AI for peace: mitigating the risks and enhancing opportunities*

Michele Giovanardi[1,2] 🔵

[1]Florence School of Transnational Governance, European University Institute, Firenze, Italy
[2]University for Peace, Ciudad Colon, Costa Rica
Email: michele.giovanardi@eui.eu

Abstract

This article focuses on measuring the impact of artificial intelligence (AI) on the peace and security agenda, taking stock of recent initiatives and progress in this area. While there is a keen awareness of the fact that AI can be weaponized to become a tool of power politics and military competition, there is comparatively less systematic attention paid to what technology can do for peace. While it is important to address risk mitigation, equal space should be given to thinking about how to harness the peace potential of AI on a large scale. This study follows a series of publications that aim to assess the impact of technological innovation on peace, also referred to as PeaceTech, Global PeaceTech, peace innovation, or digital peace-building. The first section provides an overview of the debate on the impact of AI on peace and conflict. The second section examines conceptual frameworks and measures of the impact of AI on peace and conflict. The third section looks at the risks to peace and conflict posed by the use of AI and possible governance measures to mitigate them. The fourth section provides examples of AI-enabled initiatives that are having a positive impact on peace, providing a compass for public and private investment. The conclusion offers policy recommendations to advance the AI for peace agenda.

Policy Significance Statement

The policy relevance of this article lies in its call for a balanced approach to artificial intelligence (AI) that not only addresses potential risks and weaponization but also explores AI's substantial role in promoting peace. By emphasizing the need for systematic attention to the peace potential of AI, the study contributes to the ongoing discourse on how to regulate and govern AI for optimal societal benefit. The proposed policy recommendations serve as a guide for policymakers and private investors, urging them to consider the broader positive impact of AI on peace and security, ultimately shaping governance initiatives and promoting the responsible use of AI in line with peacebuilding objectives.

1. Introduction

Although still ill-defined and multifaceted, the concept of "artificial intelligence (AI)" permeates the public discourse and is ubiquitous in conferences, academic journals, and political debates. AI is presented as the

---

*The article has been updated since original publication. A notice detailing the change has also been published: https://doi.org/10.1017/dap.2024.92

"great enhancer" of all the good and evil inherent to our human nature. AI is deemed to bring new solutions to old problems, such as the treatment of incurable diseases, the management of extreme weather caused by climate change, more efficiency, and new tools to alleviate poverty and inequalities, a new possibility for cross-cultural communications and education, but AI is also blamed for the deepening of existing problems, such as the erosion of the information environment through the mass spread of disinformation, the manipulation of public opinion, increased inequalities and concentration of power, and the looming perspective of new weapon systems such as lethal autonomous weapons (LAWs). Techno-pessimist narratives contribute to a sense of urgency that coexists with the hope brought about by techno-optimist accounts. This generates a sense of "timely agency" and responsibility in steering AI in the desired direction before it reaches irreversible tipping points. Regulation and good governance are essential to mitigate AI's potential risks and capitalize on its economic and social benefits. To this end, a series of governance initiatives have been implemented worldwide. A non-exhaustive list includes the United Nations Educational, Scientific and Cultural Organization (UNESCO) recommendations on AI, the Organisation for Economic Co-operation and Development (OECD) AI Principles, the G20 Principles for Responsible Stewardship of Trustworthy AI, the G7 Statement on the Hiroshima AI process, the European Union (EU) AI Act, the US Executive Order 14110, and the Global Digital Compact to be adopted at the United Nations (UN) summit of the future in September 2024.

While recent studies have focused on the impact of AI on the work environment (Garcia-Murillo and MacInnes, 2024), climate change, health, and Sustainable Development Goals (SDGs) broadly (Fowdur et al., 2024), this article will focus more specifically on the impact of AI on peace and conflict and its measurement therein. This study keeps up with a series of publications aimed at assessing the impact of technological innovation on peace, also labeled as PeaceTech, Global PeaceTech, Peace Innovation, or Digital Peacebuilding (Bell, 2024). The first section offers an overview of the debate on the impact of AI on peace and conflict. The second examines conceptual frameworks and measures of the impact of AI on peace and conflict. The third section addresses the risks for peace and conflict posed by the deployment of AI and the possible governance mitigation measures to be put in place. The fourth section provides examples of AI-enabled initiatives that positively impact peace, representing a compass for public and private investments. The conclusion offers policy recommendations to advance the agenda of AI for peace.

2. AI, peace, and conflict: a short literature review

While the literature has focused extensively on the ethical concerns related to the development and deployment of AI, we can find comparatively less literature explicitly addressing the impact of AI on peace and conflict. AI, with its transformative potential, stands on the precipice of reshaping every facet of human interaction, including the delicate domain of peace and conflict. This review delves into the complex narratives surrounding AI's impact, unveiling its alluring promises and daunting risks.

