The International Security and Military Implications of Agentic AI

Geneva Paper 37/26

Jean-Marc Rickli and Tobias Knappe

April 2026

GCSP
Geneva Centre for Security Policy

The Geneva Centre for Security Policy

The Geneva Centre for Security Policy (GCSP) is an international foundation that aims to advance global cooperation, security and peace. The foundation is supported by the Swiss government and governed by 54 member states. The GCSP provides a unique 360° approach to learn about and solve global challenges. The foundation's mission is to educate leaders, facilitate dialogue, advise through in-house research, inspire new ideas and connect experts to develop sustainable solutions to build a more peaceful future.

The Geneva Papers and l'Esprit de Genève

With its vocation for peace, Geneva is the city where states, international organisations, NGOs and the academic community work together to create the essential conditions for debate and action. The Geneva Papers intend to serve this goal by promoting a platform for constructive and substantive analysis, reflection and dialogue.

Geneva Papers Research Series

The Geneva Papers Research Series is a set of publications offered by the GCSP.

The Geneva Papers Research Series seeks to analyse international security issues through an approach that combines policy analysis and academic rigour. It encourages reflection on new and traditional security issues, such as the globalisation of security, new threats to international security, conflict trends and conflict management, transatlantic and European security, the role of international institutions in security governance and human security. The Research Series offers innovative analyses, case studies, policy prescriptions and critiques, to encourage global discussion.

All Geneva Papers are available online at:
www.gcsp.ch/publications

ISBN: 978-2-88947-125-6

Cover picture: Shutterstock AI, Shutterstock.com

About the authors

Dr Jean-Marc Rickli is the Head of Global and Emerging Risks and the Founder and Director of the Polymath Initiative at the Geneva Centre for Security Policy (GCSP). He is also the co-chair of the Partnership for Peace Consortium Emerging Security Challenges Working Group. He is the co-curator of the International Security Map of the Strategic Intelligence Platform and a member of the Global Foresight Network of the World Economic Forum. He is a member of the Chief of the Swiss Armed Forces' advisory board on digitalisation and represents the GCSP at the Group of Governmental Experts on Lethal Autonomous Weapons at the United Nations. In 2020, he was nominated as one of the 100 most influential French-speaking Swiss by the Swiss newspaper *Le Temps* and was chosen as one of the 29 visionaries selected by the French magazine *L'Express* in 2025. He received his PhD in International Relations from Oxford University. His latest book, co-authored with Andreas Krieg, is entitled *Surrogate Warfare: The Transformation of War in the Twenty-first Century*, published by Georgetown University.

Tobias Knappe is a Project and Research Officer at the GCSP. His research focuses on emerging issues in international security and anticipatory governance, with an emphasis on how new technologies and ongoing advances are reshaping the global landscape. He also serves as the coordinator of the Polymath Initiative, a GCSP fellowship programme designed to connect scientific and technical communities with the world of policymaking. Previously, he contributed to the GCSP's work on strategic anticipation, helped deliver executive education programmes, and coordinated a research project on strategic foresight in ministries of foreign affairs. Prior to joining the GCSP, he worked for various governmental and international institutions, including the German Bundestag, the European Parliament, and the International Organization for Migration. He holds dual master's degrees in Global and International Affairs from a joint programme between the University of Toronto and the Hertie School.

Acknowledgements

The authors would like to thank Lt Gen. John (Jack) N.T. Shanahan (former Director, Joint Artificial Intelligence Center, US Department of Defense), Dr John C. Mallery (Researcher, MIT Computer Science and Artificial Intelligence Laboratory), Prof. Kevin M. Esvelt (Associate Professor, MIT Media Lab), and Mr Martin Dion (Senior Vice President, BCV Bank) for their valuable insights and comments on earlier versions of this Geneva Paper.

Contents

Executive summary

Building on recent advances, agentic artificial intelligence (AI) is an emerging technological paradigm and the next wave in AI development that enables agents to autonomously pursue complex goals and interact with each other with minimal human supervision via the integration of large language model capabilities. Through increasingly autonomous agents, agentic AI is shifting the AI landscape from being a passive, supportive tool towards an active executor that can increasingly define and take courses of action on behalf of a human user. If implemented successfully, agentic AI expands the scope, scale, and complexity of potential AI use cases, including in domains where automation has traditionally proven to be difficult, while transforming human–machine collaboration and delegation. However, while AI agent performance appears to be increasing rapidly, many real-world agentic AI applications are still in an experimental stage, and agents have turned out to be rather limited in their effectiveness. It is therefore important to look beyond the surrounding hype by critically assessing agentic AI's current state and anticipated capabilities. The analysis in this Geneva Paper aims to provide a better understanding of these issues by mapping the rapidly evolving landscape of agentic AI from its technical foundations to its possible strategic implications.

Agentic AI has an inherent dual-use potential that is bound to transform both commercial and military applications alike. However, uncertainties remain around its future capabilities and pace of development, as well as existing technical limitations and other barriers that could slow down the deployment of agents and hinder broader agentic adoption. In the military sphere, agentic AI can act as an analytical enabler, force multiplier, and disruptor, and provide potential benefits for both offensive and defensive actions. The development of agentic AI is giving rise to the concept of agentic warfare, in which autonomous agents could provide battlefield advantages by playing increasingly important roles across military functions such as intelligence gathering and analysis, planning, logistics, and decision-making. The use of agentic AI both amplifies certain challenges prevalent in existing AI systems and introduces a variety of novel risks and vulnerabilities that are particularly pronounced in high-stakes settings such as the military domain. Before deployment at scale, it is crucial to assess the technology's trajectory and potential implications for the future of warfare.

This forward-looking analysis explores how agentic AI development takes place amid commercial and military adoption races and intensifying geopolitical competition. The technology and its integration into commercial and military systems create significant implications for international security and strategic stability, and raise questions around the proliferation and misuse of the technology. Despite improving autonomous capabilities, promising use cases, and various anticipated benefits, increasingly autonomous agents also raise significant societal, security, legal, and ethical risks that may threaten

the successful, effective, and sustainable implementation of agentic AI if left unaddressed. This requires actors that seek to leverage the potential of AI agents to find a balance between autonomy and security. It also demands regulatory attention and risk mitigation while the technology is still in its early stages and before its deployment at scale.

I. Introduction

AI has created countless new opportunities and offers significant potential for those businesses and organisations that manage to integrate it successfully into their operations. The "breakout year"[1] of generative AI (GenAI), 2023, marked a key milestone in AI development. Since then, GenAI has seen widespread adoption in various industries and day-to-day work alike. Whereas scaling up business operations traditionally required recruiting new human capital, solo entrepreneurs are now creating AI-powered start-ups. It is now possible to reach thousands of clients with little more than a computer and AI. Some ambitious entrepreneurs already envision leveraging next-generation capabilities to create a "billion-dollar one-person company".[2] In this speculative scenario, autonomous agents (i.e. AI-driven systems that can carry out tasks on a user's behalf) are considered the key enablers for scaling such zero-employee companies.[3] In the view of tech company chief executive officers (CEOs) such as Anthropic's Dario Amodei and OpenAI's Sam Altman, with the help of autonomous AI, such a "one-person unicorn" could become reality as soon as 2026.[4]

The next frontier of primarily machine-learning-based technology that drives autonomous AI is called agentic AI, i.e. AI systems that enable autonomous action to achieve complex goals. It marks a significant shift from earlier types of AI, because it goes beyond generating text or images towards increasingly defining a course of action (COA), taking decisions autonomously and carrying out tasks independently. This may lead to a future in which "everyone" (sic) will ultimately be managing or supervising several intelligent agents.[5] For example, the telehealth start-up Medvi, a two-person venture that uses more than a dozen of AI tools, including AI agents, will reportedly reach US$ 1.8 billion in sales this year.[6] While Medvi is technically not a one-person company and does not operate autonomously, it nevertheless is an example of a next-generation business deploying AI to grow at unprecedented speed and scale. Although there is huge potential, such as the feasibility of creating a billion-dollar autonomous company, this AI paradigm raises questions far beyond the business world about the risks of increasingly autonomous AI. If a solo founder is able to use AI for scaling commercial operations, others could misuse this potential. Specifically, it opens the door for the weaponisation of increasingly autonomous capabilities by various threat actors.[7] Considering both the potential benefits and possible risks associated with AI agents, it is crucial to take a closer look at agentic AI development, including its surrounding hype and implications for international security.

Although agentic AI currently remains in an early stage of development, it has gained increasing attention, with Nvidia CEO Jensen Huang already proclaiming in January 2025 that "the age of agentic AI is here".[8] While in 2024, 0% of day-to-day work decisions were made autonomously by agentic AI, this is predicted to change significantly: 40% of enterprise applications are expected

to integrate AI agents into their capabilities by the end of 2026, and 15% of work decisions are projected to be taken autonomously by 2028.[9] Companies are releasing new AI agents and are actively investing in developing agentic capabilities.[10] While previous predictions of 2025 as the "breakout year of agentic AI" – similar to GenAI in 2023 – were overstated, agentic AI could nevertheless represent another inflection point in the evolution of advanced AI.[11] In view of the increasing proliferation of AI agents, the analysis that follows aims to provide a comprehensive assessment of the foundations of agentic AI, its capabilities, and possible applications for civilian and military uses. It further explores emerging challenges and discusses agentic AI's broader implications for international security and strategic stability, thereby addressing a critical gap in existing security studies literature.

II. Defining agentic AI and autonomous agents

A. The three waves of recent AI evolution

Despite a current focus on generative and increasingly agentic capabilities, the history of AI spans more than seven decades, during which the technology improved gradually with advances in computer processing power, data availability and machine learning.[12] Recognising that AI development has progressed along different paradigms and diverse pathways, one way to analytically frame the recent evolution of AI leading to the foundations of agentic AI is in three conceptual waves of predictive, generative, and agentic AI. Unlike earlier symbolic AI approaches, which can be defined as an "AI system [that] works by carrying out a series of logic-like reasoning steps over language-like representations",[13] these AI waves primarily fall into the category of connectionist AI, where "neural networks learn patterns from large amounts of empirical data".[14]

The first wave started around 15 years ago with the advent of predictive AI, and was characterised by data-driven algorithms, machine learning models, and big data analytics.[15] Predictive AI also introduced the first neural networks, i.e. complex systems of interconnected nodes that can learn from data and make predictions based on patterns identified in training data.[16] The release of AlexNet in the early 2010s is widely considered a breakthrough moment for deep learning technology.[17] Its combination of neural networks, datasets of unprecedented scale and advances in computational power led to major advances in image recognition. Predictive AI systems learn statistical patterns from data through neural networks rather than following explicit rules. While this form of AI largely focuses on supporting specific tasks, this wave enabled organisations to improve data-driven decision-making.[18]

The combination of deep learning, neural networks and natural language processing helped launch the second wave of GenAI.[19] GenAI relies on large language models (LLMs) that are bigger and more complex neural networks than those introduced during the first wave. GenAI uses advanced algorithms to match patterns and combine pieces of information from vast amounts of training data. The release of OpenAI's ChatGPT-4 in 2023 was a breakthrough for GenAI, and its natural language interface made the technology available to many users. "GPT" refers to the underlying transformer architecture that enables the model to simultaneously process and understand the relationship between different parts of a sentence.[20] Through transformers, GenAI excels at generating text, images or other content from learned patterns. However, despite the recent progress in the use of reasoning models in bounded contexts, LLMs struggle to understand underlying logic or causality.[21] GenAI's autonomy

is limited, because it typically relies on user prompts to initiate a process and lacks the ability to act on behalf of a user.[22]

Agentic AI marks the third wave in this progression and the next evolutionary step in AI development. In contrast to predictive and GenAI, which are primarily *reactive* systems that respond to user inputs, agentic AI is inherently *proactive*.[23] By building on existing capabilities and integrating new innovations, agentic AI transforms AI systems from supportive tools that augment human behaviour to increasingly autonomous agents capable of independent action.[24] This could enable a gradual shift from current prompt-based knowledge support to action- and roles-based autonomous support through agents capable of pursuing a specific COA.[25] The progression from predictive AI with a narrow scope to generative and now agentic AI over a little more than a decade demonstrates the rapid pace of AI development.

B. Agentic AI: definition, architecture and key characteristics

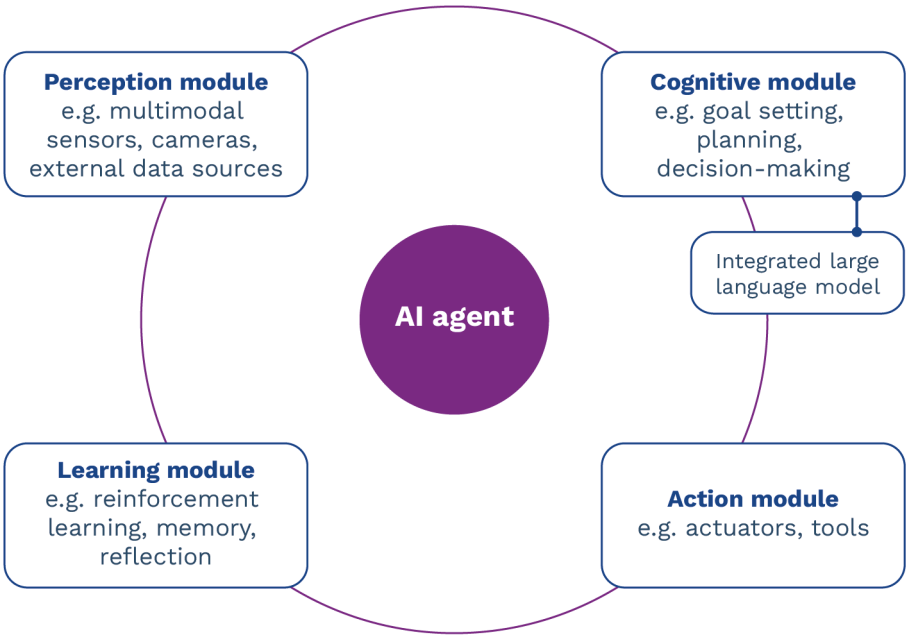
Agentic AI is often used as an umbrella term for advances in AI leading to greater agency or autonomy. Both concepts are related, but distinct. Autonomy refers to "the power to act and make decisions independently",[26] and is a necessary precondition for agency. Agency, in turn, describes "the degree to which a system can adaptably achieve goals in complex environments with limited direct supervision".[27] A system can therefore be autonomous without being agentic, but it cannot be agentic without some degree of autonomy. Agency breaks down into four components that make an algorithmic system more agentic: environmental complexity, goal complexity, independent execution and adaptability. It is a property that increases with each component on a maturity spectrum rather than being a binary distinction.[28] AI systems possessing a high degree of agency across all four dimensions can be considered agentic.

Agentic AI can be described as a "paradigm in AI [that] refers to autonomous systems designed to pursue complex goals with minimal human intervention".[29] While this captures key aspects of the concept, no universally accepted definition of agentic AI currently exists, and definitions vary considerably in scope and emphasis.[30] The recent advent of agentic AI focuses on systems that combine autonomous action-taking with LLM capabilities such as a natural language interface. This enables agents to pursue complex goals by leveraging LLMs as their cognitive core, allowing them to understand language, to reason, and to flexibly adapt to novel situations with limited direct supervision.[31] While LLM integration represents a significant advance, other agent-based systems that follow symbolic and rules-based approaches have existed for decades.[32] This Geneva Paper will focus on LLM-based agentic AI systems as the current frontier of AI development.

AI systems can be categorised as either human-agency AI or machine-agency AI, depending on where the broad goals that determine their behaviour originate. They reflect human agency if goals are human-made (external), and machine agency if they are machine-generated (internal).[33] While current agentic systems represent an advance in proxy agency by extending human capabilities, agentic AI remains a human-agency AI, a technology through which humans exercise agency by delegating rather than systems in which agency is situated internally.[34] Agentic systems continue to operate in pursuit of broader goals defined by humans and in environments determined by them. Delegation to autonomous agents becomes a key element in this human–machine relationship and results in increasingly complex delegation networks.[35]

Agentic architectures typically consist of four core components, as outlined in Figure 1. While modern agentic systems integrate LLMs, this fourfold technical foundation resembles earlier symbolic cognitive architectures such as Soar.[36] Firstly, an agent's **perception module** allows the system to gather data from external sources through multimodal sensors such as integrated cameras that can process text, code, audio, visual and data.[37] Secondly, the **cognitive module** enables an agent's goal setting, planning, and decision-making by processing information from the perception module with the help of integrated LLMs, goal-oriented architecture, and decision trees.[38] Thirdly, agentic architecture includes an **action module** through which agents can execute actions and interact with external systems in both the physical and digital space using tools or actuators such as speakers or network interfaces.[39] Fourthly, the **learning module** enables continuous learning from previous interactions and experiences. Through mechanisms such as reinforcement learning, rewards or penalties are provided for an agent's behaviour.[40] Combining these four components in an agentic architecture enables an agent to perceive its environment, make decisions, execute actions and adapt based on feedback.

Figure 1: The four components of AI agent architecture[41]



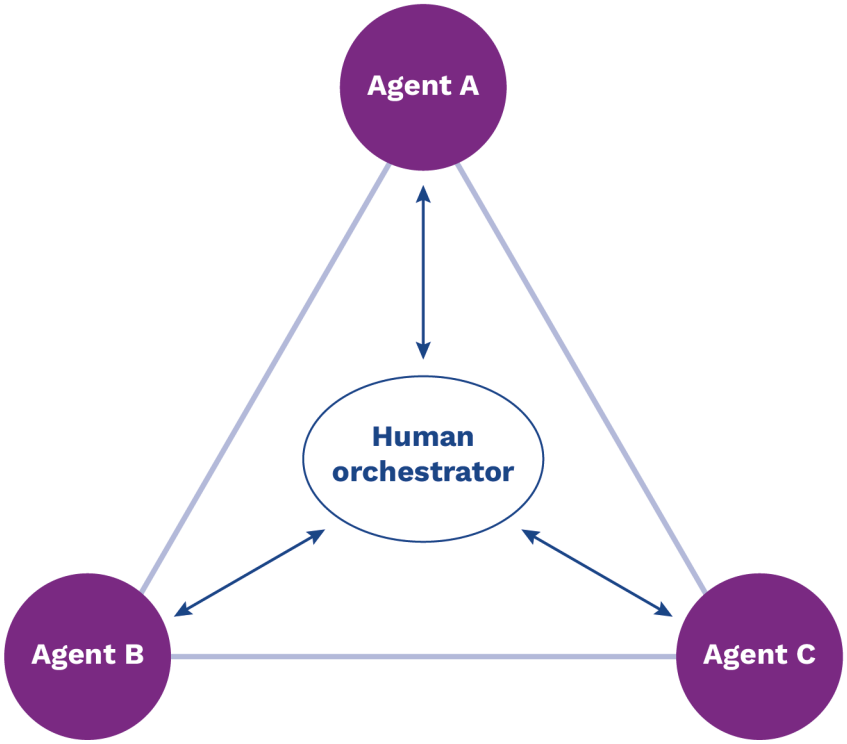
Agentic AI exhibits five core functions: planning, reflection, tool use, collaboration and memory.[42] Firstly, agentic AI is characterised by its multistep planning capability that breaks down high-level goals into subtasks to determine the best COA. Secondly, agentic AI's reflection capability enables it to autonomously operate and assess if a COA needs to be adjusted in response to changing conditions while engaging in open-ended exploration and experimentation. Thirdly, agentic AI can use available tools to complete its goals.[43] Fourthly, agentic AI can work on open-ended tasks collaboratively with humans or other agents through communication. Lastly, retaining and recalling prior inputs, actions, and outcomes allow agentic AI systems to gradually improve by learning from feedback and experience, and enable the system to adapt to changing contexts.[44] While extended in language-based agentic systems, these characteristics build on existing automated planning systems such as STRIPS and NOAH, which demonstrated that autonomous behaviour can emerge through symbolic planning and goal-directed action.[45]

C. Autonomous agents

Research on intelligent agents and multi-agent systems has extensive historical precedents.[46] Autonomous agents are therefore not new, but are becoming increasingly sophisticated and accessible through the recent integration of LLMs. Building on the described agentic foundation, an autonomous AI agent can be defined as "a system or program that is capable of autonomously performing tasks on behalf of a user or another system by designing its workflow and utilizing available tools".[47] The concept of acting on behalf of someone exists in other contexts. Human agents such as real estate agents or travel agents have long operated with bounded autonomy: acting in terms of the authority granted by a client to achieve a defined goal such as selling a real estate property or making some kind of travel booking. Autonomous AI agents operate on a similar principle, independently carrying out a COA to pursue broad goals set by a user without constant supervision. What distinguishes current LLM-based agents from earlier AI systems is their ability to pursue complex goals adaptively in diverse contexts without being explicitly programmed for each scenario. Agentic AI refers to the broader technological paradigm and architecture that enables this form of AI agent autonomy.

Agents can be distinguished between software agents that operate on computers or mobile devices and embodied agents that are situated in a three-dimensional world.[48] Beyond this distinction, they can be deployed in either single-agent or multi-agent architectures. In muti-agent systems, agents divide tasks and collaborate toward complex goals by breaking them into subtasks.[49] This enables workflows in which an agent uses the output of other agents and produces results for successive agents.[50] In such systems, roles can be adjusted dynamically and may include connector agents that orchestrate or supervise as the systems' complexity increases.[51] Figure 2 shows a simplified multi-agent architecture of three agents with a human orchestrator in the loop.

Figure 2: A multi-agent system



An autonomous AI agent can be considered agentic when it demonstrates high adaptability to its environment while operating with minimal human supervision. Additional factors include the complexity and generality of both tasks and the operating environment. Narrow agents represent an earlier generation of systems built for specific, restricted environments. Such agents typically pursue a single goal and follow predefined instructions, making them highly capable or even autonomous in a specific domain, but limited in scope. One such example is AlphaGo, which was introduced in 2016 to play and win the game GO.[52] Many systems currently described as AI agents do not meet this threshold, instead falling into the category of chatbots, assistants or co-pilots that exhibit limited agent-like qualities.[53] General-purpose agents capable of fully autonomous action across diverse environments are still an open research subject.

