Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

armasuisse

GCSP
Geneva Centre for Security Policy

What, Why and When? A Review of the Key Issues in the Development and Deployment of Military Human-Machine Teams

Jean-Marc Rickli, Federico Mantellassi, Quentin Ladetto

Tailored Study

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.

armasuisse Science and Technology

armasuisse Science and Technology is the technology competence center of the Federal Department of Defence, Civil Protection and Sport (DDPS). With its expertise in the areas of testing, research and innovation, it develops science-based foundations for its partners.

Known as deftech – defence future technologies - the technology foresight research program of armasuisse S+T inspires, informs and instructs the armed forces and its various stakeholders about the opportunities and threats posed by the use of technology.

Through its products and activities, it contributes to a collaborative and participative adventure that strengthens Switzerland's defence capability.

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
E-mail: info@gcsp.ch
www.gcsp.ch

ISBN: 978-2-88947-011-2

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 GCSP. He is also the co-chair of the Partnership for Peace Consortium (PfPC) Emerging Security Challenges Working Group and a senior advisor for the Artificial Intelligence Initiative at the Future Society. He is the co-curator of the International Security Map of the Strategic Intelligence Platform of the World Economic Forum. He is also a member of the Geneva University Committee for Ethical Research and of the advisory board of Tech4Trust, the first Swiss startup acceleration program in the field of digital trust and cybersecurity. Prior to these appointments, Dr Rickli was an assistant professor at the Department of Defence Studies of King's College London and at the Institute for International and Civil Security at Khalifa University in Abu Dhabi. In 2020, he was nominated as one of the 100 most influential French-speaking Swiss by the Swiss newspaper *Le Temps*. Dr Rickli received his PhD in International Relations from Oxford University. His latest book published by Georgetown University is entitled *Surrogate Warfare: The Transformation of War in the Twenty-first Century*.

Mr Federico Mantellassi is a Research and Project Officer at the Geneva Centre for Security Policy where he has worked since 2018. Federico's research and writing focuses on how emerging technologies impact international security and warfare, as well as on the societal implications of their development and use. Federico is also the project coordinator of the GCSP's Polymath Initiative; an effort to create a community of scientists able bridge the gap between the scientific and technological community and the world of policy making. Previously, he assisted in the organisation of executive education activities at the GCSP and was the project coordinator of the annual Geneva Cyber 9/12 Strategy Challenge. He holds a Master's Degree in Intelligence and International Security from King's College London, and a Bachelor's Degree in International Studies from the University of Leiden. Federico speaks English, French and Italian.

Dr Quentin Ladetto is the creator and director of the armasuisse Science and Technology foresight program - https://deftech.ch. Its aim is to identify disruptive technology use cases and trends and assess their implications in a defence context. He holds a PhD in geomatics from the Ecole Polytechnique Fédérale de Lausanne (EPFL), an Executive MBA in Finance from HEC Lausanne and a degree in Technology Management from the IMD. Quentin is the co-founder of atelierdesfuturs.org and a founding member of the association *Futurs*.

About this publication

Study conducted by the Geneva Centre for Security Policy for the technology foresight programme of armasuisse Science and Technology: Deftech.ch.

Study commissioned by: Dr Quentin Ladetto
Study conducted by: Dr Jean-Marc Rickli and Mr Federico Mantellassi

Preface

Dear reader,

You've probably already used terms whose meaning you – or anyone else, for that matter – don’t fully grasp, but which you nevertheless choose to use to keep up with the times. Among these terms, "artificial intelligence" and its abbreviation, "AI", will likely occupy first place. Whether describing a product, a service or a function, artificial intelligence/AI is ever more frequently used as the guarantor of what could be perceived as “super efficiency”. The absence of reference to it is perceived today as a sign of “inefficiency”. These terms almost always polarise and provoke debate, arousing both fears and expectations, mistrust and hope.

The subject of this paper is not AI as such, but "human-machine teaming", or HMT, which we consider as a possible milestone in the evolution of weaponry. In the course of the paper we aim to verify the meaning that is attributed to this increasingly fashionable term. It seems to me that this is a matter of urgent concern, because it is influencing a growing number of research programmes and R&D budgets.

While for AI the notion of “intelligence” is the subject of debate, for HMT it is the notion of “team” that is problematic. Usually, “team” refers to a group of *people* sharing the same goal. The spirit that drives each team member is one of solidarity with the other members and of surpassing oneself while working in the team. If one team member is of a quite different nature – in this case a machine – can we still speak of a "team" and, if so, can we expect the same values from it? In short, is it a question of abandoning this word in favour of another, such as “interaction” or “use”, with the effect of excluding the machine from the human dynamic, of which it would be no more than an instrument, however complex and capable it may be?

These questions might seem trivial if they did not give rise to a crucial reflection on the relationship between humans and machines, the possible expectations placed on these machines, their degree of “empathy”, their understanding of the environment in which they function, and even their possible ability to express feelings.

From technological issues we therefore find ourselves sliding into concerns that touch on philosophy and sociology; in turn, these concerns challenge the technologies that give rise to such enquiries.

By endorsing the words of Albert Camus, who said that "to name an object incorrectly is to add to the unhappiness of the world", we hope that this research project will help to clarify certain concepts and establish an interim assessment of the meaning of much-discussed notions.

We wish you an exciting journey.

Yours in Foresight,

Dr Quentin Ladetto
Head of Technology Foresight, armasuisse Science & Technology

Contents

Introduction

The modernisation of the military domain has resulted in sophisticated robotics and digital technological innovations increasingly making up an ever-growing aspect of modern military operations. Indeed, recent technological advances in robotics, artificial intelligence (AI), cognitive sciences and related fields have brought about both a *quantitative* increase in the use of robotics on the battlefield and a *qualitative* increase in their capabilities.[1] In this context, AI is seen as playing a key role in future battlefield success. Hence, armed forces globally are forging ahead with the integration of AI into the military domain, with the expectation that the successful operationalisation of ever-increasing AI-controlled partly or fully autonomous robotic systems on the battlefield will confer decisive advantages on those who adopt them.

Human-machine teaming (HMT) – the integration of machines and humans in *teams* – is seen as the optimal way of combining human soldiers with these increasingly numerous, autonomous and capable robotic systems. Indeed, largely due to novel capabilities and complex machine behaviours enabled by AI, machines could come to be considered as *teammates*. This would therefore represent an evolution from more traditional relationships between humans and machines, where the machine is a *tool* operated by humans. According to this view, *teaming* will enable armed forces to leverage the best aspects of humans and machines by successfully combining them in hybrid systems. For some, HMT represent the best way for humans to operate in tomorrow's faster, more data-heavy and more autonomous battlefield.[2]

However, several complicating factors persist. As a still-developing set of technologies and concepts, HMT is often misunderstood, resulting in important definitional challenges, risks, confusion, hype, and uncertainty over its current and future capabilities. While a great deal is known about how to successfully create teams of humans, the same cannot be said about machines – especially ones with high degrees of autonomy – partly because of the novelty of this concept. Issues around trust in automation, bidirectional communication, interfacing and many more areas remain unresolved. Additionally, because *teaming* is an inherently social activity, HMT is not only a technical endeavour, but a social one too. However, very few have engaged with this aspect of the HMT process.

Therefore, due to the speed with which autonomous systems are permeating the military domain, there is an urgent need for militaries to gain a comprehensive understanding of the issues affecting HMT. What does it mean to *team* with a machine? What complexities does the military domain introduce to this endeavour? Is *teaming* the correct framework to understand these new interactions between humans and intelligent machines? What are the implications that stem from this novel conceptualisation of human-machine interaction?