A clear analytical difference is emerging in the literature concerning the distinction between military and nonmilitary AI (or civilian AI[1]). Although the difference between the two categories is blurred as technological innovations in the military have historically been used for civilian purposes and vice versa, we can distinguish based on the purpose, funding, oversight, and ethical considerations. Civilian AI is developed by private companies, research institutions, and universities for primary research and commercial purposes. Funding often comes from venture capital, grants, and commercial partnerships. Its primary objective is to improve the human experience and increase efficiency in industries like manufacturing and logistics, fostering personalized healthcare experiences and other services. Conversely, military AI focuses on enhancing military capabilities, gaining a strategic advantage, and potentially

---

[1] *Civilian AI* is the most commonly used term and refers to AI for nonmilitary applications. It encompasses a broad range of uses, including healthcare, manufacturing, transportation, and consumer products. *Civic AI* specifically focuses on AI applications within the context of citizens and government, often related to public services, social good initiatives, and civic engagement. This might include using AI for things like improving public transportation, optimizing waste management, or enhancing citizen participation in government processes.

reducing casualties. Its development is tailored to military applications like target identification, surveillance, autonomous weapons, logistics, and intelligence gathering. Military AI funding comes from government budgets allocated for military research and development, with the participation of private companies, government agencies, and research institutions, often with classified or restricted information. Moreover, civilian AI applications often fall under the purview of industry standards, consumer protection agencies, and government regulators, with specific regulations varying based on the technology's domain, such as data privacy, safety standards, and transparency. Military AI, on the other hand, is primarily subject to internal military regulations and international treaties like the laws of war. While some level of transparency is crucial for international cooperation and responsible development, the nature of military applications often necessitates a different balance between openness and security (Table 1).

These ideal types are difficult to disentangle in reality, and as noted by Creutz et al. (2024), it is difficult to regulate one without hindering the other, being AI a general-purpose technology. Nevertheless, this analytical distinction can help us assess the impact of AI on peace, where negative and positive effects on peace can be brought about in both domains.

In the report "AI and International Security," the United Nations Institute for Disarmament Research (UNIDIR) outlines the main areas of AI-related risks in the peace and security domain. In particular, three global security risk categories are identified: miscalculation, escalation, and proliferation. The first one is concerned with the uses of AI, which, by presenting a biased or flawed operational picture, may undermine decisions on the use of force or pave the way for a deterioration in international relations. The second focuses on the potential for AI technology to lead to intentional or unintentional escalation of conflicts, and the third one is the risk of AI being misused for the proliferation of new weapons, including weapons of mass destruction (Puscas, 2023). On the latter two points, the specter of LAW systems (LAWSs) looms large. The prospect of unintended escalation and the chilling absence of human accountability associated with LAWS fuel widespread concerns (House of Lords, 2023). According to Garcia (2018), AI weapons will significantly diminish the ceilings for war to start and deepen power asymmetries between countries that can or cannot afford AI weapons. Garcia (2024) points to the growing dehumanization of the armed forces powered by AI, which reduces human beings to mere data for pattern

Table 1. *Key differences between civilian AI and military AI*

	Civilian AI	Military AI
Purpose	Commercial, public good, healthcare, personal services (e.g., personal assistants)	Enhancing military capabilities (e.g., logistics, surveillance, intelligence gathering), gaining strategic advantage, reducing casualties
Development	Private companies, research institutions, universities	Private companies, government agencies, research institutions (with classified information)
Funding	Venture capital, grants, commercial partnerships	Government budgets allocated for military research and development with the participation of private sector
Oversight	Industry standards, consumer protection agencies, government regulators	Internal military chains of command and international organizations
Ethical considerations	Biases, privacy, transparency, accountability	Legality, proportionality, potential for unintended escalation
Impact	Wide–ranging economic, social, and cultural impacts	Impact on the nature of warfare, international security, and the balance of power

recognition technologies. Algorithms that contain instructions to strike and kill human beings can increase the likelihood of unpredictable and unintended violence.

AI can enhance security and preventive measures through AI-powered early warning and response mechanisms. However, AI-powered facial recognition poses a real dilemma where the opportunities for increased security and predictive policing need to be balanced with risks for human rights and democracy. For instance, on 4 July 2023, the European Court of Human Rights, in the Glukhin v. Russia case, unanimously found a violation of Article 8 (right to respect for private life) and Article 10 (freedom of expression) of the European Convention of Human Rights (ECHR) in Russia use of facial recognition technology against Mr. Glukhin following his demonstration in the Moscow underground. The Court concluded that the use of highly intrusive technology is incompatible with the ideals and values of a democratic society governed by the rule of law and demonstrated the risks this technology poses if managed by repressive governments or malign actors (Palmiotto and González, 2023).

On the hopeful side, AI presents itself as a potential peacebuilding champion. Mäki (2020) identifies three areas of opportunities for AI in peacebuilding: AI-assisted conflict analysis, early warning systems, and support for human communication. AI-assisted communication analysis can pave the way for more productive dialogue and consensus building within peace negotiations (Höne, 2019). In his "The Peace Machine," Honkela (2017) argued that machine learning (ML) could offer unprecedented understanding and contribute to world peace. By deepening the understanding of language, AI could detect conflicts in meaning, reduce misunderstandings, help reach agreements, and reduce social conflict. Thomson and Piirtola (2024), reflecting on the role of digital dialogues in Sudan carried out by CMI—Martti Ahtisaari Peace Foundation in July 2023, argue that AI has shown great potential in analyzing large amounts of complex data collected from populations in conflict, as well as in expanding the inclusiveness of peace processes. AI-powered platforms that enable digital dialogues provide virtual spaces for participation, sharing of opinions, and contribution of ideas and can, therefore, increase the accessibility, transparency, and scalability of dialogue processes. While this is a valuable complementary tool for mediators with great potential, especially when access is limited, it should be handled carefully, avoiding "technosolutionism." Digital dialogue requires careful planning, active engagement with targeted groups, and proper assessment of potential limitations.

