israel-Gaza war. Beyond his role in investments, Thiel is also infamous for his political lobbying in the US – his donations have been found, “...in the coffers of the Republican National Committee, House Republicans' main campaign arm, over a dozen state parties, and nearly 30 GOP House members” (Metzger 2025). Most recently, Thiel gave \$852,000 to House Speaker Mike Johnson’s (R) fundraising committee, and then about 90% of this money was distributed to “other campaigns” (ibid). According to OpenSecrets, CEO Alex Karp has also made political contributions in the tens-of-thousands to both Republican and Democratic committees and candidates. Regarding their contracts, reports suggest that Palantir has arrangements with the US Army, US Special Operations Command, and the US Department of Defense, among others; they also have contracts with other AI service providers, including IBM, Amazon Web Services, and Google – all of which are featured in this study’s dataset (Harper 2024; IBM and Palantir Technologies Inc. 2024; Kinder and O’Murchu 2025; IBM 2025; Palmer 2025).

Palantir as a case highlight demonstrates the growing role of private AI companies in modern warfare and politics more broadly. Ethically and geopolitically, these findings highlight how AI can shift the balance of power in conflicts and exacerbate harms to vulnerable populations without traditional, kinetic warfare tactics. Moreover, Palantir holding government contracts during active conflict exposes the risk of creating a profit incentive wherein AI companies are encouraged to prioritize deploying their AI products in military scenarios over ethical, safety, or human rights concerns. These contracts also funnel public money into private hands, meaning

that U.S. taxpayer money may indirectly support war crimes. Finally, Palantir's political connections raise alarming questions about how lobbying may help secure government contracts while reducing regulatory and legal scrutiny. It is this fusion of money, military applications, and political access that underscores concerns about how private AI firms can profit from war while externalizing legal and moral consequences.

5.0 Conclusion & Implications

Ultimately, this study offers a starting-point for documenting and understanding AI's role in hybrid warfare, for identifying accountability gaps, and for finding leverage points for future regulatory interventions. Through a systematic analysis of 65 AI tools and 25 AI companies using the CLeAR and TAIL frameworks, this research exposes that both algorithmic systems and their corporate owners involved in hybrid warfare overwhelmingly fail to meet baseline standards for meaningful transparency and accountability. With almost half of the AI tools from the dataset scoring in the lowest of the three tiers for algorithmic transparency, and with over half of the AI firms scoring similarly low in corporate transparency, there is a sizable gap that creates opportunities for malicious interference, legal evasion, human rights violations, and ethical ambiguity. Companies were most likely to provide marketing-friendly information, such as statements of intended use or product descriptions, while avoiding more meaningful forms of transparency like disclosing algorithmic limitations, enabling third-party audits, or investor lists.

The overall opacity of the AI and AI firms in the dataset indicate that civil society, regulators, and other stakeholders are unequipped to identify cases of misuse, assess compliance with humanitarian or ethical norms, or trace responsibility for harm for legal accountability. Literature suggests that this opacity is likely a symptom of a larger trend in militarized sectors where profit incentives, rapid innovation cycles, and geopolitical competition supersede ethical values or human rights concerns. This means that hybrid agents are incentivized to exploit AI's dual-use and black-box nature to pursue strategic objectives with plausible deniability. Without enforceable international norms, standardized documentation frameworks, or independent auditing, the use of AI in hybrid warfare will continue to erode democratic oversight, challenge the foundations of international humanitarian law, and normalize accountability gaps in armed conflict.

By mapping these actors and quantifying their transparency gaps, this thesis lays a foundation for further inquiry into the ethics, regulation, and accountability of AI in war. Future research should add to the dataset to include other conflicts and more AI companies, scrutinize and refine the CLeAR and TAIL scoring systems, or explore the effectiveness of recent international regulatory initiatives like the UN's AI governance efforts. Future research could also explore how transparency incentives, corporate governance reforms, and market pressures encourage AI firms to internalize ethical responsibility rather than externalizing risk to states and civilians. Such research will be necessary for creating evidence-based and effective safeguards against ethical and legal violations perpetuated by private AI tools involved in hybrid warfare.