---

1 J.-M. Rickli and F. Mantellassi, "Human-Machine Teaming in Artificial Intelligence-Driven Air Power: Future Challenges and Opportunities for the Air Force", *Air Power Journal*, Fall 2022, pp.91-100, https://www.diacc.ae/resources/2022_Jean_Marc_Rickli_Federico_Mantellassi_Human-Machine_Teaming_Air_Power.pdf.

2 T. Nurkin and J. Siegel, "Battlefield Applications for Human-Machine Teaming: Demonstrating Value, Experimenting with New Capabilities, and Accelerating Adoption", Atlantic Council, August 2023, https://www.atlanticcouncil.org/wp-content/uploads/2023/08/Battlefield-Applications-for-HMT.pdf.

This paper will serve as a springboard for understanding to what extent HMT is a realistic, useful and operationalisable framework to conceptualise future human-machine interactions in the military domain. It is structured in three parts. Part 1 will provide an overview of the literature and discussions surrounding the issue, both in academic and defence settings. Part 2 will analyse how the specificities of the military domain complicate, alter and possibly aggravate many of the already existing unresolved issues surrounding the desire to *team* humans with machines. Lastly, Part 3 will delve into some recommendations.

Part 1: What and why?

1.1 The logic behind human-machine teaming: why are militaries thinking about teaming humans with machines?

It is first worth understanding *why* armed forces have sought to understand and develop human-machine relationships through this paradigm. Why has our conceptualisation of human-machine interaction shifted from one of using a tool, to *teaming* with it? The increased attention given to HMT in the military domain can be attributed to three factors, which are discussed below.

1.1.1 The increased presence of robotics on battlefields

Recent conflicts have shown the increased prevalence of robotics on battlefields in the form of drones and a growing assortment of unmanned systems.[3] They are also further characterised by incredibly high rates of attrition for equipment, e.g. Ukraine is reportedly losing over 10,000 drones a month, while Russia has lost over 2,000 tanks since the beginning of the conflict between these countries.[4] In light of this, armed forces are under pressure to “bring mass to the battlefield”. However, they will find it difficult to dramatically increase the number of expensive traditional platforms like tanks, fighter jets or advanced artillery systems. Thus, they are under pressure to develop and field large quantities of cheap, attritable systems to complement traditional platforms. This will likely come in the form of unmanned and highly autonomous platforms of various kinds. For example, the United States is planning to field thousands of “‘small, smart, cheap, and many’” autonomous weapons to offset China’s rapid military development.[5] The high rate of attrition is not only true for systems, but is also a reality for troops, because today’s battlefields have become incredibly lethal, especially in near-peer and peer conflicts. Armed forces are therefore increasingly looking for technological surrogates, and have adopted the rationale of increasing the use of robotics in highly contested spaces to put troops as much out of harm’s way as possible.[6] The idea that robotics will take care of the “dull, dirty, and dangerous” aspects of war has been a key motivator of this trend.[7]

1.1.2 The perceived advantages of integrating increasing levels of autonomy into military systems

Aside from increasing the number of robotic systems on battlefields, armed forces are seeking to increase the autonomy of these platforms. The rationale for this is fourfold. Firstly, the increased autonomy of weapons systems entails the

3 F. Borsari and G.B. Davis, *An Urgent Matter of Drones*, Center for European Policy Analysis, 27 September 2023, https://cepa.org/comprehensive-reports/an-urgent-matter-of-drones/.

4 J. Watling and N. Reynolds, “Meatgrinder: Russian Tactics in the Second Year of Its Invasion of Ukraine”, Royal United Services Institute for Defence and Security Studies, 19 May 2023, https://static.rusi.org/403-SR-Russian-Tactics-web-final.pdf; J. Janovsky et al., “Attack on Europe: Documenting Russian Equipment Losses during the Russian Invasion of Ukraine”, Oryx, accessed 20 November 2023, https://www.oryxspioenkop.com/2022/02/attack-on-europe-documenting-equipment.html.

5 T. Newdick and T. Rogoway, “Replicator Is DoD’s Big Play to Build Thousands of Autonomous Weapons in Just Two Years”, The War Zone, 28 August 2023, https://www.thedrive.com/the-war-zone/replicator-is-dods-big-play-to-build-thousands-of-autonomous-weapons-in-just-two-years.

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

7 D. Kunertova, “From Robots to Warbots: Reality Meets Science Fiction”, CSS Analyses in Security Policy, 2021, https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse292-EN.pdf.

increased survivability of platforms in highly contested, communication-deprived environments.[8] As such, technological surrogates, like drone swarms, will act as force multipliers, particularly as their autonomy increases. Secondly, AI-enabled autonomy in weapons systems is seen as a key way to leverage current and future data-rich battlefields and accelerate the tempo and accuracy of operations. Superior analytical speed and the networking of autonomous systems are often cited as possible benefits of increased autonomy.[9] Thirdly, autonomy is seen as a key way to alleviate the potential cognitive overload that humans may experience in operating an increasing number of military robotic systems, which can sometimes make teams less effective and degrade combat effectiveness.[10] Lastly, armed forces are seeking to make robotic systems undertake an increasing variety of tasks, not least those that could spare the lives of their own personnel.[11] To this end, AI-driven autonomy in particular is seen as a key enabler. As the level of autonomy grows and AI becomes increasingly advanced, armed forces expect that robotic systems will be developed with capabilities that put them increasingly functionally on par with humans, thus outgrowing the traditional role of tools used by humans.[12] As machine behaviours become increasingly complex and autonomous, the ensuing relationship with their human counterparts becomes increasingly complex. Some consider that this necessitates a shift to considering such advanced machines as teammates rather than mere tools.

1.1.3 The still relative brittleness of AI-enabled autonomous systems

For now, however, both due to normative and practical reasons, armed forces envision these systems as coexisting with humans.[13] From a normative perspective, ethical pressures in terms of which humans still need to be involved in the use of deadly force remain. Humankind has not yet fully come to terms with the idea of fully autonomous systems, and significant international debates and pressure persist over this idea. In fact, “meaningful human control” remains an ever-present feature of the discussion surrounding the governance of military AI. From a practical perspective, while showing great promise, AI-enabled systems currently remain relatively brittle (subject to surprising failures, spoofing or other adversarial attacks) and still limited in what they can do.[14] This is especially true in the highly contested and unpredictable domain of war. Armed forces have little incentive to field weapons systems they do not understand or fully control, or whose behaviour they cannot fully predict.[15] At the current level of AI capability, this entails that humans are still the primary operators of these systems, remaining either on or in the loop. Furthermore, current AI-enabled systems are designed to play support roles and to function as analytical enablers and force multipliers

8 P. Scharre and M.C. Horowitz, “An Introduction to Autonomy in Weapon Systems”, Center for a New American Security Working Paper, 2015, https://www.files.ethz.ch/isn/188865/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf.

9 T.X. Hammes, “Autonomous Weapons Are the Moral Choice”, Atlantic Council, 2 November 2023, https://www.atlanticcouncil.org/blogs/new-atlanticist/autonomous-weapons-are-the-moral-choice/.

10 P. Hinton, “Uncrewed Ground Systems: Organisational and Tactical Realities for Integration”, Royal United Services Institute for Defence and Security Studies Occasional Paper, October 2023, https://static.rusi.org/op-uncrewed-ground-systems_0.pdf.

11 T. Vestner and A. Lusenti, “Great Powers’ Military Robotics”, *Stratos*, Vol.1(23), 2023, https://dam.gcsp.ch/files/doc/great-powers-military-robotics.

12 Rickli and Mantellassi, 2022.

13 A. Neads et al., “From Tools to Teammates: HMT and the Future of Command and Control in the Australian Army”, Australian Army Research Centre, 2021, https://researchcentre.army.gov.au/sites/default/files/AARC%20Occasional%20Paper%20No.7%20-%20From%20Tools%20to%20Teammates.pdf.