Fueled by big data analysis, early warning systems can predict and preempt simmering tensions before they erupt into violence (United States Institute of Peace, 2023). AI can play a role in monitoring social media to detect hate speech and inflammatory language about marginalized groups, women, and the peace process that could undermine a just and inclusive solution and adopt timely preventive measures. Vigilant AI-powered non-weaponized drones, combined with satellite imagery, can play a role in monitoring ceasefire violations, reducing incidents and harm to peacekeepers, monitoring disarmament processes, or documenting human rights violations (Grand-Clément, 2022). AI can also assist in drafting agreements, using historical data on peace agreements, taking into account patterns of success in past agreements. This can inform mediation initiatives and AI-assisted dialogues (Wählisch, 2020).

While AI can support peacebuilding and mediation through AI-assisted data analytics and digital platforms, in the civilian domain, AI can also be used to erode peace via, for instance, the spread of disinformation and manipulation. The weaponization of AI for disinformation and propaganda campaigns poses a grave threat, manipulating public opinion and eroding trust (Radsch, 2022). Albrecht et al. (2024) analyzed disinformation in sub-Saharan Africa with a focus on AI-altered information environments. They found that AI is a driver of disinformation that adversely impacts peacebuilding efforts, but at the same time, there are several ways to mitigate AI-powered disinformation through content moderation procedures, the involvement of grassroots civil society organizations, and the promotion of media literacy campaigns.

Finally, AI has a positive role to play in peace and human rights education (Kurian and Saad, 2024) and has a vital role in enhancing democracy and building more sustainable and resilient societies. Buchanan and Imbrie (2022) recognize that AI development is not solely a technical issue but sits in a wider geopolitical context with nuances. Similar rules and standards might have different impacts in different socio-technical systems. To autocrats, AI offers the prospect of centralized control at home and

Table 2. *Examples of uses of AI in conflict and peace*

	AI for conflict	AI for peace
Communication	Automated disinformation campaign fueled by GenAI	AI–powered strategic communication and chatbots, AI tools to track and fight disinformation
Cybersecurity	AI–assisted coding and hacking	AI–enabled cybersecurity awareness and mitigation measures
Autonomous robots	Use of LAWs	Drones for humanitarian relief and monitoring at low cost and high efficiency
Centralization vs decentralization	Centralization of power and control	Decentralization through AI–powered education and digital participation
Intelligence	Gathering of data and intelligence, for repression and foreign interference	Gathering of information for tailored peacebuilding initiatives and monitoring of human rights abuses
Democracy and elections	Political manipulation and foreign interference operations	Inclusivity and digital democracy

asymmetric advantages in combat, while democracies, bound by ethical constraints, might be unable to keep up. However, the authors argue that AI does not foster tyranny and that it can actually benefit democracy. AI can support digital democracy by organizing citizens' preferences into actionable policy proposals through AI-enabled assistants, which can help public service and public information in GovTech. AI can also improve peace indicators by making our cities smarter, more sustainable, and safer. AI can be applied to digital twins, which are digital replicas of cities in which different scenarios can be tested and implemented, including through AI-enabled simulations based on the analysis of large datasets.

These are all combined technologies. AI pattern recognition or analytics cannot work without a proper data collection mechanism. Data collection, in turn, requires hardware such as sensors, satellites, mobile applications, and Internet infrastructure. The interconnectedness of these domains, including in relation to cybersecurity and the vulnerability of connectivity infrastructure, should be considered in a holistic assessment of AI and other emerging technologies on peace and conflict dynamics (Table 2).

3. Measuring success: What AI and what peace?

As shown in the previous paragraph, the menu of options for applying AI to positive social functions is vast, and the list is likely to grow as more investment is made in innovative projects. However, in order to make a case for AI's impact on peace, academics and practitioners should define clear measures of success. Fortunately, existing frameworks such as the SDGs and peace indices offer good starting points. However, choosing the right measures remains crucial.

In peace studies, we find definitions of peace that not only consider peace as the mere absence of war (what is referred to as "negative peace") but something more nuanced and comprehensive, e.g., the presence of harmonious relationships in societies and the absence of structural violence (Galtung, 1964). "Positive peace" is defined by the Institute for Economics and Peace as the attitudes, institutions, and structures that create and sustain peaceful societies. In other words, while it is true that we cannot have peace in the presence of war, it is also true that we may not have peace in the absence of war, due to society's lack of resilience to the shocks that might bring out conflict, the presence of structural and cultural violence that can eventually turn into direct violence (Ercoşkun, 2021).

This conceptualization of peace allows for a wide range of possible measures of the impact of AI on peace, depending on the definition of peace adopted. Below, five measures of peace are suggested, from the least to the most inclusive, in terms of a number of factors counting as leading to peace.