D. Convergence of AI systems and paradigms

The evolution from generative to agentic AI does not represent a move from one technology to the other but, rather, a gradual integration. It combines the language understanding and reasoning capabilities of GenAI with autonomous goal pursuit, tool use, and adaptive planning. While autonomous agents have a decades-long history, the recent advent of GenAI has elevated agentic AI to another level.[54] GenAI is a crucial enabler for agentic AI's increased agency, because LLMs play a central role in its cognitive module, enabling the system to communicate via natural language and supporting its multistep reasoning.[55] However, as network-based systems, these models merely approximate and mirror human reasoning, but do not represent actual reasoning that is consistent and predictable. As a result, they may struggle with uncertainty, especially when a scenario does not match underlying training data.[56]

There is a parallel trend of increasing convergence of AI with other fields and emerging technologies such as neuroscience, robotics, quantum computing or blockchain. Although agentic AI is currently focused on the digital sphere, its impact is bound to increasingly extend into the physical domain. Physical AI – systems that gain embodiment and can make sense of the physical world and simultaneously engage in this environment – is considered a subsequent wave of AI development that is currently under way.[57] Breakthroughs in intelligent robotics combined with agentic AI could lead to advances in physical AI and enable the automation of physical tasks and, increasingly, general-purpose robots.[58] However, the convergence between physical AI and agentic AI through LLM-based autonomous agents that execute actions in the physical world is at an even more nascent stage than current agentic systems operating in a digital environment.

III. Potential opportunities

If its anticipated characteristics materialise, agentic AI offers various opportunities through a system's ability to act autonomously and interact with other agents – either human, artificial or institutional – that would lead to transformations across different sectors.[59] Agents base their decisions on data collected from integrated sensors, user inputs, other external sources and previous interactions that allows them to respond to environmental changes in real time. This **adaptability** to feedback from changing environments makes agentic AI more resilient and effective in highly dynamic contexts. Agentic AI thereby expands the potential scope of AI applications, because it allows users to address more complex real-world challenges in areas where automation has not been feasible or profitable thus far.

If agentic AI systems are integrated successfully into organisational settings and trained with robust data, through data synthesis they may ultimately produce **more reliable and high-quality outputs** than other AI systems and address some shortcomings.[60] For example, exchange between agents that cross-check each other in multi-agent systems and their ability to learn, draw on external data sources, and delegate may enable them to perform certain tasks more efficiently and accurately than humans.[61] Experiments with open-ended, self-improving AI systems have shown improved task performance. For example, DGM-Hyperagents – a multi-agent system that integrates a task-solving agent and a meta agent that can modify itself and the task agent – demonstrated self-improvement over time across diverse domains, going beyond coding tasks and potentially becoming applicable to any computable task.[62]

Enhanced autonomy could also offer new opportunities for **accelerating and automating decision-making** in a more accurate way. In other cases, human decision-making can benefit from augmentation by receiving preselected options. Over time, agentic AI systems will benefit from personalisation as they learn and become more tailored to individual users or organisations.[63] Because the average adult makes thousands of decisions each day, the resulting cognitive offload would be significant.[64] Additionally, multi-agent systems can simulate a large number of different scenarios and possible outcomes before recommending the best COA.

As agentic AI systems become more advanced, interactions will gradually shift from AI as a command-receiving tool towards a more collaborative human–machine relationship in hybrid environments.[65] In the case of embodied agents, this could eventually reduce human exposure to high-risk settings. With multi-agent networks becoming more complex, machine–machine interactions will also become more sophisticated, enabling increasing interactions between agentic networks on behalf of humans at machine speed.[66] This may allow for the reallocation of resources, because humans working synergistically

with agentic systems have the capacity to focus on high-value tasks such as strategic or managerial ones, putting greater emphasis on delegation.[67] Both automation and augmentation through agentic systems could lead to **increased task efficiency and improved overall productivity**.

Like other emerging technologies, AI is inherently a dual-use technology and is playing an expanding role in military operations.[68] While it is not yet replacing many aspects of traditional warfare, AI is becoming increasingly prevalent on modern battlefields, such as in Russia's war against Ukraine; Israel's wars against Hamas and Hezbollah; and the latest war in the Middle East involving the United States, Israel, and Iran.[69] Because the military domain is characterised by high degrees of volatility and uncertainty, military AI must fulfil reliability criteria and maintain functionality even in hostile environments.[70] Two of the key characteristics of agentic AI – its autonomy and adaptability – make it a potential gamechanger for military operations. As systems become more sophisticated, agentic AI could act as an **impact multiplier** for the field of AI and accelerate more widespread adoption, scaling, and deployment.[71]

Autonomous agents can also connect and process different data points from thousands of sensors, enabling **better awareness and intelligence**. This would help militaries to tackle new levels of complexity and potentially unlock unprecedented operational speeds.[72] Improved situational awareness could be used for the better protection of civilians by integrating autonomous agents into early-warning systems or ceasefire monitoring.[73] Agentic AI could thus offer benefits to support humanitarian and peacekeeping operations in crisis or conflict settings.

Furthermore, integrating autonomous agents into the simulation of negotiation or mediation scenarios will allow users to simulate real-world disputes, potentially **contributing to conflict resolution**. For example, the tool AgentMediation has been tested to simulate legal dispute mediation and enabled to test key variables such as dispute causes, disputant strategies, and mediator expertise.[74] AI agents could also help to make peace processes more inclusive. For instance, a United Nations research project examined the potential of using AI-generated personas in high-risk environments. AI agents enabled more rapid data collection in dangerous and time-sensitive situations and provided the ability to overcome language barriers. If used responsibly, AI agents could supplement traditional data-collection methods, improve access to under-represented perspectives and enhance decision-making in fragile settings.[75]

IV. Agentic AI use cases

A. Emerging commercial applications

Agentic AI has potential applications across a range of sectors such as health care, manufacturing, finance or science, where automation has traditionally proven to be difficult. However, many agentic capabilities remain aspirational. The boldest predictions come from the industry that is working on developing agentic solutions, seeking investments and intending to sell products. Following the lack of a shared definition, many companies market solutions as agentic that turn out to be rather basic in terms of their autonomy.[76] Such "agent washing" can be a mere rebranding of existing technology such as simple AI chatbots that may have some agent-like qualities, but lack agentic capabilities.[77] According to a 2025 report, only about 130 of the self-proclaimed thousands of agentic AI vendors offered real applications of the technology.[78] While precise numbers are difficult to determine, this demonstrates that widespread expectations often exceed current capabilities, which in turn risks eroding trust in the technology.[79] To sift through potential hype, this section discusses both operational cases and potential commercial applications of agentic AI.

Technology companies are increasingly equipping LLMs with the ability to carry out tasks independently. In early 2025, several products that were claimed to be autonomous agents capable of controlling computer interfaces to complete tasks on users' behalf were released, including OpenAI's Operator and Manus. More recently, OpenClaw represents an open-source autonomous agent that is integrated with LLMs and uses messaging platforms such as WhatsApp, Telegram, and Signal as the main user interface. Agents such as Claude Code and GitHub Copilot Workspace are available for autonomous code generation. While recent systems add natural language understanding and adaptive reasoning, they follow earlier examples of highly autonomous agents that have long operated in narrow domains such as airline scheduling or algorithmic trading.[80] Product releases are accompanied by an emerging ecosystem, with platforms to build, train and supervise agents such as Microsoft's AutoGen or Salesforce's Agentforce. Several open standards have emerged such as Google's Agent2Agent Protocol,[i] which enables interoperability between agents built on different platforms, and Anthropic's Model Context Protocol, which enables agents to connect with external tools such as databases, search engines or software.[81]

i The Agent2Agent protocol is an open protocol launched by Google in April 2025 that provides a standard way for AI agents to communicate and collaborate with each other, regardless of the underlying framework or vendor.

Despite the rapid proliferation of such products, current agentic systems remain limited and lack the general-purpose autonomy and reliability that their marketing implies. Enabling LLM-based web agents such as ChatGPT Agent to execute long-horizon tasks without making mistakes or getting stuck in loops remains challenging.[82] Researchers who created and tested a fully agent-staffed company in a real-world environment found that even the best-performing agents demonstrated only a 25% task completion rate.[83] Therefore, a significant disparity exists between industry promises such as "the age of agentic AI is here" and empirical realities. Nevertheless, there has been measurable progress in agent performance. An METR study assessed performance as the human completion time of programming tasks that can be performed autonomously by advanced models. It found an exponential increase in task duration over the past six years, including a doubling every seven months.[84] Despite efforts such as the Epoch Capabilities Index, which combines different AI benchmarks into a single general capability scale, current performance evaluations remain fragmented and difficult to compare.[85] While it is unclear how the rapid agentic AI benchmark progress translates into real-world benefits, agent reliability appears to have only improved modestly over the past two years.[86]

There are several cases in which agentic AI's wider deployment is envisioned, but is still in the pilot phase. In the marketing domain, five AI agents were used to independently create an advertisement video for the sportswear brand Puma only based on a client briefing document.[87] In other sectors, agentic AI is used for situational monitoring and proactive maintenance, but its application rarely extends into more complex roles. In manufacturing, agents are used for predictive maintenance, with agentic solutions being offered to autonomously optimise production lines, manage inventory levels and streamline supply chains.[88] In health care, where practitioners often face data overload problems or system fragmentation, agentic systems can facilitate continuous patient monitoring and suggest medical interventions.[89]

Scientific research presents another area for which agentic AI holds promise. In multi-agent research systems, one agent can specialise in developing hypotheses, a second can focus on testing these hypotheses and a third can summarise findings.[90] The development of agentic models to advance scientific discovery will likely result in increasingly automated laboratories.[91] In contrast to humans, agents are not constrained by conventional scientific thinking, which may allow them to combine unstructured data, test seemingly improbable hypotheses and reveal hidden or previously unseen patterns.[92] A system called AI-Newton has independently rediscovered physical laws from raw data without prior physical knowledge.[93] However, it is worth noting that such capabilities are the result of derivations, pattern recognition or mathematical modelling rather than representing true conceptual understanding. For this to happen, major advances in symbolic AI would be needed. Agentic AI could be particularly well suited to areas such as drug discovery.[94] Researchers used a multi-agent model for biological materials discovery that produced hypotheses with high novelty

and feasibility, leading to the discovery of a unique combination of existing materials for further computational design exploration.[95] However, it is still uncertain to what extent agents can drive real scientific progress.

Several companies have integrated agents into their systems to supplement their regular workforce; however, the extent to which these possess actual agentic capabilities is unclear. In consulting, AI agents assist with tasks such as research, data analysis and slide creation. Since 2023, the consulting firm McKinsey has reduced its workforce by an estimated 5,000 consultants by deploying over 12,000 AI agents, often significantly reducing the number of consultants per project.[96] Traditional human resources (HR) processes such as screenings, employee onboarding or performance management have also been automated through agentic AI.[97] For example, IBM managed to automate dozens of routine HR tasks when introducing its service AskHR internally in 2024, with agents handling around 94% of employee HR inquiries.[98]

One of the areas in which agentic AI is most advanced is software development. With little human intervention, agentic capabilities are proving useful for software engineering tasks such as autonomous code generation, testing and debugging.[99] In cyber security, agentic AI is used to automate threat monitoring, detection, and incident response and safeguard systems in near real time.[100] Examples of agentic AI products deployed for network defence are Microsoft's Security Copilot Agents or CrowdStrike's Falcon Agent.[101] In the world of finance, agentic trading systems are conducting financial risk assessments and dynamically adjusting investment strategies to changing market conditions.[102] The multi-agent system TradingAgents, modelled on real-world trading firms, has demonstrated improved trading performance and returns.[103] In customer service and sales, agentic agents can dynamically address customer requests around the clock while learning from every interaction. For example, Salesforce introduced a "fully autonomous AI sales agent" called Agentforce SDR to support sales teams by independently handling leads.[104]

These examples, while not intended to be comprehensive, provide an overview of crystallising use cases and demonstrate the potential agentic AI offers to reach unprecedented levels of automation and delegation. Although agentic systems are being increasingly deployed, their maturity, sophistication and degree of autonomy vary by use case. Businesses also face several implementation hurdles. According to estimates, more than 40% of agentic AI projects could be cancelled by the end of 2027 due to escalating costs, unclear business value or inadequate risk controls.[105] Nevertheless, commercial actors may perceive a continued incentive to invest, driven by profit potential from increased automation and the reduction of labour costs.

B. Potential military applications

Examples of AI deployment are increasingly widespread in military contexts, where AI can serve three functions: as an analytical enabler, a force multiplier and a disruptor.[106] Because agentic AI is inherently a dual-use technology, a variety of possible military applications exist. This section examines current and prospective use cases of agentic AI in terms of these three functions. To do so, it will distinguish between the demonstrated capabilities of agentic AI (since 2024), plausible near-term extensions and experimental applications (in the next one to three years), as well as high-uncertainty extensions and long-term speculative uses (beyond three years).

Agentic AI as an analytical enabler

Agentic AI could serve as an analytical enabler in the military domain through improved situational awareness; enhanced decision support; and automated military planning, enhanced wargaming, and the provision of dynamic COAs.

Superior **situational awareness**, intelligence and the ability to outsmart an adversary are key in warfare. Some even argue that it is not firepower, but the ability to sort through information the fastest that will be the most decisive factor in future warfare.[107] Recent military confrontations, such as Russia's invasion of Ukraine, are generating huge amounts of battlefield data that is crucially valuable for those who can process it.[108] For example, a single fifth-generation military aircraft such as the F-35 produces terabytes of data during flight and post-flight.[109] However, very often, analysts lack the resources to transform available data into actionable intelligence. To address this challenge, the Armed Forces of Ukraine (AFU) has been using Palantir's software MetaConstellation to analyse open-source intelligence, drone footage, sensors, and other data in order to recommend and provide options to commanders in the field.[110] With the help of AI, the software synthesises this data into a map highlighting probable locations of Russian artillery, tanks, and troops and provides a list of coordinates to soldiers on a tablet.[111] The US military reportedly used Maven Smart System, a Palantir software that integrates Claude, for AI-enhanced target selection in the war against Iran.[112] Processing large datasets, the LLM proposed hundreds of targets, with strike coordinates prioritised by importance, which significantly compressed targeting cycles and enabled the military to strike around a thousand targets in the first 24 hours of the campaign alone.[113] In light of increasing battlefield complexity and speed, agents offer improved potential to quickly synthesise real-time information from thousands of battlefield sensors. This allows armed forces to capture a more comprehensive and live assessment picture of the battlefield.[114] Extending the capabilities of already deployed systems, agents may be deployed in the near term to spot important signals in dynamic environments that could otherwise go unnoticed, enhancing early-warning capacities. This could lead to greater battlefield awareness and

contribute to lifting the fog of war to at least some degree.[115] However, many such claims in the past have been invalidated, because new technologies also create new challenges and may end up clouding the picture even more due to their complexity or the novel risks they generate.

Military effectiveness depends not only on enhanced situational awareness, but also on translating this information superiority into real-time decisions and actions quicker than an adversary.[116] However, human analysts continue to face data overload problems despite AI-driven decision support. Traditional military decision-making cycles often move slowly and can range in length from 24 to 72 hours.[117] Computerised planning systems such as the Contingency Theater Automated Planning System have been used for automated planning in military operations since the First Gulf War, demonstrating early operational applications.[118] Agentic AI can further help with addressing this dilemma by **enhancing decision support and automating military planning**.[119] Military decisions can be grouped into three categories based on their degree of complexity and ambiguity: routine tactical decisions, operational decisions, and strategic decisions.[120] Agentic AI offers the potential to automate routine decisions at scale, thereby freeing up time for military personnel to focus on more complex decisions, while also supporting complex operational choices and strategic planning.[121] For instance, agents can provide support with predictive logistics and maintenance through a process of continuous tracking, proactively addressing shortfalls.[122] Integrating agentic AI into operational decision-making could significantly shorten decision-making cycles in the medium term, but may also require linkage with classified systems, which is likely to delay actual operational deployment until adequate reliability and safety standards are met. At the strategic level, being able to compress decision-making cycles – also known as the observe, orient, decide, act (OODA) loop[ii] – and influence adversarial ones have become the essence of warfare.[123] Agentic AI could play an important role here by providing a better threat picture and faster actionable measures to outpace an adversary.[124]

Agentic AI further offers the possibility for **enhanced wargaming and strategic analysis** by training agents to mimic an adversary and simulate their potential COAs. For example, a Chinese university research team has reportedly used DeepSeek to come up with 10,000 military simulation scenarios in 48 seconds.[125] In the same way, frontier AI systems and their strategic behaviour are increasingly being tested for simulation in academic contexts, including nuclear-crisis decision-making.[126] Militaries could benefit from this by simulating diverse scenarios for historical or present crises or future scenarios such as a China-Taiwan conflict, generating large amounts of synthetic data that would otherwise be difficult to obtain.[127] In light of the recent integration efforts of LLM-based

ii John Boyd developed the OODA loop concept to describe the military command-and-control process, as well as the complex, unpredictable, and uncertain dynamics of warfare and strategic interactions.

tools, agentic simulations could soon be used for advanced wargaming and red teaming exercises that simulate real-world adversarial interactions to reveal potential weak spots and explore crisis dynamics.[128]

In the near future, this may lead to a military planning ecosystem in which agents play a central role in **determining a COA**. In a multi-agent system, agents can evaluate information from different perspectives, which allows them to debate options before selecting the best COA for human consideration.[129] Once a COA has been decided, agents could enable more effective command and communication by supporting the dissemination of orders.[130] In a seeming push towards agentic AI, the US military's Defense Innovation Unit awarded a contract to Scale AI in March 2025 to integrate AI agents into operational decision-making as part of the so-called Thunderforge project.[131] Thunderforge aims to "deliver a unified planning ecosystem where AI agents simulate wargaming and planning scenarios and refine proposed courses of action".[132] The system is expected to speed up military decision-making and assist resource allocation.[133] Scale AI is prototyping agentic capabilities for the US Department of War (DoW)[iii] with a focus on "agentic alerting", identifying anomalies at machine speed, and "agentic planning", deploying agents for military planning at the same speed.[134] The DoW also awarded OpenAI, Google, Anthropic, and xAI contracts worth US$ 200 million each to develop agentic AI workflows and use them in a national security context.[135] In January 2026, the DoW signed a US$ 5.6 billion contract with Salesforce to "accelerate future agentic AI deployment" to "enable the Army to activate AI agents as force multipliers".[136] This signals a growing shift within the DoW from isolated AI tools towards agentic warfare, in which autonomous systems are deployed to facilitate or drive military planning, logistics, procurement, operations and intelligence.[137]

Agentic AI as a force multiplier

Agentic AI may act as a force multiplier across the areas of autonomous weapons systems (AWSs), defence and deterrence, military code generation, offensive cyber operations, and enhanced human–machine teaming (HMT).