14 S. Scott-Hayward, “Securing AI-based Security Systems”, Geneva Centre for Security Policy, Strategic Security Analysis Issue 25, June 2022, https://dam.gcsp.ch/files/doc/ssa-2022-issue25.

15 A. Holland Michel, “The Black Box, Unlocked: Predictability and Understandability in Military AI”, United Nations Institute for Disarmament Research (UNIDIR), 2020, https://unidir.org/files/2020-09/BlackBoxUnlocked.pdf.

whose role it is to assist and enhance human activities and not replace humans. Because warfare is an inherently human affair, where humans inflict violence on other humans to compel them to change their behaviour, humans will remain the central actors of warfare for some time to come.

Taken together, these three factors create an environment made up of increasingly numerous and capable machines that will coexist with humans with some degree of oversight responsibility over their actions.[16] In an effort to leverage the best out of both humans and machines, as well as the *relationship itself*, armed forces are seeking to operationalise human-machine teams.

1.2 What is a human-machine team?

The field of HMT suffers from a deep definitional crisis.[17] Substantive differences persist across various sectors (e.g. academia and industry) in defining precisely what a human-machine team *is*.[18] This leads the term to be used very inconsistently to describe a host of different human-machine interactions. The field's definitional problem stems principally from the fact that "across the literature, there isn't a cohesive, and coherent, understanding of the differences between different types of relationships between humans and technology, from cooperation to teaming".[19] As Berretta et al. explain, "In most of the literature, terms underlining the collaborative element such as *partner*, *symbiosis* or *teammate* are used as buzzwords without further explanation without really understanding humans and AI as a sociotechnical system acting as a team" (original emphasis).[20] The result is a breadth of literature focused on studying relationships between humans and technological artefacts that are *not* teams. In essence, therefore, there is a gap between the meaning of the term "human-machine *team(ing)*" and what it is often used to describe.

This is principally because *"teaming"* is a term borrowed from sociology, but regularly used in the engineering, robotics and defence literature in ways that are inconsistent with its sociological meaning.[21] Indeed, few engage with the implications of the use of the term and what it entails, focusing mostly on practical and technical aspects of machine behaviour. Discussions with an operational human factors expert revealed a scepticism over the need to define teams in line with sociological meanings. Because machines are inherently different from humans, focusing on the requirements of human *teaming* as the standard for them is not a fruitful endeavour.[22] In the authors' view, however, building a *team* of humans and machines also requires engaging with the philosophical and sociological dimensions of having machines as teammates.[23] A team is a "set

16 Rickli and Mantellassi, 2022.

17 S. Berretta et al., "Defining Human-AI Teaming the Human-Centered Way: A Scoping Review and Network Analysis", *Frontiers in Artificial Intelligence*, 2023, https://doi.org/10.3389/frai.2023.1250725.

18 B. Walker-Munro and Z. Assaad, "The Guilty (Silicon) Mind: Blameworthiness and Liability in Human-Machine Teaming", *Cambridge Law Review*, Vol.8(1), 2023, pp.1-24, https://heinonline.org/HOL/LandingPage?handle=hein.journals/cambrilv8&div=6&id=&page=.

19 Ibid.

20 Berretta et al., 2023.

21 A.M. Greenberg and J.L. Marble, "Foundational Concepts in Person-Machine Teaming", *Frontiers in Physics*, 2023, https://doi.org/10.3389/fphy.2022.1080132.

22 Interview with Dr S. Huber, 24 November 2023.

23 M. Black et al., "Supporting the Royal Australian Navy's Campaign Plan for Robotics and Autonomous Systems", RAND Australia, 2022, https://www.rand.org/content/dam/rand/pubs/research_reports/RRA1300/RRA1377-2/RAND_RRA1377-2.pdf.

of two or more people who interact dynamically, interdependently, and adaptively, toward a common and valued goal, each member having specific roles or functions to perform, and a limited life-span of membership".[24] As seen by the language of the definition, our understanding of teaming is that of an inherently human activity, a relationship between or among humans.[25] Therefore, true human-machine *teaming* should require a machine partner capable of exhibiting and understanding behaviours that have until now been exclusively human. In fact, a Defense Advanced Research Projects Agency programme focused on HMT seeks to "demonstrate the basic machine *social* skills needed to generate effective human-machine collaboration" (emphasis added).[26] It notes that "The inability of artificial intelligence ... to represent and model human partners is the single biggest challenge preventing effective human-machine teaming today".[27] Some experts posit that due to this, teaming is the prerogative of humans, and advocate a shift away from using anthropomorphic language to describe human-machine relations.[28]

Irrespective of the lack of common definition, there are some commonly agreed variables, dynamics and elements that would characterise human-machine teams. Berretta et al. provide a comprehensive definition of a human-AI team, which, because AI is a foundational requirement for teaming humans with with machines, the current authors believe can serve as a definition for HMT:

> Human-AI teaming is a process between one or more human(s) and one or more (partially) autonomous AI system(s) acting as team members with unique and complementary capabilities, who work interdependently toward a common goal. The team members' roles are dynamically adapting throughout the collaboration, requiring coordination and mutual communication to meet each other's and the task's requirements. For this, a mutual sharing of intents, shared situational awareness and developing shared mental models are necessary, as well as trust within the team.[29]

This definition captures key characteristics of HMT and imposes on it a significant relational complexity, involving bidirectional communication, adaptiveness, fluid tasking, shared mental models and situational awareness, interdependency, coordination, trust, and advanced AI autonomy. However, some experts note that perhaps *distributed* rather than *shared* situational awareness is a more useful characterisation of the division of knowledge in HMT.[30] In any case, HMT is therefore an interaction between humans and technology that fulfils both the technical and social and relational requirements of teaming.[31] It is worth noting that while the word "machine" in HMT seems to imply that the non-human

24 E. Salas et al., "Toward an Understanding of Team Performance and Training", in R.W. Swezey and E. Salas (eds), *Teams: Their Training and Performance*, Ablex Publishing, 1992, pp.3-29.

25 Greenberg and Marble, 2023.

26 DARPA (Defense Advanced Research Projects Agency), "Using AI to Build Better Human-Machine Teams", 21 March 2019, https://www.darpa.mil/news-events/2019-03-21b.

27 Ibid.

28 A. Sapadaro. "A Weapon Is No Subordinate: Autonomous Weapon Systems and the Scope of Superior Responsibility", *Journal of International Criminal Justice*, Vol.1(18), 2023, https://doi.org/10.1093/jicj/mqad025; K.D. Evans et al., "Do We Collaborate with What We Design?", *Topics in Cognitive Science*, Vol.1(20), 2023, https://doi.org/10.1111/tops.12682.

29 Berretta et al., 2023.

30 Interview with Dr S. Huber, 23 November 2023.

31 Greenberg and Marble, 2023.

element of the team is a physical object such as a robotic system, this is not necessarily the case. In HMT the non-human counterpart could just as well be non-physical or software based, such as an algorithm.

Teams are social artefacts, therefore real human-machine teams are ones that reflect and integrate the social nature of teaming and the relational complexity this implies. Anything short of this is not a true *team*, but an instance of human-machine interaction in which the machine is a (sometimes very advanced) tool, but not a teammate. For it to be a true teammate, advanced AI is foundational. By advanced AI we refer to what is sometimes called Frontier AI, i.e. "Highly capable general purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models".[32] Indeed, O'Neill et al. note that "significant advances in AI, machine learning and cognitive modeling ... ha[ve] resulted in the shifting of human–autonomy collaboration from purely conceptual to increasingly practical and applicable".[33] Advanced AI is a key technology that will allow the levels of autonomy necessary for true HMT and the ability to communicate, understand contexts and situations, share mental models with humans, and other such complex social and relational behaviours worthy of a teammate.