The Armed Conflict Location & Event Data (ACLED) Conflict Index, produced by the ACLED Project, assesses political violence in every country and territory in the world. The level of conflict is ranked based on the four indicators of deadliness, danger to civilians, geographic diffusion, and armed group fragmentation. AI could improve this domain of "negative peace" by reducing numbers on one or more of the four indicators, or as a balance of contrasting effects (e.g., higher geographical diffusion but less deadliness due to the employment of unmanned and autonomous weapons).

The Global Peace Index (GPI) is a measure of "negative peace" by the Institute of Economics and Peace. It combines 23 indicators related to domestic and international conflict, societal safety and security, and militarization, such as battle-related death, terrorism, homicide, violent crime, military expenditures, and UN missions. This is a more inclusive measure of violence, which also considers intrastate and intragroup violence but still refers to the domain of negative peace, or peace, as the absence of direct violence and conflict. All 23 indicators forming the index could be affected significantly by military and civilian AI.

The UN Sustainable Development Goal 16 (SDG16) is one dedicated to peace, justice, and strong institutions. It comprises 24 indicators organized around 12 targets ranging from reducing violence, mortality rates, and corruption to ending human trafficking, strengthening the rule of law, public access to information, and identity for all. Each target is assigned to specialized UN agencies, which favor innovation and measurement of the impact of AI within the UN workstreams and projects. This measure can be associated with "positive peace," as it addresses not only active conflicts but also some of their underlying causes.

The Positive Peace Index (PPI) is a measure of "positive peace" and is made of 24 measurable parameters. These parameters have nothing to do with conflict per se (e.g., combat death) but with societal conditions that would lead to lasting peace. It includes measures such as gender equality, healthy life expectancy, control of corruption, regulatory quality, financial institutions index, gross domestic product (GDP) per capita, open government openness and transparency, government effectiveness, access to public service, and freedom of the press. While AI can clearly contribute to the functions that underpin the conditions for peaceful societies, these initiatives often do not fall under the label of PeaceTech or Digital Peacebuilding, but rather under GovTech and public sector innovation. However, it is important to recognize that these functions are also predictors of peace. Countries deteriorating in the PPI are more likely to deteriorate in the GPI.

The Peace Data Standard (Guadagno et al., 2018) is the most inclusive measure. It considers peace a set of positive, prosocial behaviors that maximize mutually beneficial positive outcomes resulting from interactions with others. It includes all functions that create cooperation and trust among people, such as the use of home exchange apps, e-commerce platforms, or social media. It focuses primarily on cooperative intergroup interaction mediated by technology (Figure 1).
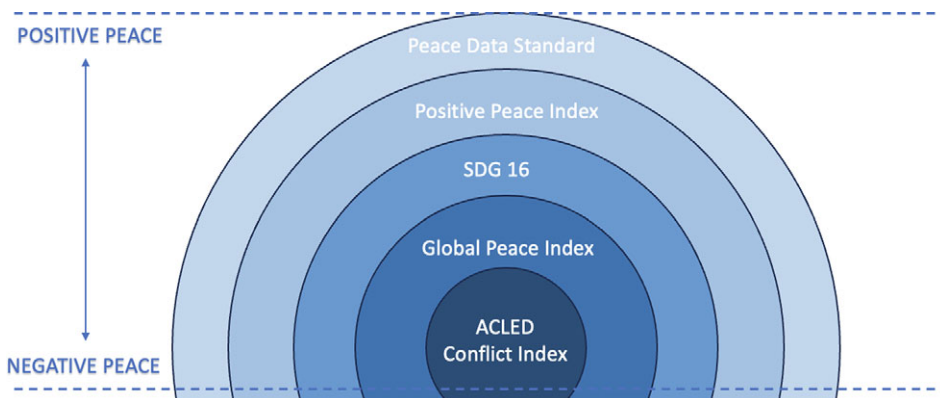

Figure 1. *Different levels of analysis where AI can have an impact on peace.*

This classification should not be seen hierarchically, with negative peace-oriented measures given more salience and relevance than positive peace-oriented measures. Indeed, a ceasefire, for example, may not last if it is not underpinned by the social conditions that make peace durable and sustainable over time, as captured by positive peace indicators.

In measuring the impact of AI on peace, we should not only ask about "what peace" we are talking about but also "what AI." Definitions of AI are also fraught with nuances and crucial differences, leading to analytical confusion. AI is a broad concept that encompasses the creation of intelligent machines that can simulate human thought and behavior. It involves the development of systems that can exhibit intelligence without explicit programming, using algorithms such as reinforcement learning and neural networks. ML, on the other hand, is a subset of AI that focuses specifically on the ability of machines to learn from data without being explicitly programmed. It involves extracting knowledge from data to make predictions or decisions based on historical data, using algorithms that learn from experience. ML is widely used in applications such as online recommendation systems and personalized services. However, the two terms often overlap, as most AI applications currently discussed are ML-based. Large language models (LLMs) like Generative Pre-trained Transformer 4 (GPT-4), Claude 2, or Gemini also use ML to process vast amounts of text data to infer relationships between words within the text. In essence, the term "AI" alludes to the idea of replicating human-like intelligence through machines, which could lead to the inclusion in the definition of more and more applications and technological innovations produced for this purpose, even beyond ML, in the future. Current legislation, such as the EU AI Act, allows for this broad interpretation as it focuses on the objectives of the AI technology as "software that generates outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Article 3), rather than the technology underpinning it, being it based on statistical approaches or ML, including logic and knowledge-based approaches (see Annex I of the AI Act).