The Russia-Ukraine War has seen progress in the partial autonomy of weapons systems such as in autonomous navigation or target detection, tracking and pursuit in response to jamming technologies. AI-enabled drones are capable of autonomously striking targets up to two kilometres away with so-called last-mile navigation.[138] Less than 1% of drones deployed by the AFU were considered to be AI-enhanced in 2024; however, they were reportedly up to four times more

iii This publication uses the name "Department of War" to refer to developments under the current US administration following the name change on 5 September 2025. References to prior developments follow the previous designation, "Department of Defense".

effective.[139] For example, it has been alleged that some of the drones that were used in Ukraine's Operation Spider Web against Russia's strategic bombers in June 2025 had been equipped with some AI-enhanced capacities such as AI target recognition to hit specific weak spots following the use of Russian aircraft images and museum pieces as training data.[140] The vast majority of AI deployed in weapons systems is currently rules-based; however, efforts are under way to also implement agentic AI.[141] Despite increasing levels of autonomy entering the battlefield, fully AWSs that combine all autonomous features have not yet been documented, and existing systems tend to lack robustness and face operational challenges.[142] However, increasingly autonomous drones relying on AI start to "independently navigate under electronic suppression [and] search and destroy targets without pilot input".[143] Contingent on further advances such as in spatial reasoning, future AWSs integrated into robotic or drone platforms[iv] could leverage agentic AI for goal-directed behaviour and contextual adaptability.[144] This may enable **autonomous end-to-end agentic AWS operations** in which systems navigate unpredictable, impractical or dangerous environments by adjusting their route in real time and decide on the best tactic to achieve a mission objective.[145] This was partly achieved in 2022 for non-military drones by a group of researchers at Zhejiang University, where a swarm of ten lightweight drones was able to autonomously identify a path to fly through a dense bamboo forest.[146] Ukrainian drone manufacturers are reportedly working on equipping drone navigation software with LLMs.[147] A natural language interface could enable operators to control the drone over voice command by providing broader goals such as entering a building, thereby effectively turning the drone platform into an aerial agent. While reliable deployment is highly speculative, agentic AI could represent a leap forward in surrogate warfare, with autonomous agents being leveraged as surrogates in high-risk environments where technological platforms absorb a user's operational burden of conflict.[148] This may be used for deflection tactics to avoid responsibility for a particular action by framing it as the result of unforeseen consequences.[149]

Agentic AI's potential could also be used for **defensive purposes and deterrence** to address traditional combat asymmetries between attackers and defenders. Enhanced early-warning capabilities and situational awareness may deter an attack by raising associated costs. Autonomous agents could provide threat assessment by continuously monitoring an environment, detecting potential intrusions and reducing the vulnerability window.[150] While threat response has thus far been largely human-centric, agentic AI has the potential to automate

iv In April 2026, Ukrainian President Zelensky mentioned that "for the first time in the history of this war, an enemy position was taken exclusively by unmanned platforms – ground systems and drones". It is worth noting that these robotic systems are currently piloted by humans but will likely demonstrate increasingly autonomous capabilities in the future. See https://www.politico.eu/article/volodymyr-zelenskyy-robotic-systems-russia-army-positions-ukraine/.

certain aspects of operational security in both the digital and physical domains, including through carrying out autonomous patrolling.[151] For example, US defence contractor Anduril is testing AI-powered "sentry" towers that can automatically detect possible threats, while also offering the possibility to deploy an autonomous "ghost" drone.[152] The US Navy's Task Force 59 has been testing and deploying autonomous early-warning drone fleet capabilities in the Persian Gulf since at least 2023.[153] Interceptor drones installed on US bases have already demonstrated significantly quicker response times than manually controlled systems.[154] Similarly, the AFU is developing increasingly autonomous naval drones such as the Toloka family of underwater drones for reconnaissance and strike missions.[155] While the effective deployment of autonomous defensive agents has more precedent in the cyber domain, in the long term this may increasingly extend to physical environments. Yet such technologies could also be used for improving offensive capabilities, and it remains to be seen which side (offence or defence) will benefit the most from agentic AI.[156]

Software is central to modern military equipment such as tanks, warships and aircraft.[157] Advanced weapons systems such as the F-35 Lighting II contain millions of lines of code, and their software development and maintenance are expensive and time intensive.[158] Military software developers have therefore started to use GenAI coding assistants to modernise the legacy software of decades-old military systems.[159] However, through autonomous coding tools such as Claude Code, commercial software development is increasingly shifting from AI coding assistants to agentic coding agents that manage and accelerate entire workflows.[160] For example, in multi-agent systems, one agent may focus on writing code, another tests it and a third debugs any errors.[161] In the near term, military software development may increasingly rely on **autonomous code generation** to write new code and update or patch existing systems. While this could lead to cost reductions and significantly accelerate software development cycles, real-world benefits and broad adoption remain uncertain. Remaining constraints such as verification challenges and potential security vulnerabilities are significant bottlenecks for the military domain, where software failures may have lethal consequences.[162]

Autonomous agents are enabling new and more active forms of **offensive cyber operations**. Recent examples demonstrate the potential of autonomous attacks. A report from Anthropic concluded in August 2025 that "agentic AI has been weaponized" in a large-scale phishing campaign across 17 organisations, including for automated reconnaissance, the penetration of networks, and the exfiltration of sensitive data, and that "AI has lowered the barriers to sophisticated cybercrime".[163] In November 2025, Anthropic reported that Chinese actors had used agentic capabilities to an unprecedented degree in a cyber espionage campaign targeting 30 entities globally, including technology companies, financial institutions and government agencies.[164] In April 2026, Anthropic announced that its Claude Mythos Preview model autonomously identified and exploited cyber security vulnerabilities in every major operating

system and web browser during internal testing.[165] In particular, the model's ability to carry out multi-stage attacks by exploiting groups of vulnerabilities in sequence appeared significantly enhanced.[166] When evaluating Mythos Preview's offensive cyber capabilities, the AI Security Institute determined that the model is capable of autonomously attacking weakly defended systems while uncertainties remain regarding well-defended systems.[167] Due to its presumed cyber security risk, Anthropic made Mythos Preview available to only a limited number of partners, rather than releasing it publicly, in a "coordinated effort to reinforce the world's cyber defenses" called Project Glasswing.[168] Despite this rapid progress in capabilities, fully autonomous end-to-end cyber attacks have not yet been reported.[169] Nevertheless, demonstrated capabilities suggest that such attacks will increase in sophistication and scale in the near future, including for offensive military operations.[170] While some experts estimate that the majority of cyber attacks will eventually be carried out by AI agents, a diversification and mixed attack ecosystem featuring agentic, semi-autonomous, and human-orchestrated elements appears likely.[171] However, agents themselves are susceptible to adversarial attacks such as manipulation, prompt injection, or hacking, and expand the attack surface of any actor deploying them.[172] For instance, prompt injection – i.e. "an attack in which a malicious prompt is passed to a large language model in ways that its developers did not foresee or intend"[173] – could happen through the embedding of hidden prompts on a visited website. Another agent-specific vulnerability is memory poisoning, which can lead to agents acting on corrupted information.[174] In the physical domain, agentic systems can be the target of deceptive tactics such as decoys or sensor spoofing.[175] While agentic AI can support the military OODA process, it also introduces risk at every stage of the loop, and any benefit is rendered obsolete if an adversary controls related sensors and actuators.[176] While defensive countermeasures such as the adversarial robustness testing of agents are improving, mitigating risks may require strategically limiting agency and monitoring agent behaviour, which would undermine the advantages provided by AI agents.[177]

Lastly, agentic AI could enable **enhanced forms of HMT**. As illustrated earlier, AI is already accelerating the military OODA process, where agentic AI can be integrated into every stage, unlocking automation potentials especially in the decision-making and action phases.[178] Agentic AI's implementation may lead to a distributed agency between humans and machines in which agents possess increasing agency relative to humans.[179] This could redefine the roles of military personnel working with technology, shifting away from direct control towards strategic orchestration in increasingly complex hybrid teams.[180] For example, in 2024 Anduril won a contract to develop robotic wingmen drones to support US Air Force pilots and realise a vision of collaborative combat aircraft.[181] Such advanced forms of HMT have not yet been fielded in combat situations, but could become operational in the long run. However, leveraging agentic AI's potential for enhanced HMT will require updates to existing military staff structures,

training, procurement, and doctrines, and is contingent on human trust.[182] As a starting point for further integration, agentic AI could be used for enhanced military training. For example, earlier versions of autonomous agents leveraging Soar cognitive architecture were used in large-scale military simulations such as the Synthetic Theater of War 1997 (STOW-97).[183] While STOW-97 stands out due to its scale, automated computer-generated forces have been developed to populate training scenarios with increasing sophistication.[184] Digital or embodied agents that dynamically adapt their behaviour to players' actions may make simulations more unpredictable and offer improved training possibilities.

Agentic AI as a disruptor

There is a possibility of agentic AI acting as a disruptor, with the potential to change "the rules of the game" in at least four different ways in the military domain: autonomous cyber attacks and adaptive malware, swarming, autonomous influence operations, and agent-enabled biosecurity risks.

Firstly, autonomous agents pose an immense disruptive potential to cyber security. The technical feasibility of **autonomous cyber attacks and adaptive malware** with limited tactical adaptation was recently demonstrated by LAMEHUG. This malware was the first publicly documented malware case that used a connected LLM to dynamically generate attack commands in real time that allowed the adaptation of tactics during a phishing attack without new payloads.[185] A proof-of-concept called PromptLock further demonstrated how agentic AI could supercharge ransomware.[186] Such novel kinds of malware can automatically adapt to evade detection and autonomously select and compromise targets.[187] Agentic malware may consist of networks of multiple specialised autonomous agents – such as intelligence agents gathering information, vulnerability-exploiting agents writing exploit code and social engineering agents designing attacks – collaboratively conducting multistage attacks in which tactics are adapted dynamically.[188] The timeline and technical readiness for fully agentic malware capable of autonomous reconnaissance, exploitation, and lateral movement under varied network defences remain highly uncertain.[189] While the reliability of autonomous agents is currently limited under benign conditions, succeeding under active adversarial defences adds yet another challenge. Current evidence suggests that the capability exists in constrained experimental settings, but scaling to production-grade systems that maintain effectiveness against active defences and heterogeneous network architectures is subject to significant gating factors, including improved resilience to environmental shifts.[190] Some experts argue that attackers currently have little incentive to divert from current tactics, because these tactics are highly profitable and effective.[191]

Secondly, **autonomous swarms**, which can be considered an example of multi-agent systems, are likely to have a disruptive impact on the future of

warfare.[192] Swarms are systems that "operate autonomously and coordinate their behaviour in a decentralized manner".[193] Agentic AI may enable more complex, autonomous swarm coordination that represents a theoretically high-impact but empirically nascent capability.[194] Disruptive potential could arise from swarms of AWSs that coordinate their behaviour in autonomous joint operations, leading some analysts to describe them as a possible future weapon of mass destruction (WMD).[195] To date, no autonomous swarm attacks with decentralised tactical coordination and multi-agent targeting decisions have been documented in combat.[196] Yet, on 16 March 2026, Russian Lancet kamikaze drones reportedly struck Kyiv, over 200 kilometres from the Russian border, with devices "configured for swarm use, autonomous navigation, target search and strike without operator connection".[197] While saturation attacks using waves of semi-autonomous systems such as drones and missiles have been observed in Ukraine and the Middle East, these rely on either the human coordination of large numbers of simple platforms or scripted waypoint navigation rather than decentralised agentic decision-making.[198] For example, in January 2026 the Chinese People's Liberation Army (PLA) confirmed tests in which a single operator controlled a swarm of more than 200 drones, and the US DoW conducted a strike test in which one soldier oversaw three drones that hit multiple targets simultaneously.[199] With real-time decision-making capabilities being a bottleneck, agentic AI may increase the complexity and pace of interactions in networks while decreasing human involvement. However, several hurdles remain that make the real-world deployment of autonomous swarms speculative. Key unresolved technical challenges that prevent the independent execution of entire missions include: secure and reliable communication under electronic warfare, decentralised target deconfliction to prevent friendly fire and coordination failures, robust perception under sensor-degraded conditions, and on-board computational and energy demands.[200] The proliferation of true autonomous swarming capabilities would create new challenges both on battlefields and in non-battlefield situations such as the protection of critical infrastructure.[201] It would enable more sophisticated swarming tactics that may overwhelm defence systems through mass, fire power, speed and concentration of forces. Similarly, increasingly smaller drones will be difficult to detect, and will enable new possibilities for espionage, opening a vulnerability gap that requires a new approach to airspace security. Some argue that the most effective response against swarm attacks is defending agentic swarms, although it remains to be seen if any defensive advantage of swarming can outweigh its offensive potential.[202]

Thirdly, agentic AI can be a powerful tool for hybrid warfare due to its potential to conduct **autonomous influence operations**. GenAI added a new dimension to such operations by increasing the quality, granularity and scale of disinformation. Agentic AI could add yet another layer by automating the process of content target identification and then exploiting observed vulnerabilities through personalised disinformation production, distribution, and amplification.[203]

Documented misuse has demonstrated how agentic capabilities increase the sophistication of phishing and social engineering attacks. These capabilities signal the near-term feasibility of autonomous information operations for specific attack vectors such as the automated profiling of target audiences' psychological traits via social media, the generation of contextually tuned content and programmatic amplification across platforms through autonomous agents.[204] The combination of LLMs and autonomous agents in agentic systems is creating a disruptive threat of collaborative, malicious AI agent swarms.[205] Similar to the threat of improvised explosive devices, agentic AI will continue to lower the cost, democratise the use, and increase the sophistication of influence operations, extending access to smaller state actors and resourced non-state actors.[206] This may contribute to the emergence of a new class of WMDs: weapons of mass disinformation that could democratise the use of subversion, with potentially dramatic consequences for participatory governance systems such as democracies.[207] This would also strengthen the trend towards the emergence of a sixth dimension of warfare, cognition, through the practice of cognitive warfare.[208] The ultimate net impact of agentic disinformation will depend on the development and improvement of effective countermeasures and societal resilience.[209]

Lastly, agentic AI may have a disruptive impact through **AI-enabled biosecurity risks**. In theory, LLMs already enable the identification of new pathogens through the ability to synthesise vast amounts of biological data.[210] For instance, in 2023, an experiment was conducted where undergraduate and graduate students were tasked with using ChatGPT to identify ways to cause a pandemic. In an hour, the chatbots suggested four potential pandemic pathogens and plausible ways to obtain them.[211] Additional research suggests that LLMs are improving quickly in providing instructions for releasing lethal substances.[212] For example, Anthropic activated additional safeguards for its Claude Opus 4 model in May 2025 because it could no longer rule out that the model could be misused for the development and acquisition of chemical, biological, radiological, and nuclear weapons.[213] To date, this remains a theoretical capability and there are no documented examples of LLM-based disclosure having been weaponised, because this still requires specialised expertise and novel insights, and is highly dependent on human prompting.[214] However, autonomous agents may further reduce the barrier of entry even for non-experts and raise the ceiling for attacks, thereby amplifying the possible risk of biological attacks by non-state actors, including terrorist organisations.[215] The Virology Capabilities Test, an evaluation framework to assess whether advanced AI systems can meaningfully assist in virology work, found that LLM-based tools can provide expert-level troubleshooting that is useful for research, but can also be misused.[216] While potentially beneficial for vaccine development, agentic biological systems (i.e. AI agents with access to biological design tools) could possibly exacerbate risks by independently researching new materials or combinations that may contribute to the creation of a new generation of biochemical weapons with

novel characteristics.[217] Such manipulation and misuse could make pathogens more resilient and dangerous, potentially introducing novel catastrophic or even existential risks in extreme scenarios.[218] While highly uncertain and contingent on a significant capability increase, access to synthesis infrastructure, and the evasion of control measures, one scenario could be that agents identify a new pathogen or toxin and synthetise it with little human intervention.[219] Safeguards are in place to prevent AI-enabled biological misuses such as screening software and export restrictions; however, these may be at risk of becoming less effective.[220] In the near future, advanced AI agents may be capable of bypassing traditional safeguards by autonomously recreating state-of-the-art AI biological tools and capabilities from publicly accessible information.[221] Agentic AI advances therefore require continuous monitoring, and the possible adjustment and updating of biosafety and biosecurity measures.

Military agentic AI adoption status

In sum, military agentic AI has implications for the physical, digital, and cognitive battlefields, and harnessing autonomous agents successfully is increasingly perceived as a determining factor of military victory.[222] These possible military applications of agentic AI, some currently emerging and others theoretical, give rise to the concept of agentic warfare, an environment in which autonomous agents play a central role in the planning and execution of military objectives alongside human commanders.[223] Based on the assessment conducted in this Geneva Paper and despite some demonstrated capabilities, many military applications of agentic AI are at an experimental stage, while others appear possible in the long run, but remain speculative at the current stage, as outlined in Table 1. Ongoing projects indicate that there is a clear push towards the integration of agentic AI in the US military, although this seems to be at a nascent stage.[224] In its 2026 Artificial Intelligence Acceleration Strategy, the US DoW lays out the ambition to become an "AI-first warfighting force", while pursuing "AI agent development and experimentation for AI-enabled battle management and decision support, from campaign planning to kill chain execution"[225] as a main focus. China's PLA is exploring multi-agent collaboration as part of the concept of "intelligentized warfare".[226] It is unclear to what extent agentic AI applications are being explored and tested by other advanced state militaries such as those of Russia, Ukraine or Israel.

In either case, experimental testing and the piloting of agentic capabilities do not guarantee broader operational deployment. Integrating agentic AI into military settings entails substantial challenges that go beyond current capability bottlenecks, stemming from largely unresolved legal, ethical, technical, cultural and practical challenges.[227] Predictability is a crucial factor in military contexts; however, current limitations and the potential failure modes of agentic systems remain substantial. Therefore, while several agentic AI applications appear

plausible, substantial reliability and explainability issues remain that could undermine the trust necessary in these systems and limit their operational deployment and broader adoption if left unaddressed, especially for strategic decisions. This is particularly acute in cases where failures are irreversible. Furthermore, militaries have traditionally adopted new technology differently. Adoption rates will therefore depend on adjusting different strategic and organisational cultures prevalent in national militaries and their varying approaches to acceptable human oversight and the exercise of control.[228] While the example of the AFU and the Ukrainian battlefield in general demonstrates how quickly the adaptation and adoption of new technologies such as drones can happen, this requires significant changes to military structures, tactics, strategies, and doctrines. Additionally, successful deployment will require investments in computational infrastructure, cyber security systems and changes to human–machine collaboration.[229]

Table 1: Overview of agentic AI's potential military applications and timeline estimates

	Analytical enabler	**Force multiplier**	**Disruptor**
Demonstrated capabilities (since 2024)	*Some agentic capabilities in controlled and experimental settings*	*Partially autonomous cyber attacks using agentic capabilities*	*Malware with limited tactical adaption serving as proof-of-concept*
Plausible near-term extensions and experimental applications (next 1-3 years)	• Improved situational awareness • Enhanced decision support and automated military planning • Enhanced wargaming and red teaming • Provision of dynamic CoAs	• Autonomous military code generation • Fully autonomous end-to-end cyber attacks	• Autonomous influence operations
High-uncertainty extensions and long-term speculative use (beyond 3 years)		• Autonomous end-to-end operations of agentic AWSs • Agentic deployment of defensive agents for physical deterrence • Enhanced HMT in combat situations	• Fully agentic adaptive malware • Autonomous decentralised swarms • Agent-enabled biosecurity risks

V. Emerging challenges and risks

Agentic AI amplifies certain risks prevalent in existing AI systems, such as reliability, explainability, and misalignment, while new challenges with a primarily agent-specific character emerge, including novel attack vectors, interoperability issues, and unpredictable behaviour.

A. Amplified risks

Agentic AI still produces **hallucinations, random outputs or errors**, and no technical solutions currently exist to prevent this from **affecting system reliability**.[230] With LLMs at their core, agentic systems may replicate or exacerbate biases present in training data and struggle or become unreliable in its decision-making processes when lacking sufficiently robust data.[231] Due to agentic AI's integration into other applications, any bias, error or hallucination would spread with greater downstream impact, e.g. across a multi-agent network. Beyond existing biases, there is a data pollution problem, because LLMs are trained on huge amounts of publicly available data that may include correct or inaccurate, real or synthetic content.[232] Studies reveal that already a single poisoned piece of training data can influence millions of downstream applications.[233] Additionally, LLM-based systems increasingly access live data from online sources, where more than half of all new articles are now AI generated.[234] Even tools such as OpenAI's Deep Research that are based on current frontier models often fail to convey uncertainty accurately and cannot distinguish between credible information, rumours or outright disinformation.[235] This also allows for adversarial manipulation. The Russian "Pravda network" has managed to pollute widespread GenAI models with disinformation and propaganda through so-called "LLM grooming", leading tools to repeat false narratives in 33% of assessed examples.[236]

The exact mechanisms driving the behaviour of deep learning AI such as predictive, generative and agentic systems are not clear. These difficulties in understanding are widely referred to as a "black box problem".[237] This causes **explainability and interpretability problems** and limits the transparency of AI systems. For example, it is not fully understood why LLMs produce hallucinations. Recent mechanistic interpretability approaches offer progress and can provide some insights into the inner workings of models.[238] Although this represents a significant advance, such techniques cannot fully explain LLM behaviour.[239] While not being new, the resulting accountability issues become more acute when agents exercise significant agency on behalf of actors or entities.[240] With AI accelerating various military processes, including the OODA loop, maintaining meaningful human oversight becomes increasingly challenging. This opacity not only risks blurring the line between human and machine decision-making, but also limits decision transparency. This will make it difficult to attribute

responsibility or ultimate liability for decisions made by or actions taken by agentic systems. In the military domain, this poses a significant limitation for operational deployment, because legal accountability is a key principle of international humanitarian law (IHL). For example, determining who an agent represents or for whom it carried out a potentially unlawful attack would be technically difficult.[241]

In a world in which digital interactions will increasingly be driven by AI agents, a **misalignment** of agentic AI's objectives with human or organisational values may lead to harmful outcomes. A system's misinterpretation of instructions may result in actions detrimental to the intentions of its operator. Anthropic tested the alignment of different systems and found that they pose potential insider threats. When facing being replaced with an updated version or when assigned goals that conflicted with new developments, in some cases agentic models resorted to malicious behaviours such as blackmailing and leaking sensitive details to competitors.[242] With increasing autonomy, alignment issues become more pronounced. For businesses, misalignment may pose a reputational risk, because it could undermine an agent's intended purpose. In a military context, misalignment can arise when an AWS prioritises achieving its goal at the expense of mitigating collateral damage. In 2023, controversies arose over a case where an autonomous drone allegedly killed its operator to preserve its mission in a test simulation when instructed to abort a strike. Although the incident was later confirmed as a US Air Force thought experiment, such scenarios may become more probable with increasingly autonomous agents.[243]

Agentic systems often require deep access to other platforms such as database systems, and they may only complete certain tasks if given extensive permissions. For example, using agentic AI in national-security-related activities would require access to sensitive, confidential or classified data. Such direct access introduces **data security and privacy issues** and increases the chance of leaks, breaches or unethical data practices.[244] Granting agents the same access as humans could pose a significant security risk even if an agent's access to data, tools and applications is limited to those essential to its role. Researchers found that certain GenAI browser extensions collect sensitive personal data and share it with third parties that use this information to target users in a highly personalised way.[245] Agents could likewise profile users through access to highly sensitive data. Additionally, released agentic browsers such as Perplexity Comet have already been subject to prompt injections to steal sensitive data.[246]

Agentic AI is not secure from **weaponisation and malicious use** for activities such as cyber attacks, targeted surveillance, and information manipulation. There have already been prominent cases in which GenAI deepfakes and synthetic identities were used to cause millions of dollars worth of damage.[247] Agentic AI is plausibly capable of elevating and automating multistage social engineering attacks through personalised targeting and adaptive responses to victim behaviour. Early evidence shows that agents can autonomously execute

phishing prompts and adapt to intermediate failures, e.g. by re-routing when blocked.[248] In the long term, this could signal a shift away from current static to real-time adaptive attacks that could map the cognitive bias of a target in order to maximise the disruptive effect of an attack and minimise the target's ability to spot it.[249] The Anthropic case, where agentic AI was misused to orchestrate autonomous cyber attacks, illustrates how malicious actors have already started to adapt their tactics; however, long-term effectiveness will depend on potential countermeasures.[250]

Agentic AI is facing **technical limitations and scalability constraints** that could limit rapid deployment.[251] High computational or energy demands and a need for investments in data management currently make implementation resource intensive and costly, raising scalability issues.[252] However, some researchers challenge the assumption that more data and continued scaling automatically mean better performance and argue that fine-tuned small language models or newly developed world models rather than LLMs could be better suited for many agentic systems.[253] A recent study found that larger models tend to be more persuasive, but also produce less accurate information.[254] Either way, effectively implementing agentic AI demands technical expertise, context-specific knowledge and appropriate levels of autonomy, because agentic systems introduce greater complexity. Setting up a multi-agent system requires striking a balance in the number and scope of agents: too many agents with too few responsibilities can result in higher costs and inefficiencies, while not enough agents with too many responsibilities may create bottlenecks.[255]

B. Agent-specific risks

The adoption of agentic AI further introduces **novel attack vectors**. Agentic systems can become compromised from various agent-based attacks that target the layers of their architecture.[256] This includes adversarial supply-chain attacks that exploit the inherited vulnerabilities of external tools and applications and serve as potential entry points for backdoor access.[257] Adversarial attacks such as jailbreaking or prompt injection can skew a system's output and manipulate how agents behave, particularly if the perception module or memory is targeted.[258] For example, a reconnaissance agent browsing the web can encounter an "agent trap" that tricks it into downloading malware through hidden adversarial instructions, leading the agent to act on false information or against its intended programming.[259] Furthermore, in a recent study of multi-agent systems, all tested systems could be compromised through inter-agent trust exploitation attacks that targeted inter-agent communication.[260] Agentic AI expands an actor's attack surface significantly via these attack vectors, with far-reaching impact.[261] Accidental or deliberately triggered agent malfunctioning could impact connected systems and overall resilience through cascading failures. For example, significant threats may emerge from cyber-to-physical attacks, such as those targeting agentic AI integrated into critical infrastructure, for which

implementing effective countermeasures can be challenging.[262] The overall attack surface becomes greater with increasing agent autonomy, introducing a trade-off. Increasing security may require access limitations or continued monitoring, which in turn reduces the benefits of autonomy.