Central to human-machine teams is the role of trust, and the complex act of building and calibrating it.[34] Trust can be understood as "*the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party*" (original emphasis).[35] In relation to machines, trust can be thought of as "the human's confidence in the reliability of the system's conclusions, and its ability to perform specified tasks and accomplish defined goals".[36] System transparency, explainability, experience, training, mechanical understanding and efficient interfaces have been shown to contribute to appropriate trust calibration.[37] However, trust is a careful balancing act, and humans are susceptible to both overtrusting and undertrusting a system. They have been shown to sometimes give undue trust to autonomous machines, deferring to their judgement over their own (known as automation bias).[38] Inversely, increasingly performant and complex AI-enabled autonomy has also been shown to potentially adversely affect trust, due to AI's "black box" problem.[39] Trust is also highly contextual, because a human operator might calibrate their trust in a system according to external factors. It is also highly interrelated with other elements of HMT. For example, trust both enables

32 United Kingdom Government, "Safety and Security Risks of Generative Artificial Intelligence to 2025", n.d., https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf.

33 T. O'Neill et al., "Human-Autonomy Teaming: A Review and Analysis of the Empirical Literature", *Human Factors*, Vol.67(5), 2022, https://doi.org/10.1177/0018720820960865.

34 M.A. Ibrahim et al., "Trust and Communication in Human-Machine Teaming", *Frontiers in Physics*, 2022, https://doi.org/10.3389/fphy.2022.942896; M. Konaev et al., "Trusted Partners: Human-Machine Teaming and the Future of Military AI", Center for Security and Emerging Technology Issue Brief, 2021. https://cset.georgetown.edu/publication/trusted-partners/.

35 R.C. Mayer et al., "An Integrative Model of Organizational Trust", *Academy of Management Review*, Vol.20(3), 1995, pp.709-734, https://doi.org/10.2307/258792.

36 Konaev et al., 2021.

37 Rickli and Mantellassi, 2022.

38 For a more detailed overview of the relationship between trust and autonomous systems, see I. Puscas, "Human-Machine Interfaces in Autonomous Weapon Systems: Considerations for Human Control", UNIDIR, 2022, pp.9-10; https://unidir.org/wp-content/uploads/2023/05/UNIDIR_Human-Machine-Interfaces.pdf; I. Puscas, "AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures", UNIDIR, 2023, pp.32-33, https://unidir.org/wp-content/uploads/2023/10/UNIDIR_AI-international-security_understanding_risks_paving_the_path_for_confidence_building_measures.pdf; Rickli and Mantellassi, 2022.

39 Puscas, 2022.

effective communication, shared mental models and successful coordination, while also being affected by these same elements.[40] Without proper trust calibration, human operators will either (1) underuse machine counterparts, thereby negating the benefits of HMT, or (2) overuse their machine counterparts in inappropriate situations and for inappropriate purposes, thereby dangerously reducing the agency and control of the human input. Hence, the appropriate calibration of trust in machine teammates will be a key determinant of success in human-machine teams.

1.2.1 Misuse of the term HMT

As mentioned, the military community often uses the term HMT to describe any relationship between a machine using some unspecified degree of AI and autonomy, or advanced robotics and humans, without considering the above-mentioned elements. In these cases, the term HMT often denotes a relationship that exhibits unspecified characteristics that would make it more complex than a simple interaction and therefore worthy of a different terminology – here "team". However, it is unclear why these human-machine interactions warrant the use of a different terminology.

Taking as an example the characterisation of the United Kingdom's (UK) Ministry or Defence (MoD) is a good way to illustrate this point. In its Joint Concept Note 1/18 on HMT, the MoD avoids a direct definition of HMT, but describes it as "the effective integration of humans, artificial intelligence ... and robotics into warfighting systems".[41] The issue is that this way of defining HMT could also be used to define many types of human-machine interactions that do not constitute teams. For illustration purposes, one can use the example of a now-familiar scene in modern conflicts. A drone operator is flying a drone that uses image recognition software to identify targets. This information is relayed to artillery units or directly used by the operator to deliver a payload attached to the drone, and the target is destroyed.[42] This scenario fits the UK MoD's definition of HMT, because there is a human, robotics (the drone), AI (image recognition software) and efficient integration of the system's parts (i.e. the different parts worked effectively to achieve the desired result).

However, this instance of human-machine interaction is a far cry from a conceptualisation of a human-machine *team* as described above that includes elements such as:

- shared mental models between teammates (human and non-human);
- real autonomy (understood as some level of machine self-determination, self-government and task attribution);
- bidirectional coordination and communication;
- fluid task attribution; and
- shared situational awareness.

40 O'Neill et al., 2022.

41 United Kingdom Ministry of Defence, "Joint Concept Note 1/18: Human-Machine Teaming", 2018, https://assets.publishing.service.gov.uk/media/5b02f398e5274a0d7fa9a7c0/20180517-concepts_uk_human_machine_teaming_jcn_1_18.pdf.

42 A.E. Kramer and L. Addario, "Budget Drones Prove Their Value in Billion Dollar War", *New York Times*, 23 September 2023, https://www.nytimes.com/2023/09/22/world/europe/ukraine-budget-drones-russia.html.

This is in line with the way in which the term is used at large by think tank, military, and even academic communities in ways that fall short of what is necessary to define a relationship as that of a team.[43] However, if any human interactions with a machine are depicted as constituting a team, then the term HMT is void of meaning and therefore not analytically useful.

1.3 Human-machine teaming: a continuum

To offer some clarity, conceptualisations of HMT can therefore be placed on a continuum (see Figure 1). This continuum ranges from notions of simple human-tool interactions through more complex collaborations to real teams. Movement to the right along the continuum implies two things. Firstly, there is a complexification of the interaction and the relational and social elements of such interaction, from one-way interactions and binary processes to complex, interdependent, and fluid behaviours requiring shared mental models and efficient bidirectional communication. Secondly, it also implies increasing levels of technological maturity to achieve such relational complexity. This, in turn, closely mirrors the discussion surrounding autonomy, which is also often represented on a continuum from automated to autonomous. Movement towards a *team* on the HMT continuum most certainly requires movement towards autonomous on the autonomy continuum. Where one places oneself on this continuum therefore determines both the level of complexity of the relationship that is described and the level of technological maturity necessary for its realisation.

For example, the capabilities enabled by Project Maven, an often-cited example of the efforts of the US Department of Defense (DoD) to integrate AI into its operations, fall towards the left-hand side of the continuum, somewhere between a "tool" and "collaboration" type of relationship. As an algorithmic suite intended to autonomously detect, categorise, and flag objects of interest from surveillance footage, Project Maven seeks to get people and computers to "work symbiotically to increase the ability of weapon systems to detect objects".[44] However, such capabilities, and the interactions with humans that ensue, remain relatively rudimentary. The AI does not possess a level of self-determination beyond the confines of its tasks, and does not participate in the dynamic attribution of and shift tasks with its human overseers. It is, for example, not capable of "supporting, taking over, cooperating, or setting borders for the human as needed in specific situation [sic]".[45] In this case, therefore, one cannot say that the machine is capable of *shared* situational awareness and mental modelling.