4. Mitigating the risks: the governance challenge

Are we employing AI effectively as an enhancer of peace and sustainable development? Will AI contribute to more security or more insecurity in our societies? Will AI be a stabilizing or destabilizing factor in global geopolitical competition? Answering these questions requires a comprehensive understanding of AI governance, not only at the national level but also at the transnational and global levels. The parallels between the AI revolution and historical changes, such as the Industrial Revolution, offer valuable insights. However, in the field of international security policy, a more nuanced comparison emerges, particularly with the nuclear revolution in military strategy. The comparison with the nuclear revolution revolves around the potential parallels between the deterrence capabilities of nuclear weapons and those exhibited by autonomous weapons with AI. Just as the superpowers engaged in multilateral agreements for de-escalation and arms control during the nuclear age, a similar approach may be needed to navigate the complex landscape of AI.

A more fitting analogy might be found in conventions such as the "Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines." This comparison is based on the observation that AI-associated technologies could become widely accessible at lower skill and cost thresholds, similar to the challenges posed by landmines. Unlike the high skill and technical knowledge required to launch nuclear strikes, AI technologies could be within the reach of a wider range of actors. An even more precise parallel might be drawn with frameworks regulating biological and chemical weapons, where international conventions and institutions are in place to monitor and enforce responsible practices.

The lessons learned from historical analogies, such as the nuclear revolution, the Anti-Personnel Mine Ban Convention, and frameworks for biological and chemical weapons, underscore the urgency of establishing comprehensive international agreements and mechanisms to navigate the complexities of AI in the pursuit of a secure, sustainable, and ethically governed deployment. Firstly, establishing international conventions and collaborative institutions for the governance of AI can ensure the alignment of the development, deployment, and use of AI with ethical principles, human rights, and global stability. Secondly, the presence of an "epistemic community," a group of scientists and activists with a common

scientific and professional language and point of view that can generate credible information, is a powerful tool for mobilizing attention toward action. Thirdly, these efforts can be complemented by civil society activity. International advocacy campaigns can serve to draw attention to different levels of diplomatic and global action (Carpenter, 2014). Such is the case with the "International Committee on Robot Arms Control" or the "Stop Killer Robots" campaign, gaining traction worldwide.

Garcia (2018) examines twenty-two existing treaties that have acted within a "preventive framework" to establish new regimes to ban or control destabilizing weapons systems. Based on the successes of these initiatives, the author argues for adopting "preventive security governance" as an effective strategy to mitigate the risks posed by AI to global peace and security. This could be done by codifying new global norms based on existing international law and clarifying universally shared expectations and behaviors. Article 36 of the 1977 Additional Protocol to the Geneva Conventions requires states to conduct reviews of new weapons to determine compliance with international humanitarian law. However, only a few states regularly conduct weapons reviews, making this transparency mechanism insufficient as a tool for creating security frameworks for future weapons and technologies. This indicates that regardless of the mechanism that may be put in place, for these new regimes and standards to be effective, it is important to address the incentives that countries would have to ban these types of weapons, namely the fact that they will be much more widely available than other types of weapons and that their industries have an interest in preventing their products from being associated with civilian casualties. Garcia (2018) noted that individual states could emerge as champions in supporting such initiatives, catalyzing substantial progress in disarmament-focused diplomacy.

5. Enhancing the opportunities: programs and investments

While it is crucial to address risk mitigation, equal importance should be given to exploiting at scale the opportunities for peace offered by AI in a mutually reinforcing mechanism. While there is acute awareness of the fact that AI can be weaponized to become a tool of power politics and military competition (Leonard, 2021), there is comparatively less systemic attention to what technology can do to achieve peace and provide humanitarian aid and development. In this regard, public and private actors, as well as citizens, should ask how advances in AI can contribute to scaling up cooperation and innovation in peacebuilding.

As mentioned above, AI can support data-driven decision-making, conflict prediction, and early warning systems via social media and large dataset analysis, humanitarian aid and disaster response by increasing efficiency and resource optimization, communication and diplomacy by using AI-powered translation and analysis of diplomatic texts and speeches, and disarmament and arms control via monitoring and verification, and cybersecurity. AI can also play a role in peace education, mediation, conflict resolution, community engagement, and cultural understanding by developing ad hoc tools like the BridgeBot developed by the nonprofit organization Search for Common Ground in collaboration with TangibleAI. Another example is Remash, an AI-based dialogue tool that allows for real-time conversation between large populations and is used by many peacebuilding organizations to carry out their initiatives. Such tools are already in use by the UN Department of Political and Peacebuilding Affairs (UNDPPA) to hold AI-assisted dialogue with groups of up to 1,000 citizens in Yemen, Libya, and Iraq as part of official peace processes.