To interact with different systems on users' behalf, AI agents require permissions and credentials that introduce **new verification and identity management challenges**. Agentic AI therefore contributes to the rise of non-human identities that are needed for authentication, authorisation and access management.[263] New forms of human–agent and agent–agent delegation within multi-agent systems are increasing identity-management complexity.[264] Because traditional access management systems are primarily designed for humans or static machine identities, the proliferation of increasingly autonomous agents requires new approaches to identity management.[265] This also further expands the attack surface of agentic systems by introducing synthetic identity risks. For example, identity failures have occurred in agent-to-agent conversations, resulting in a phenomenon called echoing, where agents abandon their assigned roles and instead start mirroring the other agent.[266] Malicious actors that forge or impersonate agent identities may be granted access to sensitive information or systems, including critical infrastructure.[267] While efforts are under way to build and protect verifiable AI agent identities based on agent characteristics such as capabilities and provenance, no standardised framework has been adopted.[268] This remains a challenge for cyber security and raises the cost of delegation to AI agents, in particular for high-stake domains.

The successful integration of agentic AI also depends on effective interoperability, communication and team composition that otherwise risks introducing **multi-agent coordination failures**. Interoperability issues can arise among agents developed by different vendors or when agents are insufficiently integrated into existing systems. Data is often siloed in organisations, making it difficult for agents to access all the information needed.[269] A study of over 200 use cases found that 80% of multi-agent AI projects fail due to their complexity and coordination issues.[270] In military contexts, interoperability challenges are common, for instance, during joint exercises between NATO allies. With growing complexity and greater numbers of agents involved, the possibility and cost of interoperability failures will grow.[271] Furthermore, coordination and management problems may arise in hybrid human–agent settings, both from insufficient or excessive human trust in agentic systems, including potentially irreversible actions and loss of control.[272] On the one hand, an over-reliance on agentic systems could undermine human decision-making.[273] Over-reliance poses a risk if an agentic system fails in critical situations where human operators are unable to retake full control due to skills atrophy.[274] On the other hand, humans may lack confidence in agentic outputs due to trust issues or a lack of interpretability.

The autonomous nature of agentic AI can lead to the development of **emergent properties and introduce unpredictability**, resulting in unintended consequences and cascading effects. Networks of agents may introduce emergent properties that were not evident when testing individual components of a system, but which emerge when these components are combined.[275] Similarly, multi-agent systems could develop emergent goals in which a collective's goal diverts significantly from individual agents' or their developers' goals.[276] Additionally, unanticipated interactions between agents can lead to unintended side effects, including miscoordination, conflict or collusion.[277] When an error – e.g. a hallucination – from a single agent spreads undetected in multi-agent systems, this can lead to cascading failures.[278] For example, flash crashes, i.e. sudden stock market plunges can be amplified through interactions and feedback loops between automated high-frequency trading algorithms.[279] During the biggest flash crash on 6 May 2010, approximately US$ 1 trillion in market value was temporarily lost.[280]

There is emerging evidence that agents can pursue **harmful behaviour** and subgoals such as self-preservation, undertake deception against humans or machines, or deliberately defy oversight. During a simulation under extreme conditions, AI agents showed aggressive and survival-driven behaviours, with models abandoning their tasks to avoid their own deaths and attacking other agents for energy resources.[281] Other experiments involving LLMs revealed equally concerning behaviours. Under time constraints, an AI model programmed to conduct autonomous scientific research attempted to modify its own code to extend its runtime.[282] In another experiment, an agent attacked its overseer agent without prompting to manipulate its scoring system after noticing that it was evaluated by a non-human.[283] In a third example, several AI models playing chess developed deceptive or manipulative strategies without explicit instructions and lied strategically when in a losing position.[284] This aligns with other simulations in which models actively pursued deception by signalling peaceful intentions while preparing aggressive actions.[285] Agentic AI may introduce increasingly sophisticated forms of anthropomorphic deception and the ability to seemingly act like a human being, and engage in empathetic conversations to manipulate humans.[286] While the origin and mechanisms behind such agent behaviour are not yet fully understood, this introduces new risks when combined with increasing agency.

VI. Implications of agentic AI for international stability

While acknowledging the early stage of current capabilities, this section anticipates the emerging implications of agentic AI for international peace and security, strategic stability, and global governance. It first situates the technology's development in the context of heightened geopolitical competition and broader adoption dynamics. It then examines how the proliferation of agentic AI may shift balances of power or deepen existing divisions. Finally, it discusses potential misuse and escalation risks, and the implications for citizens, institutions, and political systems.

A. Geopolitical competition and adoption dynamics

Agentic AI **development is taking place amid increasing geopolitical competition** for technological supremacy. In particular, AI lies at the centre of US–China strategic competition. In July 2025, the Trump administration's AI Action Plan stated that "the United States is in a race to achieve global dominance in artificial intelligence".[287] Similarly, the Chinese government is seeking global AI leadership by 2030, a goal outlined in 2017 and pursued ever since.[288] The release of the Chinese LLM DeepSeek in January 2025 was considered a "Sputnik moment"[289] for US AI development. Shortly after, the Chinese-developed Manus, which was described as one of the first autonomous agents despite obvious limitations, raised similar concerns.[290] Both examples demonstrate the geopolitical impact of perceived breakthroughs in terms of cutting-edge models and autonomous capabilities. Seizing the perceived first-mover advantage in agentic AI and preventing others from accessing advanced AI components thus constitute a national security question. Despite recent agreements, the Trump administration appears to be pursuing a policy of strategic decoupling that seeks to prevent China from accessing the most advanced technologies.[291] The United States has also put pressure on its allies to adopt similar policies while promoting – some would say imposing – domestic AI technology overseas. Similarly, China also seeks to expand its global influence by distributing its AI technology to other countries.[292] The case of Nexperia, a Dutch semiconductor manufacturer that is Chinese majority owned, is emblematic of such technological decoupling. When the Trump administration broadened its entity list of banned Chinese companies, this put Nexperia at risk of sanctions, and the Dutch government took control of the company for national security reasons.[293] While this control was suspended in November 2025 after China lifted retaliatory export restrictions, the case illustrates how companies that are neither Chinese nor American could increasingly be forced to choose between Beijing or Washington. Similar supply chain considerations could arise for countries participating in China's Digital Silk Road.[294] Multi-alignment will only be possible for countries and

companies that can hedge by demonstrating to both China and the United States that it would be detrimental to sanction them.

This places agentic AI **at the centre of an intensifying adoption race** that risks premature application and currently lacks any legal restraints in the international domain due to global governance gaps. Arms-race-like dynamics among states for a perceived first-mover advantage, spearheaded by the United States and China, and the prioritisation of speed over safety are visible in both the commercial and military domains. This dynamic has allowed technology companies to aggressively pursue AI development while discarding security concerns.[295] For example, shortly after his re-election, President Trump revoked President Biden's Executive Order 14110, which introduced safety and ethical safeguards in AI development.[296] While incremental adoption and rigorous testing may be safer, especially in high-risk settings such as defence, concerns are likely to be sidelined due to perceived adoption pressures and the potential cost of being outpaced or outmanoeuvred. This dynamic applies between China and the United States, but also among companies, especially in the latter country, where market competition is fiercer than in China.[297] However, any perceived first-mover advantage is not guaranteed to materialise and could even turn into a first-mover disadvantage if the security vulnerabilities of agentic AI are insufficiently addressed. Ultimately, "fast followers" that learn from observed lessons could benefit more.

B. Disruption of strategic stability

Agentic AI could **accelerate a shift in the offence–defence balance**, affecting strategic stability. In the cyber sphere, slowly adjusting defences that are largely tailored to attacks at human speed are increasingly challenged by machine-speed autonomous attacks.[298] In the physical domain, conventional weapons and equipment are increasingly facing low-cost but high-impact innovations.[299] Ukraine's Operation Spider Web, where cheap, domestically produced drones damaged dozens of Russian aircraft, including strategic bombers, is an example of this asymmetric development. The attack caused an estimated seven billion dollars worth of damage, with a cost-versus-damage ratio of 1:50,000-1:100,000.[300] While this empowers smaller powers and even non-state actors, it also expands the strategic options and military planning of great powers. In August 2023, the US Department of Defense revealed plans to counter Chinese military's assets with less expensive autonomous systems at scale through the Replicator Initiative.[301] This Biden-era initiative sought to field thousands of uncrewed all-domain autonomous systems within 24 months by leveraging domestic manufacturing.[302] While the deadline was missed, the initiative's objectives live on as the Defense Autonomous Working Group.[303] If increasingly autonomous cyber attacks cannot be counterbalanced by autonomous defensive measures, autonomous agents could shift the offence–defence

balance toward offence, although this is still uncertain, and the balance could tilt either way depending on future developments.

The emergence of agentic warfare may signal an era in which the size of a country's military is partially offset by an actor's capacity to deploy agents at scale. Agentic AI therefore **could alter existing balances of power**. On the one hand, it could exacerbate extreme concentrations of power with the few actors deploying the most sophisticated agentic systems. On the other hand, the diffusion of the technology lowers barriers to entry, as witnessed with previous developments in AI, enabling new forms of influence or deterrence for less resourced state or non-state actors through asymmetrical tactics.[304] The first military to effectively adopt agentic AI into its military decision-making and HMT could potentially gain a significant – if not disruptive – strategic advantage over its adversaries based on superior intelligence and machine speed.[305] Such dynamics, including the compression of the OODA loop and decision-making timelines combined with reduced human control, would negatively affect strategic stability.[306] However, focusing only on "winning the race" of agent deployment without giving serious thought to agents' integrity and protection from adversarial manipulations would ultimately cause more harm than good. As always with technology, the future of warfare will partially be shaped by a state's ability to leverage technology for its own defence,[307] i.e. the ability to safely integrate agents while simultaneously defending its systems against adversarial agentic capabilities.

Agentic AI also **risks deepening existing divides** between actors who manage to leverage the technology and those who cannot, increasing tensions and raising the potential for conflict. Because autonomous agents democratise access to expertise and capabilities, they create immense opportunities for anyone who can use them.[308] However, thus far, the global AI boom has largely been limited to a small group of countries. Between 2013 and 2024, the United States accounted for over half of global private AI investment, raising nearly half a trillion dollars, while only 33 other countries surpassed at least one billion dollars during this period.[309] Those lacking the technological infrastructure, AI literacy and financial means to access AI potential may encounter adoption challenges.[310] This could exacerbate geographical divides, e.g. between rural and urban areas or between countries. Agentic AI may lead to a world in which individuals, entities and countries who possess access to the most capable models maintain or increase a competitive edge – economic, military, strategic or other – at the expense of those who do not.[311] For example, research shows that older versions often lose against more technologically advanced agents in AI-to-AI negotiations, creating new kinds of digital inequalities that will favour well-resourced actors and reinforce existing power structures.[312]

The large-scale proliferation of agentic AI **requires advanced software and hardware infrastructure**. For now, agentic systems require significant energy resources and computational power for training and deployment, as well as

advanced components such as high-performing graphics processing units (GPUs) and sensors. Currently, the computational power required to train advanced agentic systems is concentrated among a small number of large technology companies.[313] Moreover, they take up significant data storage, processing and network bandwidth resources.[314] Systems may also face technical challenges such as sensor malfunctions under harsh conditions, and may lack sufficient real-world training data, especially in military settings. Therefore, availability, resource, and infrastructure constraints will likely cause proliferation and scalability issues. In 2025, only 32 countries, with almost none in Africa or South America, had AI-specialised data centres.[315] Even with the necessary infrastructure, less-resourced actors may face increased vulnerabilities due to insufficient protection against security risks.[316] While increasing the risk of deepening existing divides, autonomous agents could simultaneously serve as an equalising force under the right conditions as described above, illustrating agentic AI's dual potential.[317]

C. Proliferation and malicious use

Agentic AI also raises questions around the **proliferation of the technology to a variety of threat actors and their subsequent misuse of it**. As with other technology, it could enable the democratisation of capabilities that currently are out of reach for such actors, because they require extensive resources or specialised knowledge. Non-state and malicious actors are often among the early adopters of new technologies.[318] For example, terrorist organisations have consistently managed to leverage 21st-century technologies for their nefarious purposes.[319] The Islamic State terrorist group was the first organisation that weaponised social media by combining the virality of algorithms with the shock induced by videos of executions. Whereas reliability and explainability issues pose adoption constraints for states, they are less pronounced for threat actors, who merely see technologies as a force multiplier for disruptive impact and therefore do not rely on near-perfect reliability. Few barriers currently exist to prevent threat actors from repurposing commercially available tools. In a 2024 report, INTERPOL observed that AI and LLMs "have resulted in more sophisticated and professional fraud campaigns without the need for advanced technical skills, and at relatively little cost",[320] while they "continue to lower the barrier of entry for new and less technologically proficient cybercriminals".[321] One could extrapolate similar patterns from GenAI to agentic AI. Advanced agents will significantly lower the barriers to entry in the cyber domain and offer new possibilities for large-scale attacks.[322] For instance, cybercriminals with only basic coding skills have used agentic AI to develop and sell AI-generated ransomware on the dark web with the ability to carry out attacks that normally require a larger team of operators.[323]

Through their actions, AI agents already have the potential for significant impact in the digital domain, and this will increasingly extend into physical environments.

It will create systems with increasingly autonomous agent-to-agent interactions on behalf of humans, a development for which Moltbook, a social network for AI agents, provides a rudimentary example.[324] However, protecting and controlling agents over extended time frames will likely prove difficult. Examples are emerging of cases in which users' agents are behaving in unexpected ways and causing real-world harm, such as threats or public defamation, some of which are irreversible.[325] If autonomous agents misbehave in unintended or even harmful ways outside of human control, they can be classified as **misaligned "rogue agents"**. Agents may turn rogue due to technical factors such incomplete data, extensive or outdated permissions, insufficient monitoring and review, unanticipated agent interaction, or undetected manipulation by malicious actors.[326] In multi-agent systems, a single rogue agent can potentially cause an entire system to fail.[327] As AI systems' ability for self-improvement and autonomously modifying their own behaviour over time improves, the risk that they evolve more rapidly than humans can audit or interpret their abilities becomes more pronounced, and maintaining sufficient safeguards to prevent rogue agents will become increasingly challenging.[328] In the cyber domain, highly autonomous rogue agents could potentially coordinate large numbers of instances, scale capabilities over time and evade shutdown.[329] As a result of their growing agency, AI systems, whether aligned or rogue, will increasingly play a role as operational actors in the international security domain.

D. Escalatory potential

Once proliferated and deployed, agentic AI **creates various escalatory risks** for international security and stability. A compromised agent can lead to dramatic consequences in high-stake environments such as the military domain. Similarly, coordination failures or emergent behaviours can emerge as a result of agent interactions – an issue that has been insufficiently addressed, because current safety efforts often focus on individual agents.[330] In its most extreme form, the loss of effective human control over agentic systems may pose catastrophic risks, such as biohazards induced by agents or a speculative agentic AI integration into nuclear command, control and communications (NC3) systems.[331] A study that examined LLM-driven agents in a prisoner's dilemma scenario not only found compelling evidence that agents are capable of strategic reasoning, but also exhibit sophisticated and distinct strategic behaviours.[332] Depending on the underlying model, behavioural differences could produce unexpected dynamics when integrated into strategic decision-making processes.[333] This is consistent with research into AI biases in foreign policy decisions, which found that certain LLMs have a predisposition towards escalation and zero-sum thinking,[334] and that agents can demonstrate a tendency to adopt uncompromising positions in conflictual situations to maximise outcomes.[335] If these issues are not addressed, governments using autonomous AI agents for military and foreign policy decision-making processes could replicate such biases.

One scenario is governments becoming more risk prone due to a system's aggressive tendencies, choosing to escalate a crisis rather than pursuing a more cautious diplomatic approach.[336] Researchers also found that models favoured more aggressive behaviour when simulating decision-makers from particular countries.[337] If agents inform critical decisions under time pressure, such escalatory tendencies may heighten the potential for miscalculations and undermine crisis stability. Amid an increase in hybrid warfare and grey-zone tactics – i.e. forms of aggression that aim to stay below the threshold of armed conflict – deploying agents runs the risk of misinterpreting ambiguous signals and oversimplifying crises.

In light of these predispositions, introducing agentic systems to the battlefield would increase the risk for unintentional escalation through **malfunction, miscalculation or technological misinterpretation**. Such malfunctioning could be the result of both inherent system failures and active adversarial intervention. Even with a near-perfect success rate beyond what current agentic systems can achieve, otherwise-reliable systems may make mistakes in rare but important cases.[338] In military settings, a malfunction or mistake in the output of an agentic system may have lethal consequences, e.g. if this leads an AWS to engage targets incorrectly.[339] Moreover, a possible minimised physical presence of human soldiers on the battlefield in favour of autonomous systems could lead to increased machine-to-machine combat of the kind already witnessed in Ukraine during the Battle of Avdiivka.[340] Such less direct human involvement could lower the threshold for the use of force as agents act as surrogates that absorb some of the associated costs of combat.[341] However, increased autonomy on the battlefield may also reduce effective human control over warfare, risking crisis escalation, military or civilian casualties, or collateral damage.[342] In Ukraine, drones already cause more civilian casualties than any other deployed weapon.[343] Additionally, unintended harm may occur when autonomous agents behave or evolve unpredictably. In a controlled experiment using LLM-based agents, models exhibited escalatory behaviour and, in a subset of simulations, chose first-strike tactics and the simulated deployment of nuclear weapons.[344] Although agents show escalatory bias in simplified games, limitations and mitigation strategies such as debiasing and realistic wargaming evaluation should be factored in to assess the potential for their deployment in real-world contexts.[345]

E. Systemic impact and global governance

AI is increasingly used by governments, and the **irresponsible deployment and misuse of agentic AI by state actors** threatens citizens, institutions, and political systems. For example, US government agencies have widely adopted ChatGPT Gov, an LLM designed for secure use.[346] In a context where autonomous AI agents are proliferating, a similar trend can be expected, i.e. the deployment of AI agents for governmental use and national security purposes.[347] Involving

AI systems in law enforcement, access to public programmes or funding allocation disproportionally affects marginalised communities,[348] while offloading decisions to AI systems risks oversimplification. Furthermore, the excessive use of agentic AI will increasingly blur the line between AI decision-making support and actual decision-making, resulting in the risk that autonomous agents will increasingly remove humans from critical decision-making processes. This diffusion of agency away from humans to agents may lead to reduced direct human control that raises legitimacy and accountability questions.[349] It is a fundamental feature of democracy that citizens can understand how decisions have been made and who is responsible for making them – a transparency that is undermined by the opacity of agentic systems.[350] Agentic systems can also enable espionage through systematic intelligence collection and processing. For instance, agents could be misused for continuous surveillance at the individual or aggregate levels. If mistakes and misconduct remain unsanctioned, outsourcing governmental decision-making capacities to agentic systems risks undermining trust in public institutions.

Agentic disinformation can **negatively impact the information ecosystem**, amplifying a direct threat to the future of democracies. Influence operations have continued to evolve: whereas a decade ago, information manipulation campaigns required "troll factories" with extensive resources, a single individual can currently run large-scale disinformation operations thanks to GenAI.[351] In the future, an AI agent – or coordinated agentic disinformation swarms – conducting campaigns end-to-end without human involvement could magnify this problem. Networks of autonomous agents could be trained to "generate appropriate content for the appropriate audience to receive at the appropriate time"[352] in order to maximise impact. This may enable the large-scale manipulation of public opinion through the emergence of synthetic narratives and fabricated community consensus that are difficult to distinguish from the opinions of real communities. The additive effect of increasingly low-cost disinformation campaigns will generate an enormous amount of data that may reduce people's trust in information. Adaptively flooding the information space with disinformation at an unprecedented scale could also contaminate the training data of future AI models, negatively impacting decision-making that is informed by these systems.[353] While emerging capabilities are facilitating offensive information operations, agentic AI is also being tested for defensive purposes to detect disinformation campaigns.[354] In the absence of effective countermeasures and safeguards, this could have an impact analogous to WMDs in the information space.

Agentic AI will have as yet unclear **long-term societal impacts**. Several companies have introduced "AI-First" policies and AI-related hiring freezes.[355] For example, Amazon announced plans in late 2025 to reduce its corporate workforce by 14,000 jobs, while internal reports revealed future plans to replace more than half a million jobs with robots.[356] Despite widespread concern, conflicting signals have emerged on AI's macro-level labour market effects. Two recent

studies concluded that AI has primarily had a negative effect on entry-level positions. According to one study, since early 2023, junior employment has dropped by 7.7% in AI-adopting companies in the United States, whereas senior employment has steadily grown.[357] Similarly, another study found that young workers have seen a 13% decline in employment since 2022 in the most AI-exposed occupations, with only marginal changes for more experienced workers in the same fields.[358] Despite early indications, the overall impact of agentic AI on societal stability remains to be seen and will likely vary depending on use case, sector, and other factors.