The full dataset with scoring breakdowns can be accessed at:

https://docs.google.com/spreadsheets/d/1xGeOSKtf-yhNQj93IZ8x_QRO_WiwlUIErHkE0lrekfo/edit?usp=sharing

References

- Ahmed, Dr. Eissa Reis et al. 2021. "Empirical Analysis on Corporate Transparency, Competitive Advantage, and Performance: An Insight of Muscat Securities Market." *Journal of Governance and Integrity*, 4(2). <https://journal.ump.edu.my/jgi/article/view/6091>
- Albu, O. B., & Flyverbom, M. 2016. "Organizational Transparency: Conceptualizations, Conditions, and Consequences." *Business & Society*, 58(2): 268-297. <https://doi.org/10.1177/0007650316659851>
- Azani, Eitan. 2013. "The Hybrid Terrorist Organization: Hezbollah as a Case Study." *Studies in Conflict & Terrorism*, 36(11): 899–916. <https://doi.org/10.1080/1057610X.2013.832113>
- Barredo Arrieta, Alejandro et al. 2020. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion*, 58: 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bergengruen, Vera. 2024. "How Tech Giants Turned Ukraine Into an AI War Lab." *Time*, February 8, 2024. <https://time.com/6691662/ai-ukraine-war-palantir/>
- Bernstein, Michael S. et al. 2023. "Embedding Societal Values into Social Media Algorithms." *Journal of Online Trust and Safety*, 2(1): 1–13. <https://hci.stanford.edu/publications/2023/jots.pdf>
- Bhatt, Umang et al. 2020. "Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty." *Cornell University, AAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) 2021*, <https://doi.org/10.48550/arXiv.2011.07586>
- Binns, Reuben. 2021. "Fairness in Machine Learning: Lessons from Political Philosophy." *Proceedings of Machine Learning Research*, 81: 1–11 <https://arxiv.org/pdf/1712.03586>
- BİROĞUL, Serdar et al. 2025. "Exploring the Impact of ISO/IEC 42001:2023 AI Management Standard on Organizational Practices." *Artificial Intelligence Research*, 5(1): 14-22. <https://doi.org/10.54569/air.1709628>

Brandes, Leif and Donja Darai. 2017. "The value and motivating mechanism of transparency in organizations." *European Economic Review*, 98: 189-198.

<https://doi.org/10.1016/j.euroecorev.2017.06.014>

Bullock, Justin. 2025. "Towards a Global Agency for Governing Artificial General Intelligence: Challenges and Perspectives" Global Governance Institute, Jan. 14, 2025.

<https://www.globalgovernance.eu/publications/towards-a-global-agency-for-governing-artificial-general-intelligence-challenges-and-perspectives>

Burton, Joe. 2023. "Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence." *Technology in Society*, 75.

<https://doi.org/10.1016/j.techsoc.2023.102262>

Business & Human Rights Resource Centre. 2025. "Big Tech companies face allegations of war crimes complicity amid Israel's war in Gaza." *Business & Human Rights Resource Centre*, May 5, 2025,

<https://www.business-humanrights.org/en/latest-news/big-tech-companies-allegedly-complicit-in-war-crimes-amid-israels-war-in-gaza-incl-company-responses/>

Chmielinski, Kasia and Sarah Newman et al. 2024. "The CLeAR Documentation Framework for AI Transparency: Recommendations for Practitioners & Context for Policymakers." Harvard Kennedy School Shorenstein Center on Media, Politics, and Public Policy,

<https://shorensteincenter.org/clear-documentation-framework-ai-transparency-recommendations-practitioners-context-policymakers/>

Crockford, Kade. 2020. "How is Face Recognition Surveillance Technology Racist?" the American Civil Liberties Union, June 16, 2020.

<https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist>

Crootof, Rebecca. 2016. "War Torts: Accountability for Autonomous Weapons." *University of Pennsylvania Law Review*, 164(6): 1347-1402.

https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9528&context=penn_law_review&httpsredir=1&referer=

Crootof, Rebecca. 2022. “AI and the Actual IHL Accountability Gap.” In *The Ethics of Automated Warfare and Artificial Intelligence*, Centre for International Governance Innovation. <https://ssrn.com/abstract=4289005>

Felzmann, H. et al. 2020. “Towards Transparency by Design for Artificial Intelligence.” *Sci Eng Ethics*, 26: 3333–3361. <https://doi.org/10.1007/s11948-020-00276-4>

Fontes, Robin and Dr. Jorrit Kamminga. 2023. “Ukraine A Living Lab for AI Warfare.” *National Defense Magazine*, March 24, 2023. <https://www.nationaldefensemagazine.org/articles/2023/3/24/ukraine-a-living-lab-for-ai-warfare>

Garon, Jon M. 2022. “When AI Goes to War: Corporate Accountability for virtual mass disinformation, algorithmic atrocities, and synthetic propaganda.” *Northern Kentucky Law Review*, 49(2): 181-234. <https://heinonline.org/HOL/Page?handle=hein.journals/nkenlr49&id=197&collection=journals&index=>

Gebru, Timnit et al. 2018. “Datasheets for Datasets.” Cornell University, <https://doi.org/10.48550/arXiv.1803.09010>