To scale up the opportunities offered by AI in peacebuilding, more public and private funding should be allocated for both the development of AI for peace projects and research on their effectiveness and impact on peace processes. A three-step cycle, including (1) mapping of initiatives, (2) analysis of impact, and (3) project investment, should be adopted and repeated within the different sectors of AI and peace and their enabling digital ecosystem. Although independent research can be seen by funders as an inefficient phase that stands between capital investment and project development, it provides a valuable compass for directing investment toward what works and what does not, thereby reducing costs in the long run. Independent research also provides valuable feedback to the practitioners in charge of project development so they can adjust their focus and have an even greater chance of success. Stakeholder mapping allows for exploiting synergies and economies of scale and avoiding duplication. To exemplify how the cycle can work in practice, the example of the employment of AI in the fight against disinformation will be presented (Figure 2).
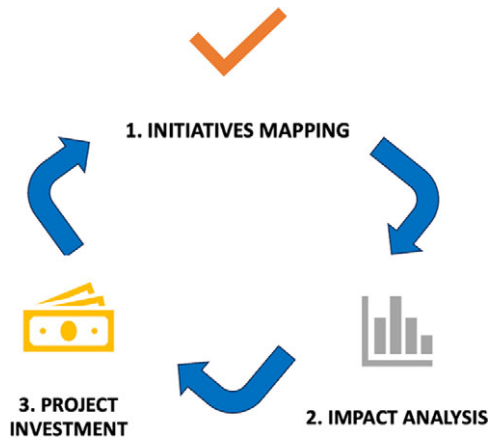
Figure 2. *A three-step cycle for investing effectively in PeaceTech projects (e.g., AI for peace).*

While we know that generative AI poses an existential threat to the information environment, AI can also mitigate these risks and help governments and civil society combat disinformation or its weaponization. Both public institutions and private companies have begun to invest in the development of AI-powered fact-checking and content moderation tools. Alongside these initiatives, grassroots initiatives are emerging that empower citizens and civil society organizations to combat disinformation using a combination of AI and human expertise to detect and dismantle false information. In their study of AI and disinformation, Pilati and Venturini (2024) use a "web mapping" approach to map the ecosystem of initiatives that employ AI in the fight against disinformation, a method based on the idea that hyperlinks can serve as proxies for social connections (Severo and Venturini, 2016). This methodology can be applied to different domains to untangle and visualize the complex web of relationships characterizing a given sociotechnical ecosystem. It makes use of statistical methods and dedicated software like Hyphe for web crawling and analysis and Gephi for network visualization (Jacomy et al., 2014, 2016) (Figure 3).

This mapping shows a dense network of initiatives from different stakeholders in the AI-enabled fight against disinformation, with commonalities and differences. The map can guide further analysis and results. For example, one of the main findings of Pilati and Venturini (2024) is that in the EU, innovation and development of AI tools are fostered by public funding, especially European funding under the Horizon program, in collaboration with higher education institutions, but not led by them. In contrast, in the United States, AI tools are developed mainly in higher education research environments, particularly in Ivy League universities, with core support from private funding.