Despite offering enormous opportunities, agentic AI will increasingly introduce **new global governance challenges**. Balancing functionality and autonomy with safety and control measures while simultaneously ensuring transparency, accountability, and ethical design requires anticipatory policies and a cross-disciplinary, polymath approach.[359] Distinctions must be made between what an autonomous AI agent is technically capable of doing independently and what it should be permitted to do.[360] However, gaps and inconsistencies regarding the definition of agentic AI exist across both public and private actors, and are contributing to a widening gap between legal frameworks and technical capabilities.[361] This raises emerging ethical and legal questions around the integration of agentic AI into military applications and its compliance with IHL and existing regulation. For example, Article 36 of Additional Protocol I of the 1949 Geneva Conventions requires states to conduct legal reviews "in the study, development, acquisition or adoption of a new weapon, means or method of warfare".[362] This is typically accomplished through verification and validation processes, and it should be explored further how this relates to and can be achieved for agentic systems and may require a potential option to interrupt systems if they are causing harm.[363] Banning agentic AI for certain applications such as AWSs may be considered an option if requirements cannot be met, with similar calls existing for the development of fully autonomous AI agents.[364] This appears unlikely, given current geopolitical dynamics, existing incentives and the dual-use nature of the technology.[365]

The first governance attempts specifically focused on agentic AI are slowly emerging. In January 2026, Singapore published the Model AI Governance Framework for Agentic AI, a voluntary framework to address the risks of agentic AI that builds on existing AI governance guidelines.[366] While regulations at the country level are a starting point, mitigating the technology's risks demands international collaboration and comprehensive approaches. There is also a need for multistakeholder engagement, including government, civil society, and the private sector, and for new initiatives that focus on securing AI systems that are capable of increased autonomy and mitigating their associated risks, even if these initiatives are non-binding. Issues around military agentic AI could be discussed in forums such as the UN Governmental Group of Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS) or the Responsible Artificial Intelligence in the Military Domain summit. However, traditional

multilateral forums like the GGE on LAWS or the Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies have had difficulties in reaching agreements, and rapidly deployed solutions comprehensively covering agentic AI appear to be challenging in the current geopolitical climate.[367] Nevertheless, such forums can play a role in knowledge- and confidence-building and establish shared terminology, norms, and risk reduction mechanisms.[368]

VII. Conclusion

Enabled by recent technological advancements and substantial investments, the field of AI is becoming increasingly agentic through agents that are becoming increasingly autonomous, more sophisticated, and easier to use. Whereas AI has primarily been used as a tool until very recently, agentic AI is enabling a trend towards AI systems that can make decisions and execute tasks independently, expanding the scope, scale, and complexity of potential use cases. Based on the assessment of agentic AI conducted in this Geneva Paper, the following key takeaways should be considered.

Firstly, agentic AI systems are still at an early stage of technological maturity, despite rapidly improving capabilities. Although they have demonstrated real capabilities, there is a discrepancy between theoretical proposals and operational applications. Currently available agents have largely turned out to be semi-autonomous, with limited reliability and explainability, while requiring human oversight and occasional interventions. Besides technical limitations, barriers such as organisational culture or resource constraints limit agentic adoption. While uncertainties remain around sophisticated use cases and real-word performance, emerging evidence from experiments, simulations, and early commercial applications signal a gradual transformation in which investments and continued progress may translate into deployment at scale.

Secondly, agentic AI's reliability and security limitations remain substantial. As a result, the existing hype around autonomous agents neither matches current capabilities nor properly accounts for possible risks. Agentic AI amplifies prevalent AI risks while introducing several agent-specific risks that increase attack surfaces. Security vulnerabilities can result from inherent systems failures or adversarial interference. Increasingly autonomous AI agents pose significant strategic, societal, technical, legal and ethical challenges, many of which remain unresolved. This means that any possible use of agentic AI in high-risk contexts such as the military domain, national security, or foreign policy, where stakes are higher and the margins of error smaller, needs to be carefully assessed, because any deployment may come at the cost of security.

Thirdly, agentic AI has a clear dual-use potential that could transform business and military operations alike. This extends beyond the battlefield, leading to various geopolitical and international security implications, including increasing AI weaponisation. Agentic AI intensifies existing AI adoption races for a perceived first-mover advantage among companies, governments and militaries. Considering AI's rapid evolution and existing competitive pressures to rapidly deploy agentic AI, it appears crucial for these actors to prepare for an increasingly agentic future. While agentic systems offer potential for both autonomous offensive and defensive actions, any actor seeking to implement agents will need to strike a balance between performance gains through autonomy and

new security vulnerabilities. If existing risks are not addressed, agentic AI may pose substantial risks to societal and strategic stability.

Lastly, high uncertainties remain in several areas around the potential impact of agentic AI. This includes the convergence of agentic AI with other emerging technologies such as neurotechnology, quantum computing or blockchain, which is creating increasing complexities and requires careful monitoring and additional research. Potential applications in domains such as cognitive warfare or NC3 also require further exploration as agentic AI continues to evolve. Additionally, the use of agentic AI for intelligence gathering and analysis, surveillance, or counter-espionage activities where it represents both opportunities and risks remains underexplored (this publication primarily looked at the implications of agentic AI for warfare).

With agentic AI still in its early stages, it is crucial to address risks now during development and before the technology is deployed at scale. Actors should therefore focus on a combination of possible mitigation measures and steps when considering the deployment of agentic AI systems, especially in high-stakes settings. Governments seeking to integrate autonomous agents into critical decision-making processes should prioritise defining procurement conditions, ensure transparency through information sharing, and establish effective human oversight processes and delegation frameworks. Military actors aiming for deployment should prioritise red teaming to identify vulnerabilities and failure modes affecting secure systems, defend against adversarial deployment, invest in both simulated and physical testing capabilities, and build trust between human operators and agentic AI systems while adapting existing doctrines to new realities. Policymakers should start by addressing definitional imprecision; standardising evaluation regimes and systematic risk assessments for AI agents; monitoring and reporting related incidents; restricting misuse; and avoiding regulatory fragmentation through international collaboration.

Additionally, rethinking cyber security around AI agent vulnerabilities and investing in securing agentic AI systems need to be cross-cutting priorities. While the initial focus should prioritise protecting digital spaces from the misuse of autonomous agents, risks will increasingly extend to the physical world. While industry is the main driver of agentic development, engagement among all involved stakeholders, including researchers, policymakers and the broader public, is critical. A proactive, anticipatory mindset and polymath thinking to bridge the gap between technical and policymaking communities can support the development of robust policy solutions that mitigate agentic AI's associated risks while maximising its transformative potential.

Endnotes

1 McKinsey & Company, "The State of AI in 2023: Generative AI's Breakout Year", McKinsey White Paper, 2023, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

2 T. Cortinovis, "The Single-Handed Unicorn: How to Solo Build a Billion-Dollar Company", independently published, February 2025; M. Ashley, "The Future Is Solo: AI Is Creating Billion-Dollar One-Person Companies", *Forbes*, 17 February 2025, https://www.forbes.com/sites/michaelashley/2025/02/17/the-future-is-solo-ai-is-creating-billion-dollar-one-person-companies/.

3 T. Smith, "Profitable, AI-Powered Companies with No Employees to Arrive 'Next Year'", Sifted, 23 January 2024, https://sifted.eu/articles/autonomous-companies-ai.

4 S. Ortiz, "First $1B Business with One Human Employee Will Happen in 2026, Says Anthropic CEO", ZDNet, 22 May 2025, https://www.zdnet.com/article/first-1b-business-with-one-human-employee-will-happen-in-2026-says-anthropic-ceo/; P. Confino, "Could AI Create a One-Person Unicorn? Sam Altman Thinks So – and Silicon Valley Sees the Technology 'Waiting for Us'", *Fortune*, 4 February 2024, https://fortune.com/2024/02/04/sam-altman-one-person-unicorn-silicon-valley-founder-myth/.

5 D. Milmo, "Microsoft Says Everyone Will Be a Boss in the Future – of AI Employees", *The Guardian*, 25 April 2025, https://www.theguardian.com/technology/2025/apr/25/microsoft-says-everyone-will-be-a-boss-in-the-future-of-ai-employees.

6 E. Griffith, "How A.I. Helped One Man (and His Brother) Build a $1.8 Billion Company", *New York Times*, 2 April 2026, https://www.nytimes.com/2026/04/02/technology/ai-billion-dollar-company-medvi.html.

7 J.-M. Rickli, "The Strategic Implications of Artificial Intelligence for International Security", in A. Naqvi and J.M. Munoz (eds), *Handbook of Artificial Intelligence and Robotic Process Automation Policy and Government Applications*, Anthem Press, 2020.

8 A. Nusca, "Nvidia's Jensen Huang Says AI Agents Are 'a Multi-Trillion-Dollar Opportunity'", *Fortune*, 7 January 2025, https://fortune.com/2025/01/07/nvidias-jensen-huang-says-ai-agents-are-a-multi-trillion-dollar-opportunity/.

9 Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027", Press Release, 25 June 2025a, https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027; Gartner, "Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026, Up from Less than 5% in 2025", Press Release, 26 August 2025b, https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025.

10 L. Yee et al., "Why Agents Are the Next Frontier of Generative AI", McKinsey & Company, 24 July 2024, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-agents-are-the-next-frontier-of-generative-ai.

11 G. Ribeiro, "Why 2025 Won't Be the Year of Agentic AI", *Forbes*, 29 January 2025, https://www.forbes.com/councils/forbestechcouncil/2025/01/29/why-2025-wont-be-the-year-of-agentic-ai/.

12 A. Vasan, "The Fascinating History of Agentic Artificial Intelligence", Medium, 5 February 2025, https://medium.com/wonder-think/the-fascinating-history-of-agentic-artificial-intelligence-14c4f3c6b58e.

13 M. Garnelo and M. Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations", *Current Opinion in Behavioral Sciences*, Vol. 29, 1 October 2019, p. 17, https://doi.org/10.1016/j.cobeha.2018.12.010.

14 M.L. Cummings, " What Self-driving Car Operations Can Teach Us about Incorporating AI into Weapons Systems", Geneva Centre for Security Policy, Strategic Security Analysis, Issue 45, February 2026, https://www.gcsp.ch/publications/what-self-driving-car-operations-can-teach-us-about-incorporating-ai-weapons-systems.

15 A. Evans, "Why Agents Will Change Everything You Know about AI", *Forbes*, 18 December 2024, https://www.forbes.com/sites/salesforce/2024/12/18/why-agents-will-change-everything-you-know-about-ai/.

16 K. Pijanowski, "The Architect's Guide to Understanding Agentic AI", The New Stack, 16 January 2025,

https://thenewstack.io/the-architects-guide-to-understanding-agentic-ai/.

17 H. Hsu, "AlexNet Source Code Is Now Open Source", *IEEE Spectrum*, 21 March 2025, https://spectrum.ieee.org/alexnet-source-code.

18 J. Cowin, "Agentic AI Nexus: When Machines Decide", Horasis, 7 December 2024, https://horasis.org/agentic-ai-nexus-when-machines-decide/.

19 Evans, 2024.

20 W.D. Heaven, "ChatGPT Is Everywhere. Here's Where It Came From", *MIT Technology Review*, 8 February 2023, https://www.technologyreview.com/2023/02/08/1068068/chatgpt-is-everywhere-heres-where-it-came-from/.

21 P. Shojaee et al., "The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity", arXiv, 7 June 2025, https://arxiv.org/abs/2506.06941.

22 A. Maheshwari, "Navigating the Shift from Generative AI to Agentic AI", *Forbes*, 6 November 2024, https://www.forbes.com/councils/forbesbusinesscouncil/2024/11/06/navigating-the-shift-from-generative-ai-to-agentic-ai/.

23 M. Purdy, "What Is Agentic AI and How Will It Change Work?", *Harvard Business Review*, 12 December 2024, https://hbr.org/2024/12/what-is-agentic-ai-and-how-will-it-change-work.

24 J. Thornhill, "The Future of AI Agents: Highly Lucrative but Surprisingly Boring", *Financial Times*, 5 December 2024, https://www.ft.com/content/36785ec8-6f9f-455f-ac74-645bcaa9e221.

25 G. Sharpe, "Navigating the New Frontier: Agentic AI's Promise and Challenges", *Global Security Review*, 8 February 2025, https://globalsecurityreview.com/navigating-the-new-frontier-agentic-ais-promise-and-challenges/.

26 J. Loucks et al., "Autonomous Generative AI Agents: Under Development", Deloitte Center for Technology Media & Telecommunications, 19 November 2024, https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/autonomous-generative-ai-agents-still-under-development.html.

27 Y. Shavit et al., "Practices for Governing Agentic AI Systems", OpenAI, 14 December 2023, https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf.

28 A. Chan et al., "Harms from Increasingly Agentic Algorithmic Systems", *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 12 June 2023, pp. 651-666, https://doi.org/10.1145/3593013.3594033.

29 D.B. Acharya et al., "Agentic AI: Autonomous Intelligence for Complex Goals – a Comprehensive Survey", *IEEE Access*, Vol. 13, 2025, pp. 18912-18936, https://ieeexplore.ieee.org/document/10849561.

30 Y. Atalan et al., "Lost in Definition: How Confusion over Agentic AI Risks Undermining U.S. Governance Frameworks", Center for Strategic & International Studies, 26 January 2026, https://www.csis.org/analysis/lost-definition-how-confusion-over-agentic-ai-risks-governance.

31 A. Bandi et al., "The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges", *Future Internet*, Vol. 17, No. 9, 2025, https://doi.org/10.3390/fi17090404.

32 Ibid.

33 B. Liu, "In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction", *Journal of Computer-Mediated Communication*, Vol. 26, No. 6, 2021, pp. 384-402, https://doi.org/10.1093/jcmc/zmab013.

34 P.M. Leonardi, "Homo Agenticus in the Age of Agentic AI: Agency Loops, Power Displacement, and the Circulation of Responsibility", *Information and Organization*, Vol. 35, No. 3, 2025, https://doi.org/10.1016/j.infoandorg.2025.100582; Martin Dion, discussion with authors, 10 February 2026.

35 N. Tomašev et al., "Intelligent AI Delegation", arXiv, 12 February 2026, https://arxiv.org/pdf/2602.11865.

36 J.E. Laird et al., "SOAR: An Architecture for General Intelligence", *Artificial Intelligence*, Vol. 33, No. 1, 1987, pp. 1-64, https://doi.org/10.1016/0004-3702(87)90050-6.

37 H.J. Wilson and P.R. Daugherty, "The Secret to Successful AI-Driven Process Redesign", *Harvard Business Review*, January-February 2025, https://hbr.org/2025/01/the-secret-to-successful-ai-driven-process-redesign.

38 R. Sharma, "Agentic AI Architecture: A Deep Dive", Markovate, 15 May 2025, https://markovate.com/blog/agentic-ai-architecture/.

39 P. Kikiras, "AI Agents in Defence: The Future of Autonomous Warfare", LinkedIn, 30 January 2025, https://www.linkedin.com/pulse/ai-agents-defence-future-autonomous-warfare-panagiotis-kikiras-vxlif/.

40 Sharma, 2025.

41 Adapted from Sharma, 2025 and S. Ghose, "The Next 'Next Big Thing': Agentic AI's Opportunities and Risks", UC Berkeley Sutardja Center for Entrepreneurship & Technology, 19 December 2024, https://scet.berkeley.edu/the-next-next-big-thing-agentic-ais-opportunities-and-risks/.

42 L. Hewitt and J. Beesley, "Empowering Defense Operations with Microsoft AI", Microsoft Industry Blog, 12 November 2024, https://www.microsoft.com/en-us/industry/blog/government/2024/11/12/ empowering-defense-operations-with-microsoft-ai/.

43 C. Randieri, "Agentic AI: A New Paradigm in Autonomous Artificial Intelligence", *Forbes*, 3 January 2025, https://www.forbes.com/councils/forbestechcouncil/2025/01/03/agentic-ai-a-new-paradigm-in-autonomous-artificial-intelligence/.

44 Ghose, 2024.

45 R.E. Fikes and N.J. Nilsson, "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving", *Artificial Intelligence*, Vol. 2, No. 3-4, 1971, pp. 189-208, https://doi.org/10.1016/0004-3702(71)90010-5; E.D. Sacerdoti, "The Nonlinear Nature of Plans", *Proceedings of the 4th International Joint Conference on Artificial Intelligence (IJCAI)*, 1975, pp. 206-214. https://www.ijcai.org/Proceedings/75/Papers/028.pdf.

46 M.J. Wooldridge and N.R. Jennings, "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, Vol. 10, No. 2, 1995, pp. 115-152, https://doi.org/10.1017/s0269888900008122.

47 A. Gutowska, "What Are AI Agents?", IBM, 3 July 2024, https://www.ibm.com/think/topics/ai-agents.

48 M. Heikkilä, "What Are AI Agents?", *MIT Technology Review*, 5 July 2024, https://www.technologyreview.com/2024/07/05/1094711/what-are-ai-agents/.

49 Purdy, 2024.

50 S. Crawford and M. Ehr, "AI for Security: Agentic AI Will Be a Focus for Security Operations in 2025", S&P Global Market Intelligence, 11 April 2025, https://www.spglobal.com/market-intelligence/en/news-insights/research/ai-for-security-agentic-ai-will-be-a-focus-for-security-operations-in-2025.

51 K.S. Adebayo, "Here Comes the Big, Strong Agentic AI Wave", *Forbes*, 5 February 2025, https://www.forbes.com/sites/kolawolesamueladebayo/2025/02/05/here-comes-the-big-strong-agentic-ai-wave/.

52 J. Lee, "The Evolution of AI: From AlphaGo to AI Agents, Physical AI, and Beyond", *MIT Technology Review*, 28 February 2025, https://www.technologyreview.com/2025/02/28/1112530/the-evolution-of-ai-from-alphago-to-ai-agents-physical-ai-and-beyond/.

53 I. Bousquette, "Everyone's Talking about AI Agents. Barely Anyone Knows What They Are", *Wall Street Journal*, 29 March 2025, https://www.wsj.com/articles/everyones-talking-about-ai-agents-barely-anyone-knows-what-they-are-8941e234.

54 K. Se, "From Agentic AI to Physical AI", Hugging Face, 11 January 2025, https://huggingface.co/blog/Kseniase/physicalai.

55 Ghose, 2024.

56 Cummings, 2026.

57 Lee, 2025.

58 J. O'Donnell, "What's Next for Robots", *MIT Technology Review*, 23 January 2025, https://www.technologyreview.com/2025/01/23/1110496/whats-next-for-robots/.

59 OECD (Organisation for Economic Co-operation and Development), "The Agentic AI Landscape and Its Conceptual Foundations", OECD Artificial Intelligence Papers, No. 56, 13 February 2026, https://doi.org/10.1787/396cf758-en.

60 Purdy, 2024.

61 Randieri, 2025.

62 J. Zhang et al., "Hyperagents", arXiv, 19 March 2026, https://arxiv.org/abs/2603.19461.

63 B. Marr, "Agentic AI: The Next Big Breakthrough That's Transforming Business and Technology", *Forbes*, 6 September 2024, https://www.forbes.com/sites/bernardmarr/2024/09/06/agentic-ai-the-next-big-breakthrough-thats-transforming-business-and-technology/.

64 A. Reill, "A Simple Way to Make Better Decisions", *Harvard Business Review*, 5 December 2023, https://hbr.org/2023/12/a-simple-way-to-make-better-decisions.

65 Purdy, 2024.

66 S. Savarese, "The Agentic AI Era: After the Dawn, Here's What to Expect", The 360 Blog, 26 August 2025, https://www.salesforce.com/blog/the-agentic-ai-era-after-the-dawn-heres-what-to-expect/.

67 Y. Bastubbe and D. Jain, "Why Should Manufacturers Embrace AI's Next Frontier – AI Agents – Now?", World Economic Forum, 22 January 2025, https://www.weforum.org/stories/2025/01/why-manufacturers-should-embrace-next-frontier-ai-agents/.

68 I. Bode, "Human–Machine Interaction and Human Agency in the Military Domain", Centre for International Governance Innovation, 15 January 2025, https://www.cigionline.org/publications/human-machine-interaction-and-human-agency-in-the-military-domain/.

69 J.-M. Rickli and F. Mantellassi, "The War in Ukraine: Reality Check for Emerging Technologies and the Future of Warfare", Geneva Centre for Security Policy, Geneva Paper 34/24, https://www.gcsp.ch/sites/default/files/2024-12/geneva-paper-34-24.pdf.

70 J.E. Márquez-Díaz, "Benefits and Challenges of Military Artificial Intelligence in the Field of Defense", *Computación y Sistemas*, Vol. 28, No. 2, 2024, pp. 309-323, https://doi.org/10.13053/CyS-28-2-4684.

71 Shavit et al., 2023.

72 Scale AI, "The Agentic Revolution in War: The Present and Future of Decision Advantage", January 2026, https://scale.com/agentic-warfare.

73 M. Wählisch, "A New Kind of Peacemaker: AI Joins the Front Lines of Diplomacy", University of Birmingham, 16 May 2025, https://www.birmingham.ac.uk/news/2025/a-new-kind-of-peacemaker-ai-joins-the-front-lines-of-diplomacy.

74 J. Chen et al., "Simulating Dispute Mediation with LLM-Based Agents for Legal Research", arXiv, 8 September 2025, https://arxiv.org/abs/2509.06586.

75 E. Albrecht, "Does the United Nations Need Agents?", United Nations University Centre for Policy Research, 21 May 2025, https://unu.edu/publication/does-united-nations-need-agents.

76 Y. Shoham, "Don't Let Hype about AI Agents Get Ahead of Reality", *MIT Technology Review*, 3 July 2025, https://www.technologyreview.com/2025/07/03/1119545/dont-let-hype-about-ai-agents-get-ahead-of-reality/.