Further along the scale towards (but not at) *teaming*, one can find the "loyal wingman" concept, understood as the pairing of highly autonomous, unmanned aircraft working alongside and collaboratively with manned aircraft.[46] With a first flight alongside manned aircraft in October 2023, the Kratos XQ-58A Valkyrie, which was flown entirely by an onboard AI system, showed remarkable capabilities and some degree of self-government. At least conceptually, at the current level of sophistication of the most capable sensors, computers, and AI,

43 Berretta et al., 2023.

44 C. Pellerin, "Project Maven to Deploy Computer Algorithms to War Zone by Year's End", US DoD, 21 July 2017, https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

45 Berretta et al., 2023.

46 Rickli and Mantellassi, 2022.

the loyal wingman concept will entail a complex relationship between pilot and system, characterised by complex trust dynamics, bidirectional communication, interdependence, and advanced autonomy. However, current technological capabilities do not yet enable all the social and relational requirements of a true *teammate*.[47] As Seeber et al. note (original emphasis),

> for machines to be effective teammates, they will need to be more capable than today's chatbots, social robots, or digital assistants that *support* team collaboration. They will need to *engage* in at least some of the steps in a complex problem-solving process, i.e., defining a problem, identifying root causes, proposing and evaluating solutions, choosing among options, making plans, taking actions, learning from past interactions, and participating in after-action reviews. Such machine partners would have the potential to considerably enhance team collaboration.[48]

O'Neill et al. further state that in the future, "technological advances (e.g., AI) lead to the development [of] autonomous agents with increasing levels of capabilities to interact with humans dynamically, greater self-learning capabilities, and *affective, behavioral, and cognitive capacities consistent with what humans expect from a genuine team member*" (emphasis added).[49] For Brill et al., the following conditions must be met for a true HMT:

> Teammates typically perform complementary, largely non-redundant functions. Teammates have bounded autonomy, freedom to act according to one's judgment, but within the limits of social contracts, organizational structure, or situational constraints. Autonomy mandates nondeterministic behavior; however, the aforementioned constraints should limit unpredictability of actions and outcomes. Teammates have shared knowledge and shared awareness ... and as such, they learn and change. With sufficient experience and a positive history, teammates will learn to trust one another, and their collective experiences will lead to enhanced transparency. Lastly, teammates act with benevolence to help one another and to further the team's goals.[50]

Hence, even the loyal wingman concept falls short of reaching the furthest point on the HMT continuum. For that, the development of social intelligence in machines will be necessary, enabling them to develop a "theory of mind"[51] of humans, and in turn allowing them to share mental models.[52] For now, however, "AI research has not yet produced technology capable of critical thinking and problem solving on par with human abilities".[53] This, we concur, will require advances in fundamental research in the fields of AI (frontier AI; artificial general

47 Konaev et al., 2021.

48 I. Seeber et al., "Machines as Teammates : A Research Agenda on AI in Team Collaboration", *Information and Management*, Vol.57(2), 2020, https://doi.org/10.1016/j.im.2019.103174.

49 O'Neill et al., 2022.

50 J.C. Brill et al., "Navigating the Advent of Human-Machine Teaming", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol.62(1), 2018, https://doi.org/10.1177/1541931218621104.

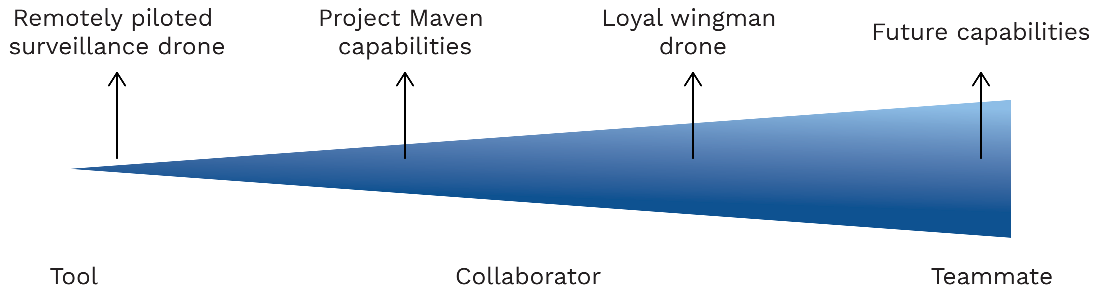
51 Theory of mind is the ability to understand and take into account another individual's mental states. For more information on this issue, see https://www.sciencedirect.com/topics/neuroscience/theory-of-mind.

52 DARPA, 2019.

53 Seeber et al., 2020.

intelligence, or AGI), cognitive sciences and neurotechnology, as well as potentially synthetic biology.

Figure 1: The HTM continuum



All in all, a review of the field shows that real, operational human-machine *teams* are not yet a reality. This necessitates looking beyond the use of the term by the military and think thanks that portray HMT as an accomplished reality. As noted, often when the term is used it is not actually describing HMT in a way that is consistent with what proper *teaming* would entail. What exists for now, and what is often described using the term HMT, are collaborations between humans and machines with high levels of autonomy, but not working in *teams*. This requires technological capabilities that do not yet exist. Understanding that discussions on HMT exist on a continuum helps provide some clarity to this debate. Placing examples of human-machine interactions on this continuum guided both by the relational complexity of the interaction between human and agent and the level of technological sophistication at hand can help us observe where any particular instance or capability is in terms of *teaming*.

Part 2: Key challenges for human-machine teaming in the military domain

HMT will not solely be confined to the military domain and will have applications across various other sectors and industries, from the medical sector to industrial assembly lines.[54] However, the military domain differs significantly from other domains. It is adversarial, dirty, complex, lethal, controversial, characterised by uncertainty (fog) and emotions, and is ethically loaded. Characteristics such as these are not trivial and are bound to impact HMT in the military domain, including its feasibility and timeline to development, and the shape that it will take.

HMT in the military domain therefore needs to be considered as distinctively different from HMT in other areas. We will do this by highlighting some key issues integral to the military domain that are mostly absent from the existing literature on military HMT, but which are almost certain to have an impact on the realisation of HMT in military settings.

Stress, time constraints and heightened emotional states have been shown to significantly affect how people engage with technology, especially autonomous technology.[55] Because teaming with machines is a social – and emotional – behaviour, these pressures are bound to affect these relationships in times of war. As Ibrahim et al. note,

> for HMTs [human-machine teams] operating in such environments, there may be unpredictable aspects of teammate interactions that emerge as a function of the HAS [highly automated system] capabilities, the human teammate, the team dynamics, and the complexity and unpredictability of the contexts they operate in. This makes the HMTs adoption in dynamic contexts risky and potentially costly for humans involved – both within the HMT, and in their environments.[56]

Communications, the availability of time, trust, interdependence and shared situational awareness (among others) are all likely to be highly affected by the realities of war, thereby degrading the combat effectiveness of human-machine teams. For example, effective bidirectional communication between human and machine is seen as a cornerstone of HMT, and one of the key tenants of *teaming* in general.[57] Communication is central to creating trust, managing coordination and situational awareness, and building shared mental models.[58] In today's highly contested battlefields, maintaining the edge in the electromagnetic battlefield is both a challenge and a key requirement.[59] It is safe to assume that electronic warfare measures will impede communications between humans and their

54 K.H. Henry et al., "Human-Machine Teaming Is Key to AI Adoption: Clinician's Experiences with a Deployed Machine Learning System", *Npj Digital Medicine*, Vol.5(97), 2022, https://www.nature.com/articles/s41746-022-00597-7.

55 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(1), 2023, pp.43-67, https://doi.org/10.1080/14702436.2022.2102486.

56 Ibrahim et al., 2022.

57 National Academies of Sciences, Engineering, and Medicine, *Human-AI Teaming: State of the Art and Research Needs*, Washington DC, National Academies Press, 2022, https://doi.org/10.17226/26355.