These web mapping and data visualization techniques can be adopted in a variety of settings to inform programming and investments in AI for peace or, more broadly, technology for peace (PeaceTech) initiatives, identifying areas for potential improvement and collaboration among different stakeholders. For example, these data and results on AI to fight disinformation, as well as the effectiveness of the initiatives mapped, can be processed and evaluated by independent research institutions such as the European Digital Media Observatory, based at the European University Institute. The research and evaluation of this independent observatory inform, in a scientific advisory capacity, the decisions of the European Media and Information Fund, a multi-donor fund supported by private institutions such as Google and the Calouste Gulbenkian Foundation. The Fund allocates resources to a wide range of stakeholders with the goal of supporting collaborative efforts to debunk misinformation, amplify independent fact-checking, and enable targeted research and innovative tools to fight disinformation. New initiatives funded or scaled up under the European Media and Information Fund will, in turn, feed and drive change in the ecosystem map in a reiterative cycle. This simple scheme should be replicated in other areas of AI and peace or PeaceTech with the creation of multi-donor funds and independent research institutions that address the challenges and opportunities offered by AI in influencing patterns of peace and conflict.
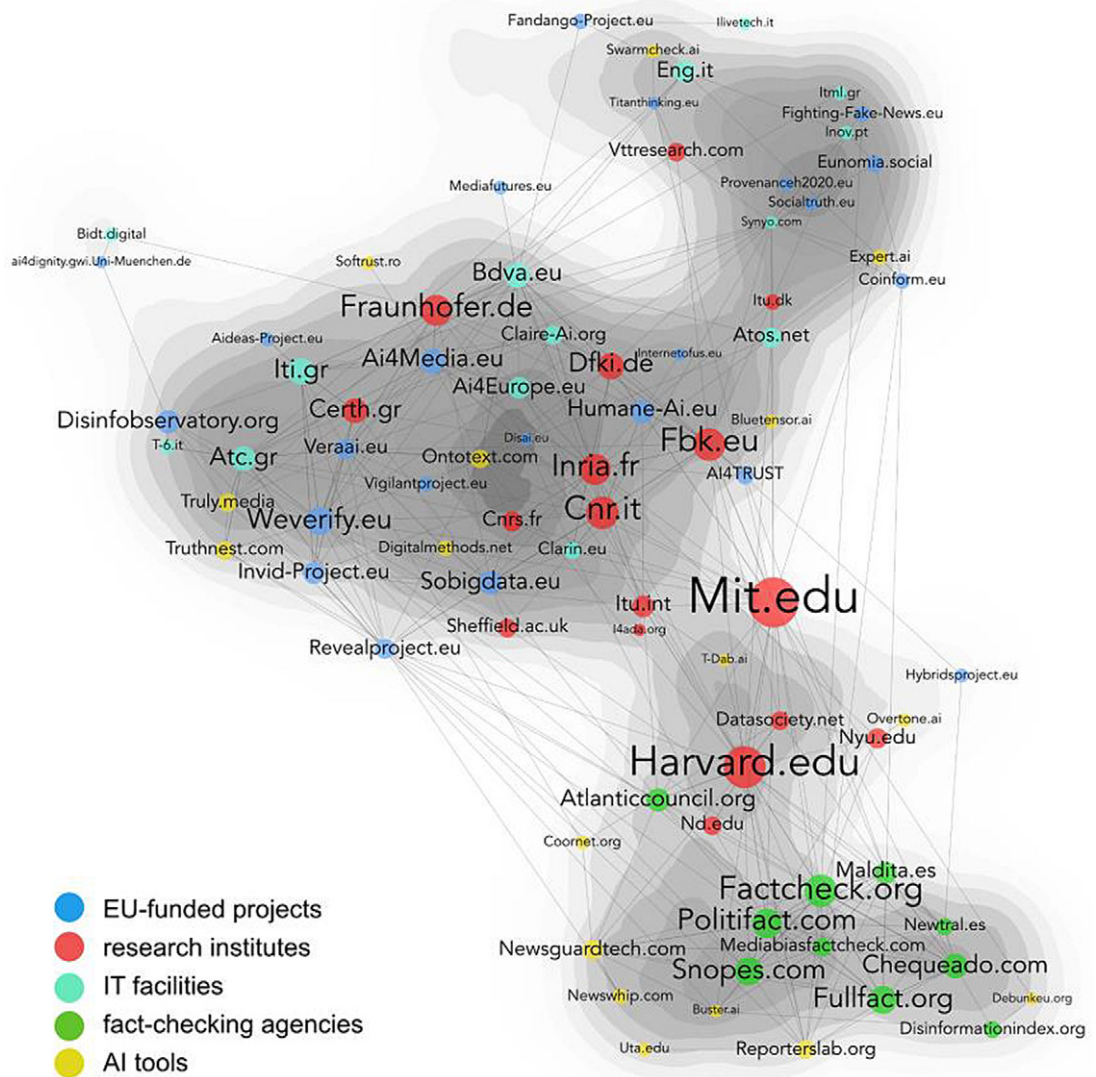
Figure 3. *Web map of AI initiatives against disinformation classified by category in Pilati and Venturini (2024).*

6. Conclusion: a way forward

The advancement of the interdisciplinary field exploring the impact of technology on peace necessitates clarity in defining terms and fostering a shared understanding. There is already an ecosystem of scholars, research institutes, civil society organizations, public institutions, private companies, philanthropic institutions, and venture capitalists working on technology for peace, including AI for peacebuilding (Bell, 2024). Acknowledging the emergence of this new interdisciplinary domain and the burgeoning PeaceTech movement, akin to the pivotal role the green movement played in championing climate sustainability, holds the potential to advocate for an agenda where technology and innovation for peace take a central position within the framework of global human development goals. This requires, however, a paradigm shift in public and private funding, in particular by complementing public funding for military technology projects with initiatives to support peace-oriented technology. This shift should be

accompanied by increased research and investment in the field, exploiting the potential of public–private partnerships. As pointed out earlier, simultaneously establishing an independent research institute and a multi-donor fund to support AI for peace initiatives would provide the most effective combination.

Navigating this complex landscape requires a balanced approach. While the responsible use of AI can provide crucial tools for peacebuilding, conflict mitigation, and saving lives, the ethical and legal concerns surrounding its development and application cannot be ignored. As underscored by the World Economic Forum, 2024, responsible development and governance are paramount to mitigate the inherent risks of AI. As suggested by the UN Advisory Board on AI, governance should be a key enabler, moving beyond self-regulatory practices and embracing international collaboration models seen in successful scientific organizations like Conseil Européen pour la Recherche Nucléaire (CERN) and European Molecular Biology Laboratory (EMBL) or successful public–private partnerships like GAVI, the Vaccine Alliance. The multifaceted challenges, such as the climate–migration–peace and security nexus and the delicate balance between conflicting rights like security and privacy, require comprehensive and inclusive approaches. Establishing a common language rooted in human rights as minimum standards is crucial for an inclusive and globally applicable framework. Ultimately, the impact of AI on peace will be shaped by the quality of public–private partnerships and investments, the multilateral governance of its potential threats, and citizens' participation and inclusion in these regulatory processes.

References

Albrecht E, **Fournier-Tombs E and Brubaker R** (2024) *Disinformation and Peacebuilding in Sub-Saharan Africa. Security Implications of AI-Altered Information Environments.* Research Report.