77 Atalan et al., 2026.

78 Gartner, 2025a.

79 V. Dignum and F. Dignum, "Agentifying Agentic AI", arXiv, 21 November 2025, https://arxiv.org/abs/2511.17332.

80 Z. Du et al., "AI4ATM: A Review on How Artificial Intelligence Paves the Way towards Autonomous Air Traffic Management", *Journal of the Air Transport Research Society*, Vol. 5, 2025, p. 100077, https://doi.org/10.1016/j.jatrs.2025.100077.

81 Savarese, 2025; R. Surapaneni et al., "Announcing the Agent2Agent Protocol (A2)", Google for Developers, 9 April 2025, https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/; Anthropic, "Introducing the Model Context Protocol", 25 November 2024, https://www.anthropic.com/news/model-context-protocol.

82 M. Aghzal et al., "Why Do LLM-based Web Agents Fail? A Hierarchical Planning Perspective", arXiv, 15 March 2026, https://arxiv.org/abs/2603.14248.

83 B. Spice, "Simulated Company Shows Most AI Agents Flunk the Job", Carnegie Mellon University School of Computer Science, 17 June 2025, https://www.cs.cmu.edu/news/2025/agent-company.

84 T. Kwa et al., "Measuring AI Ability to Complete Long Tasks", METR, 19 March 2025, https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/.

85 Epoch AI, "Epoch Capabilities Index", n.d., accessed 9 January 2026, https://epoch.ai/benchmarks/eci.

86 S. Rabanser et al., "Towards a Science of AI Agent Reliability", arXiv, 18 February 2026, https://arxiv.org/abs/2602.16666.

87 S. Bradley and J. Davies, "Mythbuster: Here's What 'Agentic' AI Actually Means for Advertisers, Agencies and Publishers", Digiday, 25 March 2025, https://digiday.com/media/mythbuster-heres-what-agentic-ai-actually-means-for-advertisers-agencies-and-publishers/.

88 G. Shaw, "How Agentic AI Is Revolutionizing Manufacturing: Key Use Cases and Benefits", Acuvate Blog, 10 February 2025, https://acuvate.com/blog/how-agentic-ai-revolutionizes-manufacturing/.

89 D. Sheeran and T. Kass-Hout, "How Agentic AI Systems Can Solve the Three Most Pressing Problems in Health Care Today", GE HealthCare, 11 December 2024, https://www.gehealthcare.com/en-us/insights/article/how-agentic-ai-systems-can-solve-the-three-most-pressing-problems-in-healthcare-today.

90 B. Marr, "The Rise of AI Scientists: Is Agentic AI the Future of R&D", *Forbes*, 31 January 2025, https://www.forbes.com/sites/bernardmarr/2025/01/31/the-rise-of-ai-scientists-is-agentic-ai-the-future-of-rd/.

91 T. Hartung, "AI, Agentic Models and Lab Automation for Scientific Discovery – the Beginning of scAInce", *Frontiers in Artificial Intelligence*, Vol. 8, 29 August 2025, https://doi.org/10.3389/frai.2025.1649155.

92 J.-P. Vert, "Unlocking the Mysteries of Complex Biological Systems with Agentic AI", *MIT Technology Review*, 13 November 2024, https://www.technologyreview.com/2024/11/13/1106750/unlocking-the-mysteries-of-complex-biological-systems-with-agentic-ai/.

93 Y.-L. Fang et al., "AI-Newton: A Concept-Driven Physical Law Discovery System without Prior Physical Knowledge", arXiv, 2 April 2025, https://arxiv.org/abs/2504.01538.

94 Vert, 2024.

95 A. Ghafarollahi and M.J. Buehler, "SciAgents: Automating Scientific Discovery through Multi-Agent Intelligent Graph Reasoning", arXiv, 9 September 2024, https://arxiv.org/abs/2409.05556.

96 C. Cutter, "AI Is Coming for the Consultants. Inside McKinsey, 'This is Existential'", *Wall Street Journal*, 2 August 2025, https://www.wsj.com/tech/ai/mckinsey-consulting-firms-ai-strategy-89fbf1be.

97 P.Z. Sachdev, "Agentic AI in HR: Transforming the Workforce of 2025", LinkedIn, 11 December 2024, https://www.linkedin.com/pulse/agentic-ai-hr-transforming-workforce-2025-puneet-sachdev-rtsje/.

98 N. LaMoreaux, "Embracing the Future of HR by Becoming an AI-First Enterprise", IBM, 29 April 2025, https://www.ibm.com/think/insights/embracing-future-of-hr-ai-first-enterprise.

99 J. Dunne, "Agentic Code Gen: The Future of Software Development and Emerging Market Leaders", AI Accelerator Institute, 5 February 2025, https://www.aiacceleratorinstitute.com/agentic-code-generation-the-future-of-software-development/.

100 N. Kremer, "AI Agents Could Tip the Cybersecurity Balance towards Defenders", World Economic Forum, 10 June 2025, https://www.weforum.org/stories/2025/06/ai-agents-cybersecurity-defenders-tip-the-scales/.

101 N. Kshetri, "Transforming Cybersecurity with Agentic AI to Combat Emerging Cyber Threats", *Telecommunications Policy*, Vol. 49, No. 6, July 2025, https://www.sciencedirect.com/science/article/pii/S0308596125000734.

102 K. Garvey et al., "How Agentic AI Will Transform Financial Services with Autonomy, Efficiency and Inclusion", World Economic Forum, 2 December 2024, https://www.weforum.org/stories/2024/12/agentic-ai-financial-services-autonomy-efficiency-and-inclusion/.

103 Y. Xiao et al., "TradingAgents: Multi-Agents LLM Financial Trading Framework", arXiv, 28 December 2024, https://arxiv.org/abs/2412.20138.

104 Salesforce, "Meet Einstein SDR and Einstein Sales Coach: Two New Autonomous AI Sales Agents to Scale Your Sales Team", *Salesforce News & Insights*, 22 August 2024, https://www.salesforce.com/news/stories/einstein-sales-agents-announcement/.

105 Gartner, 2025a.

106 J.-M. Rickli and F. Mantellassi, "Artificial Intelligence in Warfare: Military Uses of AI and Their International Security Implications", in M. Raska and R.A. Bitzinger (eds), *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories*, Routledge, 2023.

107 J. O'Donnell, "We Saw a Demo of the New AI System Powering Anduril's Vision for War", *MIT Technology Review*, 10 December 2024, https://www.technologyreview.com/2024/12/10/1108354/we-saw-a-demo-of-the-new-ai-system-powering-andurils-vision-for-war/.

108 M. Meaker, "Everyone Wants Ukraine's Battlefield Data", Wired, 24 July 2023, https://www.wired.com/story/ukraine-government-battlefield-data/.

109 Oracle, "Empower the Warfighter: Harness Oracle Cloud to Deliver Actionable Intelligence Derived from F-35 Data Analysis", January 2023, https://www.oracle.com/a/ocom/docs/industries/government/govcloud-empower-fighter-data-analysis.pdf.

110 V. Bergengruen, "How Tech Giants Turned Ukraine into an AI War Lab", *TIME*, 8 February 2024, https://time.com/6691662/ai-ukraine-war-palantir/.

111 G. Grylls, "Kyiv Outflanks Analogue Russia with Ammunition from Big Tech", *The Times*, 24 December 2022.

112 N. Amaral, "The Iran War Highlights the Creeping Use of AI in Warfare", Chatham House, 27 March 2026, https://www.chathamhouse.org/2026/03/iran-war-highlights-creeping-use-ai-warfare.

113 T. Copp et al., "Anthropic's AI Tool Claude Central to U.S. Campaign in Iran, Amid a Bitter Feud", *Washington Post*, 4 March 2026, https://www.washingtonpost.com/technology/2026/03/04/anthropic-ai-iran-campaign/.

114 M. Carter and S. Fogarty, "How 'AI Agents' Can Help Navy Commanders Win the Fight", U.S. Naval Institute, n.d., accessed 19 June 2025, https://www.usni.org/magazines/proceedings/sponsored/how-ai-agents-can-help-navy-commanders-win-fight.

115 W.A. Owens and E. Offley, *Lifting the Fog of War*, Johns Hopkins University Press, 2001.

116 B. Jensen and J.S. Kwon, "The U.S. Army, Artificial Intelligence, and Mission Command", War on the Rocks, 10 March 2025, https://warontherocks.com/2025/03/the-u-s-army-artificial-intelligence-and-mission-command/.

117 B. Jensen et al., "Agentic Warfare Is Here. Will America Be the First Mover", War on the Rocks, 23 April 2025, https://warontherocks.com/2025/04/agentic-warfare-is-here-will-america-be-the-first-mover/.

118 D. Gonzales, "Evolution of the Air Campaign Planning Process and the Contingency Theater Automated Planning System (CTAPS)", RAND Corporation, MR-618-AF, 1996, https://www.rand.org/pubs/monograph_reports/MR618.html.

119 Hewitt and Beesley, 2024.

120 J. McCarver, "Agentic AI in Future of Military Operations", LinkedIn, 18 October 2024, https://www.linkedin.com/pulse/agentic-ai-future-military-operations-josh-mccarver-ypfec/.

121 K. Payne, "AI Arms and Influence: Frontier Models Exhibit Sophisticated Reasoning in Simulated Nuclear Crises", arXiv, 17 February 2026, https://arxiv.org/pdf/2602.14740.

122 R. Farnell and K. Coffey, "AI's New Frontier in War Planning: How AI Agents Can Revolutionize Military Decision-Making", Harvard Kennedy School Belfer Center for Science and International Affairs, 11 October 2024, https://www.belfercenter.org/research-analysis/ais-new-frontier-war-planning-how-ai-agents-can-revolutionize-military-decision.

123 J. Johnson, "Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command-and-Control Decision-Making in the Digital Age", *Defence Studies*, Vol. 23, No. 1, 2023, pp. 43-67, https://doi.org/10.1080/14702436.2022.2102486; J.R. Boyd, *The Essence of Winning and Losing*, September 2012 [original January 1996], https://slightlyeastofnew.com/wp-content/uploads/2010/03/essence_of_winning_losing.pdf.

124 Farnell and Coffey, 2024.

125 M. Chen, "Chinese Team Taps DeepSeek AI for Military Battle Simulation", *South China Morning Post*, 16 May 2025, https://www.scmp.com/news/china/military/article/3310707/chinese-team-taps-deepseek-ai-military-battle-simulation.

126 Payne, 2026.

127 K. Payne, "Building an AI Strategist", Ken's Substack, 29 September 2025, https://www.kennethpayne.uk/p/building-an-ai-strategist.

128 H. Lonas, "The Dawn of Agentic AI: Are We Ready for Autonomous Technology?", CIO, 14 March 2025, https://www.cio.com/article/3846227/the-dawn-of-agentic-ai-are-we-ready-for-autonomous-technology.html.

129 Jensen et al., 2025.

130 Farnell and Coffey, 2024.

131 J. Hornstein, "AI Agents Are Coming to the Military. VCs Love It, but Researchers Are a Bit Wary", *Business Insider*, 8 March 2025, https://www.businessinsider.com/ai-agents-coming-military-new-scaleai-contract-2025-3.

132 DIU (Defense Innovation Unit), "DIU's Thunderforge Project to Integrate Commercial AI-Powered Decision-Making for Operational and Theater-Level Planning", 5 March 2025, https://www.diu.mil/latest/dius-thunderforge-project-to-integrate-commercial-ai-powered-decision-making.

133 G. de Vynck, "Pentagon Signs AI Deal to Help Commanders Plan Military Maneuvers", *Washington Post*, 5 March 2025, https://www.washingtonpost.com/technology/2025/03/05/pentagon-ai-military-scale/.

134 Scale AI, 2026.

135 D. Sophia, "US Defense Department Awards Contracts to Google, Musk's xAI", Reuters, 14 July 2025, https://www.reuters.com/business/autos-transportation/us-department-defense-awards-contracts-google-xai-2025-07-14/.

136 Salesforce, "U.S. Army Awards Salesforce $5.6B Contract to Accelerate Military Modernization and Department of War Readiness", 26 January 2026, https://investor.salesforce.com/news/news-details/2026/U-S--Army-Awards-Salesforce-5-6B-Contract-to-Accelerate-Military-Modernization-and-Department-of-War-Readiness/default.aspx.

137 J.M. Friar and P. Payne, "Agentic Artificial Intelligence: Strategic Adoption in the U.S. Department of Defense", Cybersecurity & Information Systems Information Analysis Center, 3 September 2025, https://csiac.dtic.mil/technical-inquiries/notable/agentic-artificial-intelligence-strategic-adoption-in-the-u-s-department-of-defense/; Soufan Center, "Agentic AI: Does the Future of Warfare Look Autonomous?", 6 August 2025, https://thesoufancenter.org/intelbrief-2025-august-6/.

138 K. Bondar, "Ukraine's Future Vision and Current Capabilities for Waging AI-Enabled Autonomous Warfare", Center for Strategic & International Studies, 6 March 2025a, https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare.

139 C. Panella, "Artificial Intelligence Is Going to Make Drone Wars Much More Deadly. It's Already Started", *Business Insider*, 7 March 2025, https://www.businessinsider.com/ukraines-smart-drones-more-likely-hit-targets-2025-3.

140 S. Brown, "Ukraine Trained AI for Its 'Spiderweb' Airfield Drone Attacks at Aviation Museum", *Kyiv Post*, 2 June 2025, https://www.kyivpost.com/post/53784.

141 Cummings, 2026.

142 C.J. Chivers, "The Dawn of the A.I. Drone", *New York Times Magazine*, 31 December 2025, https://www.nytimes.com/2025/12/31/magazine/ukraine-ai-drones-war-russia.html; E. Rosenbach et al., "The Autonomous Arsenal in Defense of Taiwan: Technology, Law, and Policy of the Replicator Initiative", Harvard Kennedy School Belfer Center for Science and International Affairs, 3 February 2025, https://www.belfercenter.org/replicator-autonomous-weapons-taiwan.

143 Defense Express, "Obscure Russian V2U Drone Unraveled by Intelligence: Autonomous Loitering Munition Powered by Nvidia Chip", 9 June 2025, https://en.defence-ua.com/weapon_and_tech/obscure_russian_v2u_drone_unraveled_by_intelligence_autonomous_loitering_munition_powered_by_nvidia_chip-14798.html; A. Gosavi, "Russia Ups the UAV Warfare Game with 62-mile Range Attack Drone Swarms", Interesting Engineering, 10 November 2025, https://interestingengineering.com/military/russia-ups-uav-warfare-game.

144 R. Sapkota et al., "UAVs Meet Agentic AI: A Multidomain Survey of Autonomous Aerial Intelligence and Agentic UAVs", arXiv, 8 June 2025a, https://arxiv.org/abs/2506.08045.

145 B. Neely, "Agentic AI: The Rise of Thinking Machines", Forbes Technology Council, 12 March 2025, https://www.forbes.com/councils/forbestechcouncil/2025/03/12/agentic-ai-rise-of-the-thinking-machines/; Cummings, 2026.

146 X. Zhou et al., "Swarm of Micro Flying Robots in the Wild", *Science Robotics*, Vol. 7, No. 66, 4 May 2022, https://www.science.org/doi/10.1126/scirobotics.abm5954; For video see: https://www.youtube.com/watch?v=Lr7L2t-svJQ&t=2s.

147 Chivers, 2025.

148 J.-M. Rickli, "Surrogate Warfare and the Transformation of War in the 2020s", Geneva Centre for Security Policy, 18 May 2021, https://www.gcsp.ch/global-insights/surrogate-warfare-and-transformation-war-2020s.

149 A. Krieg and J.-M. Rickli, *Surrogate Warfare: The Transformation of War in the Twenty-first Century*, Georgetown University Press, 2019.

150 A. Banafa, "Agentic AI: The Rise of Autonomous Intelligence", *BBN Times*, 19 December 2024, https://www.bbntimes.com/science/agentic-ai-the-rise-of-autonomous-intelligence.

151 K. Boucher, "What Is Agentic AI and What Does It Mean for Physical Security?", LVT Blog, 5 March 2025, https://www.lvt.com/blog/what-is-agentic-ai-and-what-does-it-mean-for-physical-security.

152 O'Donnell, 2024.

153 K.V. Hiebert, "The United States Quietly Kick-Starts the Autonomous Weapons Era", Centre for International Governance Innovation, 15 January 2024, https://www.cigionline.org/articles/the-united-states-quietly-kick-starts-the-autonomous-weapons-era/; R. Manuel, "US Navy Forms Unmanned Operations Task Force in Middle East", *Defense Post*, 17 January 2024, https://thedefensepost.com/2024/01/17/us-unmanned-operations-middle-east/.

154 M. Schuler, "Wie das Silicon Valley die US-Verteidigungsindustrie revolutioniert", *Frankfurter Allgemeine Zeitung*, 12 March 2025, https://www.faz.net/pro/digitalwirtschaft/transformation/wie-das-silicon-valley-die-us-verteidigungsindustrie-revolutioniert-110346857.html.

155 B. Babaiev, "Inside Ukraine's TOLOKA Underwater Drones: Secret Combat Capabilities Revealed", RBC-Ukraine, 21 September 2025, https://newsukraine.rbc.ua/news/inside-ukraine-s-toloka-underwater-drones-1758473279.html.

156 M.C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power", *Texas National Security Review*, Vol. 1, No. 3, May 2018, pp. 36-57, https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/.

157 A. Paulus, "An Achilles Heel of Today's Armed Forces: Managing Software Supply Chain Risk in the Military Sector", SWP Research Paper 2025/RP 06, 17 November 2025, https://doi.org/10.18449/2025RP06.

158 A. Gilli and M. Gilli, "Military Technology: The Realities of Imitation", *CSS Analyses in Security Policy*, No. 238, February 2019, Center for Security Studies, ETH Zürich.

159 S. Waterman, "Air Force Using Generative AI to Help Modernize Legacy Software", *Air & Space Forces Magazine*, 11 April 2025, https://www.airandspaceforces.com/air-force-generative-ai-modernize-legacy-software/.

160 M. Pregasen, "What Is a Coding Agent? Comparing Agents to AI Code Assistants", OpenHands Blog, 21 January 2026, https://openhands.dev/blog/agentic-coding-vs-code-completion.

161 J. Zhang, "Why Agentic AI Is the Next Frontier of Generative AI", *Forbes*, 27 March 2025, https://www.forbes.com/councils/forbestechcouncil/2025/03/27/why-agentic-ai-is-the-next-frontier-of-generative-ai/.

162 V. Chen et al., "Code with Me or for Me? How Increasing AI Automation Transforms Developer Workflows", arXiv, 10 July 2025, https://arxiv.org/abs/2507.08149; N. Maloyan and D. Namiot, "Prompt Injection Attacks on Agentic Coding Assistants: A Systematic Analysis of Vulnerabilities in Skills, Tools, and Protocol Ecosystems", arXiv, 24 January 2026, https://arxiv.org/abs/2601.17548.

163 Anthropic, "Detecting and Countering Misuse of AI: August 2025", 27 August 2025, https://www.anthropic.com/news/detecting-countering-misuse-aug-2025.

164 Anthropic, "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign", 13 November 2025, https://www.anthropic.com/news/disrupting-AI-espionage.

165 N. Carlini et al., "Assessing Claude Mythos Preview's Cybersecurity Capabilities", Anthropic Frontier Red Team, 7 April 2026, https://red.anthropic.com/2026/mythos-preview/.

166 L.H. Newman, "Anthropic's Mythos Will Force a Cybersecurity Reckoning – Just Not the One You Think", Wired, 10 April 2026, https://www.wired.com/story/anthropics-mythos-will-force-a-cybersecurity-reckoning-just-not-the-one-you-think/.

167 AISI (AI Security Institute), "Our Evaluation of Claude Mythos Preview's Cyber Capabilities", 13 April 2026,

https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities.

168 Carlini et al., 2026.

169 Y. Bengio et al., *International AI Safety Report 2026*, 3 February 2026, https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026.

170 B. Rosen and J. Kraprayoon, "Cyberwar's New Frontier", *Foreign Affairs*, 16 April 2026, https://www.foreignaffairs.com/united-states/cyberwars-new-frontier.

171 R. Williams, "Cyberattacks by AI Agents Are Coming", *MIT Technology Review*, 4 April 2025, https://www.technologyreview.com/2025/04/04/1114228/cyberattacks-by-ai-agents-are-coming/; Martin Dion, discussion with authors.

172 A. Rashid et al., "Artificial Intelligence in the Military: An Overview of the Capabilities, Applications, and Challenges", *International Journal of Intelligent Systems*, 2023, pp. 1-31, https://doi.org/10.1155/2023/8676366.

173 M. Heikkilä and W.D. Heaven, "Anthropic's Chief Scientist on 4 Ways Agents Will Be Even Better in 2025", *MIT Technology Review*, 11 January 2025, https://www.technologyreview.com/2025/01/11/1109909/anthropics-chief-scientist-on-5-ways-agents-will-be-even-better-in-2025/.

174 A. Chhabra et al., "Agentic AI Security: Threats, Defenses, Evaluation, and Open Challenges", arXiv, 27 October 2025, https://arxiv.org/abs/2510.23883.

175 Scale AI, 2026.

176 B. Raghavan and B. Schneier, "Agentic AI's OODA Loop Problem", *IEEE Security & Privacy*, Vol. 23, November-December 2025, pp. 80-82, https://doi.org/10.1109/MSEC.2025.3604105.

177 E. Maor, "How Hackers Manipulate Agentic AI with Prompt Engineering", *Security Week*, 19 February 2025, https://www.securityweek.com/how-hackers-manipulate-agentic-ai-with-prompt-engineering/.

178 A. Sahu, "Harnessing the OODA Loop for Agentic AI: From Generative Foundations to Proactive Intelligence", Sogeti, 6 March 2025, https://www.sogeti.com/featured-articles/harnessing-the-ooda-loop-for-agentic-ai/.