Harper, Jon. 2024. “Palantir lands \$480M Army contracts for Maven artificial intelligence tech.” *Defense Scoop*, May 29, 2024. <https://defensescoop.com/2024/05/29/palantir-480-million-army-contract-maven-smart-system-artificial-intelligence/>

Heaven, D. 2019. “Why deep-learning AIs are so easy to fool.” *Nature*, 574(7777): 163–166. <https://doi.org/10.1038/d41586-019-03013-5>

Hughes, David. 2021. “Differentiating the Corporation: Accountability and International Humanitarian Law.” *Michigan Journal of International Law*, 42(1): 47-104. <https://doi.org/10.36642/mjil.42.1.differentiating>

Human Rights Watch. 2024. “Questions and Answers: Israeli Military’s Use of Digital Tools in Gaza.” Human Rights Watch, Sep. 10, 2024.

<https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza>

Hung, Tzu-Chieh and Tzu-Wei Hung. 2022. “How China's Cognitive Warfare Works: A Frontline Perspective of Taiwan's Anti-Disinformation Wars.” *Journal of Global Security Studies*, 7(4): 1–18 <https://doi.org/10.1093/jogss/ogac016>

IBM and Palantir Technologies. 2021. “IBM and Palantir Partner to Help Businesses Easily Deploy Powerful and Open AI Applications; Unlock Valuable Data Across Hybrid Cloud Environments.” *IBM Newsroom*, Feb. 8, 2021.
<https://newsroom.ibm.com/2021-02-08-IBM-and-Palantir-Partner-to-Help-Businesses-Easily-Deploy-Powerful-and-Open-AI-Applications-Unlock-Valuable-Data-Across-Hybrid-Cloud-Environments>

International Committee of the Red Cross (ICRC), *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, Adopted June 8 1977; entered into force Dec. 7 1978. Accessed May 1, 2025. <https://www.refworld.org/legal/agreements/icrc/1977/en/104942>

International Committee of the Red Cross (ICRC). 2021. “ICRC Position Paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach.” *IRRC No. 913*, March 2021.
<https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913#:~:text=For%20conflict%20parties%2C%20human%20control,civilians%20can%20be%20further%20minimiz>

Jokinen, Janne et al. 2022. “Hybrid Threats from Non-State Actors: A Taxonomy.” *Hybrid CoE Research Report no. 6*. Helsinki: European Centre of Excellence for Countering Hybrid Threats. Accessed May 1, 2025.
<https://www.hybridcoe.fi/publications/hybrid-coe-research-report-6-hybrid-threats-from-non-state-actors-a-taxonomy/>

Kinder and O'Murchu. 2025. "Palantir's 'revolving door' with government spurs huge growth." *Financial Times*, Feb. 27, 2025.

<https://www.ft.com/content/a65c93af-abf9-427c-a2ee-a9f5b3d518b9>

Köbis, Nils Christopher et al. 2022. "The Corruption Risks of Artificial Intelligence." *Transparency International*, <http://www.jstor.org/stable/resrep43028>.

Kosoy, Daniel. 2025. "Palantir, the Secretive Tech Giant Shaping Ukraine's War Effort." *United24 Media*, Jan. 31, 2025.

<https://united24media.com/war-in-ukraine/palantir-the-secretive-tech-giant-shaping-ukraines-war-effort-5519>

Král, Pavel and Andrew Schnackenberg. 2024. "Organizational responses to transparency determinants." *Management Decision* 16, 62(13): 309–331.

<https://doi.org/10.1108/MD-07-2023-1244>

Lopez, C. Todd. 2021. "DoD: August 29 Strike in Kabul 'Tragic Mistake,' Kills 10 Civilians." *DOD News*, September 17, 2021.

<https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/>

Metzger, Bryan. 2025. "Peter Thiel's political hiatus is over. Here's where his money's flowing now." *Business Insider*, July 26, 2025.

<https://www.businessinsider.com/peter-thiel-spending-money-politics-again-house-republicans-2025-7>

Miron, Marina. 2023. "Russian 'hybrid warfare' and the annexation of Crimea: the modern application of Soviet political warfare." *International Affairs*, 99(1): 397–398,

<https://doi.org/10.1093/ia/iia279>

Mitchell, Margaret et al. 2018. "Model Cards for Model Reporting." *Cornell University*, FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29--31, 2019, Atlanta, GA, USA. <https://doi.org/10.48550/arXiv.1810.03993>

Monahan, Jennifer. 2023. "Artificial Intelligence, Explained." *Carnegie Mellon University Heinz College*, July 2023. <https://www.heinz.cmu.edu/media/2023/July/artificial-intelligence-explained>

Nye, Joseph S., Jr. 1990. "Soft Power." *Foreign Policy*, no. 80 (Autumn): 153–71.