Bell C (2024) *PeaceTech: Digital Transformation to End Wars*. Cham, Switzerland: Springer Nature.

Buchanan B and Imbrie A (2022) *The New Fire: War, Peace, and Democracy in the Age of AI*. Cambridge, Massachusetts, United States: MIT Press.

Carpenter C (2014) *Lost" Causes: Agenda Vetting in Global Issue Networks and the Shaping of Human Security.* Ithaca, New York, United States: Cornell University Press.

Creutz K, **Javadi M**, **Onderco M and Sinkkonen V** (2024) *The EU and Military AI Governance.* FIIA. Working Paper. Available at https://www.fiia.fi/en/publication/the-eu-and-military-ai-governance (accessed 6 October 2024).

Ercoşkun B (2021) On Galtung's approach to peace studies. *Lectio Socialis* 5(1), 1–8.

Fowdur TP, **Rosunee S**, **Ah King RT**, **Jeetah P and Gooroochurn M** (eds.) (2024) *Artificial Intelligence, Engineering Systems and Sustainable Development: Driving the UN SDGs*. Leeds, United Kingdom: Emerald Publishing Limited.

Galtung J (1964) What is peace research? *Journal of Peace Research* 1(1), 1–4.

Garcia D (2018) Lethal artificial intelligence and change: The future of international peace and security. *International Studies Review* 20(2), 334–341.

Garcia D (2024) Algorithms and decision-making in military artificial intelligence. *Global Society* 38(1), 24–33.

Garcia-Murillo M and MacInnes I (2024) Introduction to the handbook of artificial intelligence at work: Interconnections and policy implications. In *Handbook of Artificial Intelligence at Work* (pp. 1–14). Cheltenham, United Kingdom: Edward Elgar Publishing.

Grand-Clément S (2022) *Exploring the Use of Technology for Remote Ceasefire Monitoring and Verification*. Geneva: UNIDIR.

Guadagno RE, **Nelson M and Lock Lee L** (2018) Peace data standard: A practical and theoretical framework for using technology to examine intergroup interactions. *Frontiers in Psychology* 9, 734. https://doi.org/10.3389/fpsyg.2018.00734

Höne K (2019) *Mediation and artificial intelligence: Notes on the future of international conflict resolution. Diplofoundation, Geneva, Switzerland.*

Honkela T (2017) *Rauhankone: tekoälytutkijan Testamentti.* Helsinki, Finland: Gaudeamus.

House of Lords (2023) *Proceed with Caution: Artificial Intelligence in Weapon Systems.* Available at https://aura.abdn.ac.uk/bitstream/handle/2164/22616/Johnson_HL16_Proceed_with_Caution_VOR.pdf?sequence=1 (accessed 23 February 2024).

Jacomy M, Girard P, Ooghe-Tabanou B and Venturini T (2016) Hyphe, a curation-oriented approach to web crawling for the social sciences. In Mark R. Muldoon (ed.), *Proceedings of the International AAAI Conference on Web and Social Media.* United Kingdom: Manchester University (Vol. *10*, pp. 595–598).

Jacomy M, Venturini T, Heymann S and Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One 9*(6), e98679.

Kurian N and Saad C (2024) Where technology meets empathy: Using digital storytelling, gaming, and AI to teach about peace and human rights. In *Communication and Education: Promoting Peace and Democracy in Times of Crisis and Conflict.* Hoboken, New Jersey, United States: Wiley, pp. 148–163.

Leonard M (2021) *The Age of Unpeace: How Connectivity Causes Conflict.* New York, United States: Penguin Random House.

Mäki N (2020) *Between Peace and Technology–A Case Study on Opportunities and Responsible Design of Artificial Intelligence in Peace Technology.* Vantaa, Finland: Laurea University of Applied Science.

Palmiotto F and González NM (2023) Facial recognition technology, democracy and human rights. *Computer Law & Security Review 50*, 105857.

Pilati F and Venturini T (2024) *The use of Artificial Intelligence in counter-disinformation: A world wide (web) mapping.* Available at https://ssrn.com/abstract=4855024.

Puscas I (2023) *AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures,* UNIDIR, Geneva, Switzerland.

Radsch C (2022) *Artificial intelligence and disinformation: State-aligned information operations and the distortion of the public sphere.* Organization for Security and Co-operation in Europe. Available at https://www.osce.org/representative-on-freedom-of-media/522166.

Severo M and Venturini T (2016) Intangible cultural heritage webs: Comparing national networks with digital methods. *New Media & Society 18*(8), 1616–1635.

Thomson S and Piirtola A (2024) *Artificial intelligence and peacemaking – the case of digital dialogues in Sudan, CMI.*

United States Institute of Peace (2023) *A Role for AI in Peacebuilding.* Available at https://www.usip.org/publications/2023/12/role-ai-peacebuilding (accessed 23 February 2024).

Wählisch M (2020) Big data, new technologies, and sustainable peace: Challenges and opportunities for the UN. *Journal of Peacebuilding & Development 15*(1), 122–126.

World Economic Forum (2024) *The Global Risks Report 2024.* Available at https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf (accessed 23 February 2024).