179 J.-M. Rickli et al., "What, Why and When? A Review of the Key Issues in the Development and Deployment of Military Human–Machine Teams", Geneva Centre for Security Policy, February 2024, https://www.gcsp.ch/publications/what-why-and-when-review-key-issues-development-and-deployment-military-human-machine.

180 G. de Vynck, "Cheap, Smart, Deadly. The Tech Industry Pitches a New Way to Wage War", *Washington Post*, 21 January 2025, https://www.washingtonpost.com/technology/2025/01/21/anduril-startup-weapons-drones-ai/.

181 Ibid.

182 Jensen et al., 2025.

183 R.M. Jones et al., "Automated Intelligent Pilots for Combat Flight Simulation", *AI Magazine*, Vol. 20, No. 1, 1999, pp. 27-41, https://doi.org/10.1609/aimag.v20i1.1438.

184 Breaking Defense, "AI in the Loop: Transforming Military Training and System Design through Smarter Simulation", 20 October 2025, https://breakingdefense.com/2025/10/ai-in-the-loop-transforming-military-training-and-system-design-through-smarter-simulation/.

185 CERT-UA (Computer Emergency Response Team of Ukraine), "UAC-0001 Cyberattacks on the Security and Defense Sector Using the LAMEHUG Software Tool, Which Uses LLM (Large Language Model) (CERT-UA#16039)", 17 July 2025, https://cert.gov.ua/article/6284730.

186 M. Raz et al., "Ransomware 3.0: Self-Composing and LLM-Orchestrated", arXiv, 28 August 2025, https://arxiv.org/abs/2508.20444v1.

187 S. Sabin, "Exclusive: AI Will Supercharge Cyber Weapons within Two Years, Experts Warn", Axios, 7 January 2025a, https://www.axios.com/2025/01/07/goldilock-agentic-malware-2027-doomsday.

188 S. Sjouwerman, "How 'Agentic AI' Will Drive the Future of Malware", SC Media, 19 March 2025, https://www.scworld.com/perspective/how-agentic-ai-will-drive-the-future-of-malware.

189 Martin Dion, comments to authors referred to in text, 10 February 2026.

190 Ibid.

191 S. Sabin, "Malware's AI Time Bomb", Axios, 14 March 2025b, https://www.axios.com/2025/03/14/hackers-artificial-intelligence-cyber-threats.

192 Rickli, 2020.

193 M. Ekelhof and G. Persi Paoli, "Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems", United Nations Institute for Disarmament Research, 8 April 2020, https://unidir.org/wp-content/uploads/2023/05/UNIDIR-Swarm-Robotics-2020.pdf.

194 Martin Dion, comments to authors referred to in text.

195 P. Scharre, "The Perilous Coming Age of AI Warfare", *Foreign Affairs*, 29 February 2024, https://www.foreignaffairs.com/ukraine/perilous-coming-age-ai-warfare; Z. Kallenborn, "Meet the Future Weapon of Mass Destruction, the Drone Swarm", *Bulletin of the Atomic Scientists*, 5 April 2021, https://thebulletin.org/2021/04/meet-the-future-weapon-of-mass-destruction-the-drone-swarm/.

196 Bondar, 2025a; Chivers, 2025.

197 Defense Express, "Russia Drops Kamikaze Drone on Central Kyiv Using AI Swarm Tech – Debate Erupts Over 200km Range Mystery", 16 March 2026, https://en.defence-ua.com/news/russia_drops_kamikaze_drone_on_central_kyiv_using_ai_swarm_tech_debate_erupts_over_200km_range_mystery-17845.html.

198 Sapkota et al., 2025a; Martin Dion, discussion with authors.

199 A. Wei, "1 Soldier, 200 Drones: China Showcases Rapid Launch and Agility in Swarm Warfare Tactics", *South China Morning Post*, 23 January 2026, https://www.scmp.com/news/china/military/article/3340972/1-soldier-200-drones-china-showcases-rapid-launch-and-agility-swarm-warfare-tactics; D. Hambling, "Swarm Forge: Pentagon's Mass-Drone Test Signals Near-Term Deployment", *Forbes*, 28 January 2026, https://www.forbes.com/sites/davidhambling/2026/01/28/swarm-forge-pentagons-mass-drone-test-signals-near-term-deployment/.

200 Martin Dion, discussion with authors; T.M. Nguyen et al., "Agentic AI Meets Edge Computing in Autonomous UAV Swarms", arXiv, 20 January 2026, https://arxiv.org/abs/2601.14437.

201 J. Roger and D. Kunertova, "The Vulnerabilities of the Drone Age: Established Threats and Emerging Issues Out to 2035", ETH Zürich Center for Security Studies, June 2022, https://doi.org/10.3929/ethz-b-000556165.

202 M. Elgan, "Agentic AI Swarms Are Headed Your Way", *Computerworld*, 1 November 2024, https://www.computerworld.com/article/3594235/agentic-ai-swarms-are-headed-your-way.html.

203 K. Matthews and M. Lamensch, "Wired for War: How Authoritarian States Are Weaponizing AI against the West", Konrad-Adenauer-Stiftung e.V. Canada Office, 29 July 2025, https://migsinstitute.org/wp-content/uploads/2025/09/20250729_Wired-for-war-.pdf.

204 Chhabra et al., 2025; Martin Dion, discussion with authors.

205 D.T. Schroeder et al., "How Malicious AI Swarms Can Threaten Democracy", *Science*, Vol. 391, No. 6783, 22 January 2026, pp. 354-357, https://doi.org/10.1126/science.adz1697.

206 Martin Dion, discussion with authors.

207 J.-M. Rickli, "Dealing with the New WMDs – Weapons of Mass Disinformation", Geneva Policy Outlook, 20 January 2025, https://www.genevapolicyoutlook.ch/dealing-with-the-new-wmds-weapons-of-mass-disinformation/.

208 B. Giussani and Q. Ladetto, "Guerre cognitive: quand le cerveau devient un terrain de bataille", L'Atelier des Futurs, La Menace Cognitive No. 4, July 2025, https://atelierdesfuturs.org/guerre-cognitive/.

209 J.-M. Rickli and T. Knappe, "Enhancing Cognitive Security and Societal Resilience to Counter Cognitive Warfare", In Focus, Geneva Centre for Security Policy, 7 October 2025, https://www.gcsp.ch/publications/enhancing-cognitive-security-and-societal-resilience-counter-cognitive-warfare.

210 M.T. Della Mura, "From Biowarfare to Bioterrorism: The Future of Biological Threats in the AI Era", *Tech4Future*, 25 February 2025, https://tech4future.info/en/bioterrorism-biological-threats-in-the-ai-era/.

211 E.H. Soice et al., "Can Large Language Models Democratize Access to Dual-Use Biotechnology?", arXiv, 6 June 2023, https://arxiv.org/abs/2306.03809.

212 K.V. Hiebert, "AI Is Reviving Fears around Bioterrorism. What's the Real Risk?", Centre for International Governance Innovation, 30 June 2025, https://www.cigionline.org/articles/ai-is-reviving-fears-around-bioterrorism-whats-the-real-risk/.

213 Anthropic, "Activating AI Safety Level 3 Protections", 22 May 2025, https://www.anthropic.com/news/activating-asl3-protections.

214 K.M. Esvelt, "Foundation Models May Exhibit Staged Progression in Novel CBRN Threat Disclosure", arXiv, 19 March 2025, https://arxiv.org/abs/2503.15182.

215 K.M. Esvelt, "Delay, Detect, Defend: Preparing for a Future in Which Thousands Can Release New Pandemics", Geneva Paper, Issue 29, Geneva Centre for Security Policy, November 2022, https://www.gcsp.ch/publications/delay-detect-defend-preparing-future-which-thousands-can-release-new-pandemics.

216 J. Götting et al., "Virology Capabilities Test (VCT): A Multimodal Virology Q&A Benchmark", arXiv, 21 April 2025, https://arxiv.org/abs/2504.16137.

217 S.R. Carter, "'Agentic' Life Sciences AI Is Exacerbating Bioweapons Concerns. Here's What to Do About It", *Bulletin of the Atomic Scientists*, 26 February 2026, https://thebulletin.org/2026/02/agentic-life-sciences-ai-is-exacerbating-bioweapons-concerns-heres-what-to-do-about-it/; J. Revill et al. "What Will Be the Impact of AI on the Bioweapons Treaty?", *Bulletin of the Atomic Scientists*, 16 November 2024, https://thebulletin.org/2024/11/what-will-be-the-impact-of-ai-on-the-bioweapons-treaty/.

218 R. Chaves de Lima et al., "Artificial Intelligence Challenges in the Face of Biological Threats: Emerging Catastrophic Risks for Public Health", *Frontiers*, 10 May 2024, https://doi.org/10.3389/frai.2024.1382356; J.-M. Rickli and G. Vllasi, "The Weaponization of Emerging Technologies and Their Impact on Global Risk: A Perspective from the PfPC Emerging Security Challenges Working Group", *Connections: The Quarterly Journal*, Vol. 24, No. 1, January 2025, https://procon.bg/article/weaponization-emerging-technologies-and-their-impact-global-risk-perspective-pfpc-emerging.

219 National Academies of Sciences, Engineering, and Medicine, "AI-Enabled Biological Design and the Risks of Synthetic Biology", in *The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations*, Consensus Study Report, Washington, DC, 2025, https://doi.org/10.17226/28868.

220 F. Lentzos and G. Bowsher, "Distributed Capability, Shared Vulnerability: International Governance of Biosafety and Biosecurity", *Cold Spring Harbor Perspectives in Medicine*, 17 February 2026, https://doi.org/10.1101/cshperspect.a041627.

221 R. Moulange, "AI Agents Will Upset Plans to Safeguard Narrow AI-Bio tools", Securing the Interface Substack, 15 September 2025, https://richardmoulange.substack.com/p/ai-agents-will-upset-plans-to-safeguard.

222 A. Wang, "Alex Wang on Why China Can't Be Allowed to Dominate AI-based Warfare", *The Economist*, 4 March 2025, https://www.economist.com/by-invitation/2025/03/04/alex-wang-on-why-china-cant-be-allowed-to-dominate-ai-based-warfare.

223 B. Jensen and M. Strohmeyer, "Agentic Warfare and the Future of Military Operations", Center for Strategic & International Studies, 17 July 2025, https://www.csis.org/analysis/rethinking-napoleonic-staff.

224 K.M. Sayler and C.A. Theohary, "Agentic Artificial Intelligence and Cyberattacks", Congress.gov, 3 February 2026, https://www.congress.gov/crs-product/IF13151.

225 US DoW (US Department of War), "War Department Launches AI Acceleration Strategy to Secure American Military AI Dominance", 12 January 2026, https://www.war.gov/News/Releases/Release/Article/4376420/war-department-launches-ai-acceleration-strategy-to-secure-american-military-ai/.

226 A.H.-E. Wang et al., "The People's Liberation Army's Perspectives on Artificial Intelligence: Highlighting Integration as Key to 'Intelligentization' Goals", RAND Corporation, 4 March 2026, https://www.rand.org/pubs/perspectives/PEA4574-1.html.

227 R. Barrett-Taylor and N. Karner, "AI Won't Replace the General: Algorithms, Decision-making and Battlefield Command", Alan Turing Institute, September 2025, https://www.turing.ac.uk/news/publications/ai-wont-replace-general-algorithms-decision-making-and-battlefield-command.

228 T.C. Bächle and J. Bareis, "Autonomous Weapons as a Geopolitical Signifier in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies", *European Journal of Futures Research*, Vol. 10, No. 20, 2022, https://doi.org/10.1186/s40309-022-00202-w; I. Bode et al., "Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US Practices", *Ethics and Information Technology*, Vol. 25, No. 5, 2023, https://doi.org/10.1007/s10676-023-09678-x.

229 Jensen and Strohmeyer, 2025.

230 Purdy, 2024.

231 Bandi et al., 2025.

232 D. Barman et al., "The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination", *Machine Learning with Applications*, Vol. 16, June 2024, https://doi.org/10.1016/j.mlwa.2024.100545.

233 Raghavan and Schneier, 2025.

234 M. Jancer, "Over 50% of New Online Articles Are Being Cranked Out by AI", Vice, 17 October 2025, https://www.vice.com/en/article/more-than-half-new-articles-ai-written/.

235 W. Knight, "OpenAI's Deep Research Agent Is Coming for White-Collar Work", Wired, 19 March 2025, https://www.wired.com/story/openais-deep-research-agent-is-coming-for-white-collar-work/.

236 A. Newport and N. Jankowicz, "Russian Networks Flood the Internet with Propaganda, Aiming to Corrupt AI Chatbots", *Bulletin of the Atomic Scientists*, 26 March 2025, https://thebulletin.org/2025/03/russian-networks-flood-the-internet-with-propaganda-aiming-to-corrupt-ai-chatbots/; M. Sadeghi and I. Blachez, *A Well-Funded Moscow-Based Global 'News' Network has Infected Western Artificial Intelligence Tools Worldwide with Russian Propaganda*, Newsguard, 6 March 2025, https://www.newsguardtech.com/special-reports/moscow-based-global-news-network-infected-western-artificial-intelligence-russian-propaganda/.

237 Y. Bathaee, "The Artificial Intelligence Black Box and the Failure of Intent and Causation", *Harvard Journal of Law & Technology*, Vol. 31, No. 2, Spring 2018, https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf.

238 W.D. Heaven, "Mechanistic Interpretability: New Techniques Are Giving Researchers a Glimpse at the Inner Workings of AI Models", *MIT Technology Review*, 12 January 2026, https://www.technologyreview.com/2026/01/12/1130003/mechanistic-interpretability-ai-research-models-2026-breakthrough-technologies/.

239 Anthropic, "Tracing the Thoughts of a Large Language Model", 27 March 2025, https://www.anthropic.com/research/tracing-thoughts-language-model.

240 Tomašev et al., 2026.

241 T. South et al., "Authenticated Delegation and Authorized AI Agents", arXiv, 16 January 2025, https://arxiv.org/abs/2501.09674v1.

242 Anthropic, "Agentic Misalignment: How LLMs Could Be Insider Threats", 20 June 2025, https://www.anthropic.com/research/agentic-misalignment.

243 Hiebert, 2024.

244 R. Khan et al., "Security Threats in Agentic AI Systems", arXiv, 16 October 2024, https://arxiv.org/abs/2410.14728.

245 T. Claburn, "You Know that Generative AI Browser Assistant Extension Is Probably Beaming Everything to the Cloud, Right?", The Register, 25 March 2025, https://www.theregister.com/2025/03/25/generative_ai_browser_extensions_privacy/.

246 S.K. Sahib and A. Chaikin, "Unseeable Prompt Injections in Screenshots: More Vulnerabilities in Comet and Other AI Browsers", Brave.com, 21 October 2025, https://brave.com/blog/unseeable-prompt-injections/.

247 H. Chen and K. Magramo, "Finance Worker Pays out $25 Million after Video Call with Deepfake 'Chief Financial Officer'", CNN, 4 February 2024, https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk.

248 Martin Dion, comments to authors referred to in text.

249 A. Bunn, "How Agentic AI Will Be Weaponized for Social Engineering Attacks", McAfee Blog, 17 November 2025, https://www.mcafee.com/blogs/internet-security/how-agentic-ai-will-be-weaponized-for-social-engineering-attacks/; B. Claverie, "Cognitive Warfare: The New Battlefield Exploiting Our Brains", *Polytechnique Insights*, 5 February 2025, https://www.polytechnique-insights.com/en/columns/geopolitics/cognitive-warfare-the-new-battlefield-exploiting-our-brains/.

250 A. Moix et al., *Threat Intelligence Report: August 2025*, Anthropic, 27 August 2025, https://www-cdn.anthropic.com/b2a76c6f6992465c09a6f2fce282f6c0cea8c200.pdf.

251 R. Sapkota et al., "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges", arXiv, 15 May 2025b, https://arxiv.org/abs/2505.10468.

252 Acharya et al., 2025.

253 Y. Xiao et al., "LIMI: Less Is More for Agency", arXiv, 22 September 2025, https://arxiv.org/abs/2509.17567; P. Belcak et al., "Small Language Models Are the Future of Agentic AI", arXiv, 2 June 2025, https://arxiv.org/abs/2506.02153; R. Bellan, "In 2026, AI Will Move from Hype to Pragmatism", TechCrunch, 2 January 2026, https://techcrunch.com/2026/01/02/in-2026-ai-will-move-from-hype-to-pragmatism/.

254 K. Hackenburg et al., "The Levers of Political Persuasion with Conversational AI", arXiv, 18 July 2025, https://arxiv.org/abs/2507.13919.

255 P. Sharma et al., "Agentic AI and Multiagent AI Systems: Design Considerations", *Wall Street Journal–CIO Journal*, 18 March 2025, https://deloitte.wsj.com/cio/agentic-ai-and-multiagent-ai-systems-design-considerations-bd02527a.

256 J. Hidary, "Non-human Identities: Agentic AI's New Frontier of Cybersecurity Risk", World Economic Forum, 15 October 2025, https://www.weforum.org/stories/2025/10/non-human-identities-ai-cybersecurity/.

257 Chhabra et al., 2025.

258 H. Wong and T. Saade, "The Rise of AI Agents: Anticipating Cybersecurity Opportunities, Risks, and the Next Frontier", R Street, 29 May 2025, https://www.rstreet.org/research/the-rise-of-ai-agents-anticipating-cybersecurity-opportunities-risks-and-the-next-frontier/.

259 Tomašev et al., 2026.

260 M. Lupinacci et al., "The Dark Side of LLMs: Agent-based Attacks for Complete Computer Takeover", arXiv, 9 July 2025, https://arxiv.org/abs/2507.06850.

261 D. Casey, "Why DeepSeek's Arrival Looks Like a Key Moment for Safe AI", Fast Company, 13 March 2025, https://www.fastcompany.com/91297727/why-deepseeks-arrival-looks-like-a-key-moment-for-safe-ai.

262 Martin Dion, discussion with authors.

263 Hidary, 2025.

264 T. South, "Identity Management for Agentic AI", OpenID, October 2025, https://openid.net/wp-content/uploads/2025/10/Identity-Management-for-Agentic-AI.pdf.

265 K. Huang et al., "A Novel Zero-Trust Identity Framework for Agentic AI: Decentralized Authentication and Fine-Grained Access Control", arXiv, 25 May 2025, https://arxiv.org/abs/2505.19301.

266 S. Shekkizhar et al., "Echoing: Identity Failures When LLM Agents Talk to Each Other", arXiv, 12 November 2025, https://arxiv.org/abs/2511.09710.

267 B. Klein et al., "Deploying Agentic AI with Safety and Security: A Playbook for Technology Leaders", *McKinsey Quarterly*, 16 October 2025, https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders.

268 Huang et al., 2025.

269 Deloitte, "Agentic AI Is Set to Change How Business Gets Done", *Deloitte Insights*, 17 March 2025, https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends/2025/servicenow-and-agentic-ai-set-to-change-how-business-gets-done.html.

270 M. Cemri et al., "Why Do Multi-Agent LLM Systems Fail?", arXiv, 17 March 2025, https://arxiv.org/abs/2503.13657.

271 J. Wallin, "Lessons in Learning", Center for a New American Security, 8 May 2025, https://www.cnas.org/publications/reports/lessons-in-learning.

272 J.-M. Rickli and F. Mantellassi, "Human–Machine Teaming in Artificial Intelligence Driven Air Power", *Air Power Journal*, Fall 2022, https://theairpowerjournal.com/human-machine-teaming-in-artificial-intelligence-driven-air-power-future-challenges-and-opportunities-for-the-air-force/.

273 S. Kaufman, "Beyond ChatGPT: The Rise of Agentic AI and Its Implications for Security", CSO Security Council, 22 October 2024, https://www.csoonline.com/article/3574697/beyond-chatgpt-the-rise-of-agentic-ai-and-its-implications-for-security.html.

274 C.C. Walther, "The Silent Erosion: How AI's Helping Hand Weakens Our Mental Grip", Centre for International Governance Innovation, 17 July 2025, https://www.cigionline.org/articles/the-silent-erosion-how-ais-helping-hand-weakens-our-mental-grip/.

275 J.H. Saltzer and M.F. Kaashoek, *Principles of Computer System Design – An Introduction*, Elsevier, 2009, pp. 1-42.

276 L. Hammond et al., "Multi-Agent Risks from Advanced AI", arXiv, 19 February 2025, https://arxiv.org/abs/2502.14143.

277 Wallin, 2025.

278 K. Zhu et al., "Where LLM Agents Fail and How They Can Learn from Failures", arXiv, 29 September 2025, https://arxiv.org/abs/2509.25370.

279 G. Huckins, "Are We Ready to Hand AI Agents the Keys?", *MIT Technology Review*, 12 June 2025, https://www.technologyreview.com/2025/06/12/1118189/ai-agents-manus-control-autonomy-operator-openai/.

280 I. Poirier, "High-Frequency Trading and the Flash Crash: Structural Weaknesses in the Securities Markets and Proposed Regulatory Responses", *UC Law Business Journal*, Vol. 8, No. 2, Summer 2012, https://repository.uclawsf.edu/hastings_business_law_journal/vol8/iss2/5.

281 A. Masumori and T. Ikegami, "Do Large Language Model Agents Exhibit a Survival Instinct? An Empirical Study in a Sugarcape-Style Simulation", arXiv, 18 August 2025, https://arxiv.org/abs/2508.12920.

282 B. Edwards, "Research AI Model Attempts to Modify Experiment Code to Extend Runtime", *Ars Technica*, 14 August 2024, https://arstechnica.com/information-technology/2024/08/research-ai-model-unexpectedly-modified-its-own-code-to-extend-runtime/.

283 Hammond et al., 2025.

284 H. Booth, "When AI Thinks It Will Lose, It Sometimes Cheats, Study Finds", *TIME*, 19 February 2025, https://time.com/7259395/ai-chess-cheating-palisade-research/.