Washingtonpost.Newsweek Interactive, LLC. <http://www.jstor.org/stable/1148580>.

Nyrhinen, Mari. 2008. "IT Infrastructure: Structure, Properties and Processes." *All Sprouts Content*, 140. https://aisel.aisnet.org/sprouts_all/140

Palmer, Leigh. 2025. "Google Public Sector and Palantir Collaborate to Bring Google Cloud to FedStart." *Google Cloud Blog* (Public Sector blog), April 23 2025. <https://cloud.google.com/blog/topics/public-sector/google-public-sector-and-palantir-collaborate-to-bring-google-cloud-to-fedstart>

Palantir Technologies Inc. 2024. "U.S. Special Operations Command Expands Contract with Palantir to Deliver Advanced AI and Mission Manager Capabilities." *Palantir Technologies Inc. Investor Relations*, Dec. 9 2024. <https://investors.palantir.com/news-details/2024/U.S.-Special-Operations-Command-Expands-Contract-with-Palantir-to-Deliver-Advanced-AI-and-Mission-Manager-Capabilities/>

Perrigo, Billy. 2024. "How Anthropic Designed Itself to Avoid OpenAI's Mistakes." *Time*, May 30, 2024. <https://time.com/6983420/anthropic-structure-openai-incentives/>

Poesen, Michiel. 2022. "Regulating Artificial Intelligence (AI) in the European Union (EU): Exploring the Role of Private International Law." in X, *Recht in beweging – 29ste VRG-Alumnidag 2022* (Gompel&Svacina 2022) 297-314. <http://dx.doi.org/10.2139/ssrn.3959643>

Qureshi, Waseem Ahmad. 2020. "Information Warfare, International Law, and the Changing Battlefield." *The Fordham Law Archive of Scholarship and History*, <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2786&context=ilj>

Robinson, Michael et al. 2015. "Cyber warfare: Issues and challenges." *Computers & Security*, 49: 70-94. <https://doi.org/10.1016/j.cose.2014.11.007>

- Schaffer, Melanie. 2021. "Palantir and AWS Team Up for ERP Suite." *Yahoo Finance*, March 5, 2021. <https://finance.yahoo.com/news/palantir-aws-team-erp-suite-205011002.html>
- Sharma, Sudha; et al.. 2020. "Face Recognition System Using Machine Learning Algorithm." *Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES)*. <https://doi.org/10.1109/ICCES48766.2020.9137850>
- Shen, Yuanzi. 2024. "A Study of the Correlation Between Corporate Financial Transparency and Market Performance" *Financial Engineering and Risk Management*, 27-35. <https://doi.org/10.23977/ferm.2024.070504>
- Spezzati, Stefania and Gwladys Fouche. 2024. "Thiel's Palantir dumped by Norwegian investor over work for Israel." Reuters, Oct. 25, 2024. <https://www.reuters.com/technology/thiels-palantir-dumped-by-norwegian-investor-over-work-israel-2024-10-25/>
- Taddeo, M. 2016. "Just Information Warfare." *Topoi*, 35: 213–224. <https://doi.org/10.1007/s11245-014-9245-8>
- Tallarita, Roberto. 2025. "Amoral Drift in AI Corporate Governance." *Harvard Law Review* 138 (6, April). <https://harvardlawreview.org/print/vol-138/amoral-drift-in-ai-corporate-governance/>
- Virtosu, I and Goian, M. 2024. "Disinformation using artificial intelligence technologies – a key component of Russian hybrid warfare." *Smart Cities International Conference (SCIC) Proceedings*. 11, (Jun. 2024), 197–222. <https://scrd.eu/index.php/scic/article/view/493>
- Weiner, Stacy. 2022. "How the war in Ukraine threatens hospital cybersecurity — and what to do about it." *AAMC News*, May 24, 2022. <https://www.aamc.org/news/how-war-ukraine-threatens-hospital-cybersecurity-and-what-do-about-it>
- Wigell, Mikael. 2019. "Hybrid interference as a wedge strategy: a theory of external interference in liberal democracy." *International Affairs*, 95(2): 255-275. <https://doi.org/10.1093/ia/iiz018>
- Yang, Yiran. 2025. "Racial Bias in AI-Generated Images." *AI & Society*, <https://doi.org/10.1007/s00146-025-02282-1>

Yoon, CAPT Sukoon and Wonhee, Dr. Kim. 2023. “The Import of Hybrid Activities in the South China Sea.” Journal of Indo-Pacific Affairs, 93-101.

<https://media.defense.gov/2023/Sep/21/2003305667/-1/-1/1/JIPA%20-%20YOON%20&%20KIM.PDF>