58 O'Neill et al., 2022.

59 V. Zaluzhny, "The Commander-in-Chief of Ukraine's Armed Forces on How to Win the War", *The Economist*, 1 November 2023, https://www.economist.com/by-invitation/2023/11/01/the-commander-in-chief-of-ukraines-armed-forces-on-how-to-win-the-war.

machine counterparts, affecting the ability to pull and push information between teammates, attribute and change tasks, and – ultimately – operate effectively. It is also worth understanding how battlefield pressures affect the communication styles of humans and the resultant need to ensure that machine counterparts are designed to comprehend emotionally loaded, stress-induced and potentially unclear communications. Cognitive overload will surely also affect the ability of human counterparts to receive and digest information.

Humans should retain control of the relationship with their machine counterparts. In high-intensity combat situations, where time pressures are condensed and the stakes are at their highest, human operators are unlikely to have the time to meaningfully engage with their machines. However, these will be the critical times when such engagement will be the most necessary. Research demonstrates that "the more cognitively demanding, time-pressured, and stressful a situation is, the more likely humans are to defer to machine judgments".[60] Under pressure to accelerate their decision-making, soldiers might increasingly defer to their machine counterparts.[61] Additionally, in theory a successful HMT should allow for "graceful handoff", i.e. the handing over of tasks to humans when machine limits are encountered.[62] This requires both highly advanced AI and the human team member(s) to be cognitively engaged enough to appreciate that the machine's limits have been reached and successfully safely take over from the machine. In the situation described above, however, the human operator is unlikely to have either the time or the cognitive and emotional capacity to do so. Understanding how trust calibration – and ensuing changes in human behaviours – is affected by fear, lack of time and cognitive overload will therefore be of the utmost importance to effective HMT.

The warfighting domain will also exert pressures on AI-enabled autonomy itself. As a key enabler of greater machine behaviour increasingly approaching *teammate* potential, the conversation around military HMT is inextricably linked to that of the role of AI in the military domain. It is especially linked to the question of whether or not AI is well suited to the realities of warfare – which remains an open question. Some posit that AI is inherently unable to deal with the complexities of warfare, and is for now mostly suited to data-rich and linear environments.[63] For some, AI "cannot make appropriate command decision [sic] because it is engaged in a different kind of logic that can only make sense within a kitsch, gamified vision of war".[64] While AI adoption in the military domain is racing ahead, the jury is still out on whether or not it will be able to cope with war's "non-linear, chaotic, and analytically unpredictable nature".[65] As the key technology endowing autonomous agents in future human-machine teams, it is worth making serious efforts to understand if – and how – AI-enabled systems might fail in wartime. However, the current relative brittleness of AI is not necessarily an indication of its future capabilities. While AI could advance

60 Johnson, 2023.

61 Ibid.

62 Greenberg and Marble, 2023.

63 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*, London, Routledge, 2023, https://www.taylorfrancis.com/chapters/edit/10.4324/9781003218326-2/artificial-intelligence-warfare-jean-marc-rickli-federico-mantellassi; Johnson, 2023.

64 C. Hunter and B.E. Bowen, "We'll Never Have a Model for an AI Major General: AI, Command Decisions and Kitsch Visions of War", *Journal of Strategic Studies*, 2023, https://doi.org/10.1080/01402390.2023.2241648.

65 Johnson, 2023.

to the stage where it can enable human-level teammate capabilities, it is also worth considering the downside of such a scenario, i.e. the implications of the realisation of these hyper-capable AI systems (sometimes called Frontier AI) and what they entail for HMT in military contexts. This requires the ability to speculate, envisage possible scenarios and ask questions rather than provide concrete answers. For example, what happens – if and when – these systems accelerate the tempo of war and increase its complexity beyond any possible meaningful human engagement and understanding of situations and context? At this point, will the human, stripped of any possible meaningful input into machine actions and decisions, be subservient to the machine's recommendations? What does this imply for HMT, i.e. will we still be able to speak of a *team*? Will human enhancement be necessary to cope with the abilities of the machine? As we have seen with automation bias, AI does not even need to reach this stage for humans to lose critical engagement with machines.[66]

All in all, the variables of successful and safe HMT (e.g. effective communications, well-calibrated trust, coordination, shared mental models and situational awareness) presuppose an environment in which war fighters are unlikely to find themselves in wartime, precisely when HMT would be particularly useful. Unfortunately, discussions around military robotics, especially as they relate to AI, are full of rosy depictions of war that are remote from the realities of actual warfare.[67] For example, AI company Palantir's recent demonstration of its Artificial Intelligence Platform, a large-language-model-based battle management platform, has been cited as an example of this trend.[68] The demo depicts a type of frictionless warfare, where hyper performant AI systems are able to flourish unimpeded.[69]

The realisation of HMT cannot be considered independently from the brutal – and deeply human – realities of warfare. The development of human-machine teams cannot presuppose a permissive environment, data superiority, communication security and passive adversaries. Understanding HMT – and its successful operationalisation – in terms of such an idealised context will mean it will be unprepared for real warfare. Ensuring that human-machine teams are designed for the realities of war, not peace, will be foundational to their success.

2.1 Domain sensitivity

The adoption of unmanned systems is highly domain sensitive. The land, air, and sea domains[70] each has distinct characteristics that both significantly affect the ease of adoption of unmanned systems and autonomy and dictate the

66 Puscas, 2022.

67 I. Reynolds and O. Ahmet Cetin, "War Is Messy. AI Can't Handle It", *Bulletin of the Atomic Scientists*, 14 August 2023, https://thebulletin.org/2023/08/war-is-messy-ai-cant-handle-it/; Hunter and Bowen, 2023; F. Mantellassi, "Charting a New Course for Military AI: Going Beyond Militaristic Narratives and Rosy Depictions of Warfare", Stop Killer Robots, 24 October 2023, https://stopkillerrobots.medium.com/charting-a-new-course-for-military-ai-going-beyond-militaristic-narratives-and-rosy-depictions-of-8295a9075140.

68 Reynolds and Ahmet Cetin, 2023.

69 Palantir, "Palantir AIP: Defence and Military", YouTube, 25 April 2023, https://www.youtube.com/watch?v=XEM5qz__HOU&ab_channel=Palantir.

70 The authors recognise the existence of further domains, such as the cyber or cognitive domains of war. Because the argument focuses on the geospatial characteristics of the domains, the authors choose to focus on these traditional – and "physical" – domains of war. For an exploration of the cognitive domain of warfare, see J.M. Rickli et al., "Peace of Mind: Cognitive Warfare and the Governance of Subversion in the 21st Century", Geneva Centre for Security Policy, Policy Brief No. 9, August 2023, https://dam.gcsp.ch/files/misc/pb-9-rickli-mantellassi.

modalities of such adoption. For example, the land domain presents significantly more challenges than the air domain because the complexity of the terrain, the number of plausible situations and the increased ways in which humans can fool AI systems multiply the ways in which these systems can fail.[71] While advances are beginning to be made, the disproportionate majority of unmanned systems operate in the air domain, with a minority in the sea and land domains.[72] This is likely to have an impact on HMT, which will not spread consistently in the military as a whole, but at different speeds in different domains. This is likely to mirror the dynamics of unmanned systems and autonomy adoption. It is therefore not a coincidence that the bulk of HMT efforts globally are focused on air forces.[73] The US Air Force's collaborative aircraft – or loyal wingman drone – as part of its Next Generation Air Dominance programme, and manifested in its Kratos Valkyrie Drone, is one such example. Other air-force-focused efforts also exist, such as the Australian Boeing MQ-28 Ghost Bat drone, a similar collaborative loyal wingman drone.[74] This is due to both the clearer ability to model the characteristics of the air domain and the fact that complex interactions between humans and extremely demanding and sophisticated machines lie at the heart of air power.[75]