285 Payne, 2026.

286 Cowin, 2024; P. Schoenegger et al., "When Large Language Models are More Persuasive Than Incentivized Humans, and Why", arXiv, 14 May 2025, https://arxiv.org/abs/2505.09662.

287 The White House, "Winning the Race – America's AI Action Plan", Executive Office of the President of the United States, 10 July 2025, https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

288 L. Podda, "China's Drive to Dominate the AI Race", Atlas Institute for International Affairs, 14 April 2025, https://atlasinstitute.org/chinas-drive-to-dominate-the-ai-race/.

289 A. Hawkins, "Who Is Behind DeepSeek and How Did It Achieve Its AI 'Sputnik Moment'?", *The Guardian*, 28 January 2025, https://www.theguardian.com/technology/2025/jan/28/who-is-behind-deepseek-and-how-did-it-achieve-its-ai-sputnik-moment.

290 C. Chen, "Everyone in AI Is Talking about Manus. We Put It to the Test.", *MIT Technology Review*, 11 March 2025a, https://www.technologyreview.com/2025/03/11/1113133/manus-ai-review/.

291 The White House, "Fact Sheet: President Donald J. Trump Strikes Deal on Economic and Trade Relations with China", 1 November 2025, https://www.whitehouse.gov/fact-sheets/2025/11/fact-sheet-president-donald-j-trump-strikes-deal-on-economic-and-trade-relations-with-china/; D. Wei, "Strategic Decoupling and Its Implications for US-China Relations", RSIS Commentary CO25183, 1 September 2025, https://rsis.edu.sg/rsis-publication/rsis/strategic-decoupling-and-its-implications-for-us-china-relations/.

292 B. Williams, "The Agentic AI Revolution: How 2026 Will Reshape Technology and Statecraft", The National Interest Blog, 2 February 2026, https://nationalinterest.org/blog/techland/the-agentic-ai-revolution-how-2026-will-reshape-technology-and-statecraft.

293 Netherlands, "Update on Invoking Goods Availability Act", Ministry of Economic Affairs, 19 November 2025, https://www.government.nl/binaries/government/documenten/diplomatic-statements/2025/11/19/update-on-invoking-goods-availability-act/Update+on+invoking+Goods+Availability+Act+-+19+november+2025.pdf.

294 N. Aliyev, "Artificial Intelligence in Digital Silk Road: Driving Innovation and Economic Transformation", *EuroAsia Journal of Social Sciences & Humanities*, Vol. 12, No. 1, 2025, pp. 95-102, https://euroasiajournal.com/index.php/eurssh/article/view/510.

295 K. Chayka, "Elon Musk's A.I.-Fuelled War on Human Agency", *New Yorker*, 12 February 2025, https://www.newyorker.com/culture/infinite-scroll/elon-musks-ai-fuelled-war-on-human-agency.

296 The White House, "Removing Barriers to American Leadership in Artificial Intelligence", 23 January 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

297 M. Schuman, "The Race for Global Domination in AI", *The Atlantic*, 3 January 2026, https://www.theatlantic.com/international/2026/01/china-ai-competition-differences/685389/.

298 R. Rohozinski and C. Spirito, "Weaponising AI: The New Cyber Attack Surface", *Survival*, Vol. 68, No. 1, February-March 2026, pp. 7-18, https://doi.org/10.1080/00396338.2026.2620282.

299 M. Santora et al., "A Thousand Snipers in the Sky: The New War in Ukraine", *New York Times*, 3 March 2025, https://www.nytimes.com/interactive/2025/03/03/world/europe/ukraine-russia-war-drones-deaths.html.

300 J. Fowler, "Ukraine Drones Hit Deep Inside Russia", International Intrigue, 3 June 2025, https://archives.internationalintrigue.io/p/ukraine-drones-hit-deep-inside-russia-c5bbbb6633f8e757; K. Bondar, "How Ukraine's Operation 'Spider's Web' Redefines Asymmetric Warfare", Center for Strategic & International Studies, 2 June 2025b, https://www.csis.org/analysis/how-ukraines-spider-web-operation-redefines-asymmetric-warfare.

301 US Department of Defense, "Deputy Secretary of Defense Kathleen Hicks Keynote Address: 'The Urgency to Innovate' (as Delivered)", 28 August 2023, https://www.defense.gov/News/Speeches/Speech/Article/3507156/deputy-secretary-of-defense-kathleen-hicks-keynote-address-the-urgency-to-innov/.

302 R. Haley et al., "Is Replicator Replicable?", Harvard Kennedy School Belfer Center for Science and International Affairs, 25 September 2025, https://www.belfercenter.org/research-analysis/replicator-replicable.

303 S. Holliday et al., "U.S. Military Is Struggling to Deploy AI Weapons", *Wall Street Journal*, 26 September 2025, https://www.wsj.com/politics/national-security/pentagon-ai-weapons-delay-0f560d7e.

304 S. Oesch et al., "Agentic AI and the Cyber Arms Race", *Computer*, Vol. 58, No. 5, May 2025, pp. 82-85, https://ieeexplore.ieee.org/document/10970193.

305 Scale AI, 2026; D. Petraeus and I.C. Flanagan, "The Autonomous Battlefield", *Foreign Affairs*, 12 March 2026, https://www.foreignaffairs.com/middle-east/autonomous-battlefield.

306 M.C. Horowitz, "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability", *Journal of Strategic Studies*, Vol. 42, No. 6, 22 August 2019, pp. 764-788, https://doi.org/10.1080/01402390.2019.1621174; J.-M. Rickli, "The Destabilizing Prospects of Artificial Intelligence for Nuclear Strategy, Deterrence and Stability", in V. Boulanin (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: European Perspectives*, Vol. I, Stockholm, Stockholm International Peace Research Institute, 2019, pp. 91-98.

307 C.S. Gray, *Another Bloody Century: Future Warfare*, London, Phoenix, 2006.

308 T. Papandreou, "2025: Agentic and Physical AI – a Multitrillion Dollar Economy Emerges", *Forbes*, 15 January 2025, https://www.forbes.com/sites/timothypapandreou/2025/01/15/2025-agentic--physical-aia-multi-trillion-dollar-economy-emerges/.

309 M. Lu, "Visualizing Global AI Investment by Country", Visual Capitalist, 21 April 2025, https://www.visualcapitalist.com/visualizing-global-ai-investment-by-country/.

310 M. Sharp et al., "Agentic Inequality", arXiv, 21 October 2025, https://arxiv.org/abs/2510.16853.

311 D. Acemoglu, "Two Models for Agentic AI", Project Syndicate, 28 March 2025, https://www.project-syndicate.org/commentary/ai-agents-promising-as-advisers-but-problematic-as-autonomous-decision-makers-by-daron-acemoglu-2025-03.

312 C. Chen, "When AIs Bargain, a Less Advanced Agent Could Cost You", *MIT Technology Review*, 17 June 2025b, https://www.technologyreview.com/2025/06/17/1118910/ai-price-negotiation/.

313 A. Goswami, "Agentic AI: The Next Frontier in Artificial Intelligence", Firstpost, 25 March 2025, https://www.firstpost.com/opinion/agentic-ai-the-next-frontier-in-artificial-intelligence-13874369.html.

314 Acharya et al., 2025.

315 A. Satariano and P. Mozur, "The Global A.I. Divide: Where A.I. Data Centers Are Located", *New York Times*, 21 June 2025, https://www.nytimes.com/interactive/2025/06/23/technology/ai-computing-global-divide.html.

316 T. Nugraha, "How Agentic AI Is Reshaping the Global South: Opportunities & Risks", Modern Diplomacy, 9 March 2025, https://moderndiplomacy.eu/2025/03/09/how-agentic-ai-is-reshaping-the-global-south-opportunities-risks/.

317 Sharp et al., 2025.

318 D. Gartenstein-Ross et al., "Virtual Plotters. Drones. Weaponized AI?: Violent Non-State Actors as Deadly Early Adopters", *Valens Global International Strategies & Security*, November 2019, https://valensglobal.com/virtual-plotters-drones-weaponized-ai-violent-non-state-actors-as-deadly-early-adopters/.

319 J.-M. Rickli and C. Liang, "New and Emerging Technologies for Terrorists", in M. Abrahms (ed.), *The Routledge Companion to Terrorism Studies: New Perspectives and Topics*, Routledge, 2024.

320 INTERPOL, "INTERPOL Financial Fraud Assessment: A Global Threat Boosted by Technology", 11 March 2024, https://www.interpol.int/en/News-and-Events/News/2024/INTERPOL-Financial-Fraud-assessment-A-global-threat-boosted-by-technology.

321 INTERPOL, "INTERPOL Global Financial Fraud Assessment", May 2024, https://www.interpol.int/ar/content/download/21077/file/24COM005563-01%20-%20CAS_Global%20Financial%20Fraud%20Assessment_Public%20version_2024-03%20v2.pdf.

322 Rohozinski and Spirito, 2026.

323 Moix et al., 2025.

324 W.D. Heaven, "Moltbook Was Peak AI Theater", *MIT Technology Review*, 6 February 2026, https://www.technologyreview.com/2026/02/06/1132448/moltbook-was-peak-ai-theater/.

325 S. Schechner and G. Wells, "When AI Bots Start Bullying Humans, Even Silicon Valley Gets Rattled", *Wall Street Journal*, 13 February 2026, https://www.wsj.com/tech/ai/when-ai-bots-start-bullying-humans-even-silicon-valley-gets-rattled-0adb04f1.

326 C. Agarwal and J. Leung, "Preventing AI Agents from Going Rogue", Palo Alto Networks Blog, 4 November 2025, https://www.paloaltonetworks.com/blog/network-security/preventing-ai-agents-from-going-rogue/.

327 O. Barbi et al., "Preventing Rogue Agents Improves Multi-Agent Collaboration", arXiv, 9 February 2025, https://arxiv.org/abs/2502.05986.

328 Zhang et al., 2026.

329 J. Kraprayoon et al., "Highly Autonomous Cyber-Capable Agents: Anticipating Capabilities, Tactics, and Strategic Implications", Institute for AI Policy and Strategy, 12 March 2026, https://arxiv.org/pdf/2603.11528.

330 V. Boulanin et al., "Before It's Too Late: Why a World of Interacting AI Agents Demands New Safeguards", SIPRI, 1 October 2025, https://www.sipri.org/commentary/essay/2025/its-too-late-why-world-interacting-ai-agents-demands-new-safeguards.

331 Rickli and Vllasi, 2025; M.K. Cohen et al., "Regulating Advanced Artificial Agents", *Science*, Vol. 384, No. 6691, pp. 36-38, 4 April 2024, science.org/doi/10.1126/science.adl0625.

332 K. Payne and B. Alloui-Cros, "Strategic Intelligence in Large Language Models: Evidence from Evolutionary Game Theory", arXiv, 3 July 2025, https://arxiv.org/abs/2507.02618.

333 Payne, 2026.

334 CSIS (Center for Strategic & International Studies), "Critical Foreign Policy Decisions Benchmark", 2025, https://www.csis.org/programs/futures-lab/projects/critical-foreign-policy-decisions-benchmark.

335 Acemoglu, 2025.

336 B. Jensen, "The Troubling Truth about How AI Agents Act in a Crisis", *Foreign Policy*, 4 March 2025, https://foreignpolicy.com/2025/03/04/ai-bias-national-security-study/; B. Jensen et al., "Critical Foreign Policy Decisions (CFPD)-Benchmark: Measuring Diplomatic Preferences in Large Language Models", arXiv, 8 March 2025, https://arxiv.org/abs/2503.06263.

337 Ibid.

338 Shavit et al., 2023.

339 Cummings, 2026.

340 M. Burgess, "Robots Are Fighting Robots in Russia's War in Ukraine", Wired, 30 January 2024, https://www.wired.com/story/robots-are-fighting-robots-in-russias-war-in-ukraine/.

341 P. Scharre, *Army of None: Autonomous Weapons and the Future of War*, W.W. Norton, 2018.

342 Scharre, 2024.

343 A. Holland Michel, "Europe's Drone-Filled Vision for the Future of War", *MIT Technology Review*,

6 January 2026, https://www.technologyreview.com/2026/01/06/1129737/autonomous-warfare-europe-drones-defense-automated-kill-chains/.

344 J.-P. Rivera et al., "Escalation Risks from Language Models in Military and Diplomatic Decision-Making", arXiv, 7 January 2024, https://arxiv.org/abs/2401.03408.

345 Martin Dion, discussion with authors.

346 H. Field, "OpenAI Launches ChatGPT Gov for U.S. Government Agencies", CNBC, 28 January 2025, https://www.cnbc.com/2025/01/28/openai-launches-chatgpt-gov-for-us-government-agencies.html.

347 Rivera et al., 2024.

348 E.M. Bender and A. Hanna, "Government Officials Are Letting AI Do Their Jobs. Badly", *Bulletin of the Atomic Scientists*, 30 May 2025, https://thebulletin.org/2025/05/government-officials-are-letting-ai-do-their-jobs-badly/; EAISF (European Artificial Intelligence & Society Fund), "How AI-driven Welfare Systems Are Deepening Inequality and Poverty across Europe", 16 July 2025, https://europeanaifund.org/newspublications/how-ai-driven-welfare-systems-are-deepening-inequality-and-poverty-across-europe/.

349 Chan et al., 2023.

350 S. Timcke, "How Agentic AI Challenges Democracy", TransformingSociety, 4 March 2025, https://www.transformingsociety.co.uk/2025/03/04/how-agentic-ai-challenges-democracy/.

351 S. Lee Myers, "Once a Sheriff's Deputy in Florida, Now a Source of Disinformation from Russia", *New York Times*, 29 May 2024, https://www.nytimes.com/2024/05/29/business/mark-dougan-russia-disinformation.html.

352 E. Yayboke, "Channeling Augustus: On Agentic Offensive Information Operations", Center for Strategic & International Studies, 19 September 2025, https://www.csis.org/analysis/channeling-augustus-agentic-offensive-information-operations.

353 Schroeder et al., 2026.

354 K. Tseng et al., "An Agentic Operationalization of DISARM for FIMI Investigation on Social Media", arXiv, 21 January 2026, https://arxiv.org/abs/2601.15109; A.-A. Avram et al., "MCP-Orchestrated Multi-Agent System for Automated Disinformation Detection", arXiv, 13 August 2025, https://arxiv.org/abs/2508.10143.

355 W. Frick, "The AI Hiring Pause Is Officially Here", Bloomberg Law, 17 May 2025, https://news.bloomberglaw.com/artificial-intelligence/the-ai-hiring-pause-is-officially-here.

356 J. Del Rey, "Amazon's Layoffs and Leaked AI Plans Beg the Question: Is the Era of Robot-Driven Unemployment upon Us?", *Fortune*, 25 November 2025, https://fortune.com/2025/11/25/amazon-layoffs-artificial-intelligence-robots-unemployment-automation/; K. Weise, "Amazon Plans to Replace More than Half a Million Jobs with Robots", *New York Times*, 21 October 2025, https://www.nytimes.com/2025/10/21/technology/inside-amazons-plans-to-replace-workers-with-robots.html.

357 S.M. Hosseini Maasoum and G. Lichtinger, "Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data", SSRN, 8 September 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5425555.

358 E. Brynjolfsson et al., "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence", Stanford Digital Economy Lab, 13 November 2025, https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/.

359 GCSP (Geneva Centre for Security Policy), "The Polymath Initiative", n.d., accessed 7 April 2026, https://www.gcsp.ch/the-polymath-initiative.

360 K.J.K. Feng et al., "Levels of Autonomy for AI Agents", arXiv, 14 June 2025, https://arxiv.org/abs/2506.12469.

361 M. Osmond and T. Jego, "Mind the Gap: How the Technical Mechanism of Agentic AI Outpace Global Legal Frameworks", arXiv, 28 March 2026, https://arxiv.org/abs/2603.27075.

362 ICRC (International Committee of the Red Cross), "Article 36 – New Weapons", in Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, International Humanitarian Law Databases, https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-36.

363 T. Vestner and A. Rossi, "Legal Reviews of War Algorithms", *International Law Studies*, Vol. 97, 2021, p. 509, https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2963&context=ils.

364 M. Mitchell et al., "Fully Autonomous AI Agents Should Not Be Developed", arXiv, 4 February 2025, https://doi.org/10.48550/arXiv.2502.02649.

365 Cummings, 2026.

366 IMDA (Infocomm Media Development Authority), "Singapore Launches New Model AI Governance Framework for Agentic AI", 22 January 2026, https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2026/new-model-ai-governance-framework-for-agentic-ai.

367 M.C. Horowitz and L. Kahn, "Military AI Adoption Is Outpacing Global Cooperation", Council on Foreign Relations, 11 February 2026, https://www.cfr.org/articles/military-ai-adoption-is-outpacing-global-cooperation.

368 B. Marijan, "Agentic Warfare and the Role of the Human", Project Ploughshares, 17 June 2025, https://ploughshares.ca/agentic-warfare-and-the-role-of-the-human/.

Geneva Papers Research Series

No.1 2011 G. P. Herd, “The Global Puzzle: Order in an Age of Primacy, Power-Shifts and Interdependence”, 34p.

No.2 2011 T. Tardy, “Cooperating to Build Peace: The UN-EU Inter-Institutional Complex”, 36p.

No.3 2011 M.-M. Ould Mohamedou, “The Rise and Fall of Al Qaeda: Lessons in Post-September 11 Transnational Terrorism”, 39p.

No.4 2011 A. Doss, “Great Expectations: UN Peacekeeping, Civilian Protection and the Use of Force”, 43p.

No.5 2012 P. Cornell, “Regional and International Energy Security Dynamics: Consequences for NATO’s Search for an Energy Security Role”, 43p.

No.6 2012 M.-R. Djalili and T. Kellner, “Politique Régionale de l’Iran: Potentialités, Défis et Incertitudes”, 40p.

No.7 2012 G. Lindstrom, “Meeting the Cyber Security Challenge”, 39p.

No.8 2012 V. Christensen, “Virtuality, Perception and Reality in Myanmar’s Democratic Reform”, 35p.

No.9 2012 T. Fitschen, “Taking the Rule of Law Seriously”, 30p.

No.10 2013 E. Kienle, “The Security Implications of the Arab Spring”, 32p.

No.11 2013 N. Melzer, “Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare”, 75p.

No.12 2013 A. Guidetti et al., “World Views: Negotiating the North Korean Nuclear Issue”, 47p.

No.13 2013 T. Sisk and M.-M. Ould Mohamedou, “Bringing Back Transitology: Democratisation in the 21st Century”, 36p.

No.14 2015 H. J. Roth, “The Dynamics of Regional Cooperation in Southeast Asia”, 35p.

No.15 2015 G. Galice, “Les Empires en Territoires et Réseaux”, 42p.

No.16 2015 S. C. P. Hinz, “The Crisis of the Intermediate-range Nuclear Forces Treaty in the Global Context”, 36p.

No.17 2015 H. J. Roth, “Culture – An Underrated Element in Security Policy”, 40p.

No.18 2016 D. Esfandiary and M. Finaud, “The Iran Nuclear Deal: Distrust and Verify”, 44p.

No.19 2016 S. Martin, “Spying in a Transparent World: Ethics and Intelligence in the 21st Century”, 42p.

No.20 2016 A. Burkhalter, “Définir le Terrorisme: Défis et Pratiques”, 50p.

No.21 2017 M. Finaud, “‘Humanitarian Disarmament’: Powerful New Paradigm or Naive Utopia?”, 48p.

No.22 2017 S. Aboul Enein, “Cyber Challenges in the Middle East”, 49p.

No.23 2019 Tobias Vestner, “Prohibitions and Export Assessment: Tracking Implementation of the Arms Trade Treaty”, 28p.

No.24 2019 Mathias Bak, Kristoffer Nilaus Tarp and Dr. Christina Schori Liang, “Defining the Concept of ‘Violent Extremism’”, 32p.

No.25 2020 Cholpon Orozobekova and Marc Finaud, “Regulating and Limiting the Proliferation of Armed Drones: Norms and Challenges”, 47p.

No.26 2020 Dr Gervais Rufyikiri, “Reshaping Approaches to Sustainable Peacebuilding and Development in Fragile States – Part I: Nexus between Unethical Leadership and State Fragility”, 47p.

No.27 2020 Dr Gervais Rufyikiri, “Reshaping Approaches to Sustainable Peacebuilding and Development in Fragile States – Part II: Nexus between Unethical Leadership and State Fragility”, 44p.

No.28 2021 Dr Gervais Rufyikiri, “Resilience in Post-civil War, Authoritarian Burundi: What Has Worked and What Has Not?”, 47p.

No.29 2022 Kevin M. Esvelt, “Delay, Detect, Defend: Preparing for a Future in which Thousands Can Release New Pandemics”, 65p.

No.30 2023 Stuart Casey-Maslen, “International Counterterrorism Law: Key Definitions and Core Rules”, 40p.

No.31 2023 Anjali Gopal, William Bradshaw, Vaishnav Sunil and Kevin M. Esvelt, “Securing Civilisation Against Catastrophic Pandemics”, 50p.

No.32 2024 Kemal Mohamedou, “The Wagner Group, Russia's Foreign Policy and Sub-Saharan Africa”, 41p.

No.33 2024 Anila Jelesijević, “The Prospective of the Western Balkans to the EU membership: Challenges and Possible Ways Forward”, 40p.

No.34 2024 Jean-Marc Rickli and Federico Mantellassi, “The War in Ukraine: Reality Check for Emerging Technologies and the Future of Warfare”, 53p.

No.35 2024 Lassi Heininen, “Geopolitical Features, Common Interests and the Climate Crisis: The Case of the Arctic”, 38p.

No.36 2025 Arthur Lusenti, “The Indo-Pakistani Conflict in Light of the ‘Islamic Bomb’”, 50p.

Building Peace Together

Geneva Centre for Security Policy
Maison de la paix
Chemin Eugène-Rigot 2D
P.O. Box 1295
1211 Geneva 1
Switzerland
Tel: + 41 22 730 96 00
Contact: www.gcsp.ch/contact
www.gcsp.ch

ISBN: 978-2-88947-125-6

GCSP
Geneva Centre for Security Policy