The land domain is likely to present more challenges to the possibility of implementing HMT, mostly due to the comparatively greater number of possible obstacles that machines will face than in the air domain. On land, AI applications are much more dependent on their environment and therefore initially much more likely to fail. As seen in Ukraine, the high attrition rates of mechanised brigades and infantry units in large-scale conflicts are also likely to put a strain on the effective fielding of human-machine teams in times of war. In the sea domain, the physics of the domain, resulting in barriers to communications or limited data and power storage possibilities on ships, pose restrictions on the fielding of AI systems.[76] Nonetheless, research into and tests of especially underwater unmanned vehicles are being conducted in order to improve intelligence, command-and-control abilities, and combat operations.[77]

2.2 Challenges to interoperability

Twenty-first-century warfare has come to be characterised by interoperability, multidomain operations and combined arms manoeuvres. Furthermore, modern military operations have routinely been carried out by alliances or multilateral

71 W.A. Sanchez, "Unmanned Ground Vehicles Face a Rare Market Challenge in South America: Horses", Breaking Defense, 23 August 2023, https://breakingdefense.com/2023/08/unmanned-ground-vehicles-face-a-rare-market-challenge-in-south-america-horses/; European Defence Matters, "Paving the Way for Autonomy in Land Systems", 20 November 2023, https://eda.europa.eu/webzine/issue16/cover-story/paving-the-way-for-autonomy-in-land-systems.

72 U. Franke, "Drones in Ukraine and Beyond: Everything You Need to Know", European Council on Foreign Relations, 11 August 2023, https://ecfr.eu/article/drones-in-ukraine-and-beyond-everything-you-need-to-know/; Borsari and Davis, 2023; M. Hunder, "Ground Vehicles Are the New Frontier in Ukraine's Drone War", Reuters, 13 July 2023, https://www.reuters.com/world/europe/ground-vehicles-are-new-frontier-ukraines-drone-war-2023-07-13/.

73 J.A. Tirpak, "Unmanned Flying Teammates", *Air & Space Forces Magazine*, 7 October 2021, https://www.airandspaceforces.com/article/unmanned-flying-teammates/.

74 Boeing, "MQ-28", n.d., https://www.boeing.com/defense/mq28.

75 R. Briant, "La Synergie homme-machine et l'avenir des opérations aériennes", *Focus Stratégique*, September 2021, https://www.ifri.org/fr/publications/etudes-de-lifri/focus-strategique/synergie-homme-machine-lavenir-operations-aeriennes.

76 Presentation by Michael Depp at the UNIDIR 2023 Innovations Dialogue in Geneva, Switzerland.

77 W. Cai et al., "Cooperative Artificial Intelligence for Underwater Robotic Swarm", *Robotics and Autonomous Systems*, Vol.164, 2023, https://doi.org/10.1016/j.robot.2023.104410; J. Lima and N. Drozdiak, "NATO Turns to Underwater Drones to Better Deter Russia", Bloomberg, 28 September 2023, https://www.bloomberg.com/news/articles/2023-09-28/nato-turns-to-underwater-drones-and-ai-in-bid-to-deter-russia.

coalitions.[78] As such, interoperability (between systems, services and nations) considerations should be at the forefront of HMT efforts. Indeed, the Ukraine war has shown that wars are likely to be characterised by a mix of equipment from different nations with different levels of sophistication. For example, Ukraine has received various types of military aid from at least 41 nations worldwide.[79] Additionally, conflicts are likely to see the increased prevalence of civilian and dual-use technology adapted for military tasks.

The uneven development and adoption of HMT is likely to negatively affect the international interoperability of forces and further complicate multidomain and combined arms operations. HMT will need high degrees of tailoring to the user and system, the team in question, and the service the team is part of. While the technical and technological elements of HMT may be easily transferable to other team environments, the social and more relational elements may not be, e.g. the calibration of trust is built over time, highly personal, and contextual.[80] The creation of shared mental models, effective communication styles, and appropriate interfaces might vary significantly across cultures, languages, settings, and environments. Furthermore, global attitudes towards military AI are far from homogeneous, as exemplified by the variety of positions held by members of the UN Group of Governmental Experts on Lethal Autonomous Weapons Systems – the foremost forum for discussions on autonomy in weapons systems.[81] This translates to unequal military AI adoption rates and readiness levels.[82] Not all military budgets are equal, and a "'technologically-driven US military strategy is advancing so fast compared to European allies that, sooner rather than later, all-important NATO military interoperability might well become a thing of the past'".[83] This not only risks exacerbating political tensions over alliance burden sharing, but could affect operational realities in times of conflict.[84] The connection of sensors, AI-enabled weapons, and command-and-control structures learning and adapting in real time will create a suite of algorithmic-centric offensive and defensive capabilities that risk increasingly compressing the time between sensing, deciding, and acting. As a result, those not connected to the system might find themselves unable to keep up.[85] Indeed, should "AI applications become a necessity for warfighting in the future, states that lack AI capabilities may be less able to contribute to alliance operations".[86] Additionally, both technical and political challenges to data sharing will add a layer of complexity to battlefield AI interoperability.[87] Due to its reliance on advanced robotics and AI, these technological disparities will inevitably spill over into HMT, creating HMT "haves" and "have nots". The ensuing conundrums will negatively affect the capability of armed forces to be interoperable the more they rely on human-machine teams, which is true between services, systems and (mostly) other nations.

---

78 E. Lin-Greenberg, "Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making", *Texas National Security Review*, Vol.3(2), 2020, https://tnsr.org/2020/03/allies-and-artificial-intelligence-obstacles-to-operations-and-decision-making/.

79 Kiel Institute for the World Economy, "Ukraine Support Tracker", n.d., accessed 20 November 2023, https://www.ifw-kiel.de/topics/war-against-ukraine/ukraine-support-tracker/.

80 Greenberg and Marble, 2023.

81 Automated Decision Research, "State Positions", n.d., accessed 16 November 2023, https://automatedresearch.org/state-positions/.

82 Lin-Greenberg, 2020.

83 M. Dufour, "Will Artificial Intelligence Challenge NATO Interoperability", NATO Defense College Policy Brief No. 6, 2018, https://www.jstor.org/stable/resrep19838.

84 S.R. Soare, "What If... The Military AI of NATO and EU States Is not Interoperable?", in F. Gaub (ed.), *What If ... Not? The Cost of Inaction*, European Union Institute for Security Studies, 2021, pp.18-22, https://www.jstor.org/stable/resrep28677.6.

85 Dufour, 2018.

86 Lin-Greenberg, 2020.

87 Ibid.

2.3 Training and technological literacy

In a somewhat related manner, effective HMT – as well as simple coexistence with highly autonomous robotics – will require highly technologically literate personnel.[88] For the US DoD, "talent deficit … 'represents the greatest impediment to being AI-ready by 2025'".[89] This is not the case only to develop autonomous machine capabilities, but also to operate them. To maintain *meaningful* oversight over autonomous machine behaviours, suggestions and actions, war fighters will need to have a deep understanding of their machine counterparts. For example, sustained trust in machine systems typically requires high levels of knowledge about both how the system works and its limitations.[90] This entails that soldiers will "need to judge the reliability of AI-ML [machine learning] predictions and actions, determine algorithmic outputs' ethical and moral veracity, and judge in realtime whether, why, and to what degree AI systems should be recalibrated".[91] This will require the substantial retraining and upskilling of soldiers down to the lowest echelons. As some experts have noted, "The technological intensity of modern warfare recommends recruiting the technologically adept. … every scarce soldier will merit fighter-pilot-level investment in training, and accompanying robotics to extend his or her reach".[92] The challenges this presents are obvious. In case of high-attrition conflicts like that between Ukraine and Russia, force generation and regeneration are both vital and complex. A future force structure based on HMT will struggle to train and field troops well-trained enough to effectively be part of human-machine teams. This might entail that in high-intensity conflicts, such teams might not represent the majority of an armed force's structure the longer the conflict – and its attritional effects – lasts. Additionally, it is worthwhile for armed forces to further consider the effects of reliance on HMT on team dysfunction in the case of the loss of both a human and a robotic/non-human team member. Should the skills needed to interact with the machine teammate not be transferable among human team members due to lack of training, will the team still be functional? Will units based around HMT still be able to operate if their machine teammate and its capabilities are lost? This is very unlikely, and therefore represents new liabilities to take into account in planning military operations.

88 Black et al., 2022.

89 Ibid.

90 Neads et al., 2021.

91 Johnson, 2023.

92 J. Hasik, "Economy, Autonomy and Rethinking the Military", Center for European Policy Analysis, 8 November 2023, https://cepa.org/article/economy-autonomy-and-rethinking-the-military/.

Part 3: Conclusions and recommendations

All in all, human-machine *teaming* is not yet a reality. For now, today's technological realities do not enable us to achieve the level of relational – and social – complexity needed to qualify an interaction between a human and a machine as a *team*. The terminology is not trivial here. *Teaming* is a term borrowed from sociology, with a conceptual underpinning that should not be ignored. As advances in AI increasingly enable machines to exhibit previously unparalleled near-human-level capabilities, our relationship with them changes more than it has with previous technological evolutions. In a world where AGI is a theoretical possibility, with multibillion-dollar efforts directed towards achieving it, it is worth exploring this qualitative change in the relationship between humans and machines by utilising the term *teaming* in a way consistent with its sociological meaning. A continuum is therefore a useful way of representing human-machine relations. Guided by the relational complexity of an interaction and the technologies available (and needed) to enable it, one can understand where any human-machine interaction is positioned in relation to *teaming*.

Furthermore, discussions of technology in the military domain – especially as they relate to autonomy – should seek to better integrate the realities of warfare in their analysis. As an endeavour relying on the management of a relationship, HMT will be highly affected by the various battlefield stresses. If HMT requires conditions that are unlikely to be present in times of war, its benefits will be negated. Understanding and addressing the fact that war will adversely affect many of the requirements of a successful HMT, put pressures on the AI-enabled autonomy of machines, and affect how humans relate to autonomous machines will be as foundational to the success of HMT as the technological developments that will make it possible.

In light of the above findings, the following recommendations should be considered:

- The HMT continuum should be used as a framework through which to understand and analyse the spectrum of human-machine interactions, from using machines as tools to their acting as teammates. Due to the confusing use of the term HMT across academia and industry, the continuum can both help to provide clarity and contribute to managing expectations. By using the level of relational complexity of an interaction and technological maturity at hand as proxies, the continuum can help one understand where any given instance of human-machine teaming" is with respect to *teaming*, with true teaming requiring the highest level of relational and social complexity in the interaction and the most advanced level of technological maturity to allow for such complexity. When developing and acquiring systems or forecasting their capabilities, the continuum can be a guide to understand the technological needs to enable any desired level of relational complexity between humans and machines (i.e. tool, collaborator, teammate). For example, *teaming* will need more fundamental research in AI and other related fields, therefore necessitating substantial investments and incurring larger risks, while *collaboration* can be enacted with today's most advanced AI, robotics, computing and sensor capabilities (among others).

- Initiatives between technical and non-technical institutions on the issue of HMT should be fostered. Because HMT is a socio-technical artefact, maximising only the technological aspects of human-machine interactions will only be part of the answer, and will necessitate a deeper understanding of the social aspects of human-machine relations. It will be especially critical to gain a deeper understanding of humans' social attitudes towards machines that exhibit increasing degrees of autonomy and human-like behaviour. Maximising the technical side of the equation will not lead to the development of true human-machine *teams*. Research should therefore seek to better integrate the social, cultural and cognitive sciences into HMT.
- It is essential to understand the training and knowledge requirements advanced HMT entails for military personnel, down to the lowest echelons.[93] Investing in HMT and seeking to develop capabilities to that effect will entail significant personnel upskilling and strong digital/AI literacy. Developing capabilities in already highly skilled segments of the armed forces should be a priority. But as technologies democratise, interactions with highly autonomous and capable machinery will no longer be the sole prerogative of highly educated or skilled personnel. Preparing personnel at all levels for interactions with highly complex, capable and autonomous systems will be as important as developing the systems themselves. Additionally, training programmes should seek to negate the potential effects of machine dependency. As machine counterparts take on an increasing array of roles, military personnel should retain the ability to undertake these tasks alone should their machine "teammates" be compromised. As discussed above, it is relatively safe to assume that the autonomous systems comprising the "machine teammate" element of HMT will eventually fail, either due to adversarial actions or environmental pressures. To avoid team dysfunction, it is an essential requirement that personnel be appropriately trained to operate both with *and without* their machine counterparts.
- High-fidelity training and testing should be at the centre of the development of military HMT. Armed forces, developers, and other aspiring adopters of military HMT should seek to understand how the success variables of HMT are affected by wartime realities – especially time sensitivity, high stress levels, and cognitive overload. It is essential to understand how the relationship between human and machine is affected in times of war, if human-machine teams are to be successfully operationalised and fielded in combat situations. Additionally, lessons learned from high-fidelity training and testing should inform backup system requirements and cost implications. While one way to avoid team dysfunction if a system fails or is compromised is to enable military personnel to effectively operate without their machine counterpart, another way is to ensure that backup systems are available. The costs of such systems will depend on the function performed by the system or team. The cost of maintaining backup systems will increase with the criticality of the function, while the repercussions of lacking such systems will also increase. The decision to develop and deploy human-machine teams for any given tasks should therefore be made by considering the potential for system failure, the ability for a team to operate without the system, the costs, and the ability to deploy backup systems weighed against the consequences of not having them.

93 Johnson, 2023.

- Currently, the field of HMT is principally focused on AI, both as an enabler of and enhancer of HMT. While AI is still relatively brittle, advances in AI are accelerating, and are becoming functionally closer to humans in many domains. AI advances that seek to mimic human behaviours, such as natural language, will be particularly key, because they will have a direct impact on how humans perceive their machine counterparts. AGI remains a debated capability, but whether it is achievable or not, substantial efforts towards creating it are under way.[94] The technological monitoring of AGI is of the utmost importance because – at least in theory – AGI could enable many of the necessary behaviours for a machine to be a genuine teammate.
- Developments in the field of neurotechnology should be carefully monitored. Neurotechnology is a growing field that currently remains in its infancy, but is highly likely to converge with the field of HMT. AI and neurotechnology are inextricably linked, and efforts to create human-machine symbiosis to achieve HMT will inevitably come through the development of technologies that are able to interface the human brain with machines.[95] The neural control of computers and robotic systems could do away with important points of friction inhibiting HMT. For example, removing the need for an interface (via a screen, tablet, etc.) to mediate human-machine interaction could help to reduce operator cognitive overload.[96] Technological monitoring of non-military domains such as neuro-marketing to identify possible breakthroughs is also of the utmost importance.
- Armed forces should understand how HMT and the integration of ever-more advanced AI into military operations affect interoperability with potential allies on the battlefield. Being overly ahead of or behind key partners will negatively affect the interoperability of armed forces, while domain warfighting characteristics will also affect the modalities of HMT. For now, the air domain is likely to be the most receptive to AI-enabled capabilities and carries with it the most potential for the military use of HMT.

94 S. Altman, "Planning for AGI and Beyond", Open AI, 24 February 2023, https://openai.com/blog/planning-for-agi-and-beyond.

95 Rickli and Mantellassi, 2022.

96 Ibid.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

armasuisse

GCSP
Geneva Centre for
Security Policy

ISBN: 978-2-88947-011-2