



When Mediators Need Machines (and Vice Versa): Towards a Research Agenda on Hybrid Peacemaking Intelligence

*Andreas T. Hirblinger*¹ | ORCID: 0000-0003-1193-1903

Graduate Institute of International and Development Studies, Maison de la Paix – Chemin Eugène-Rigot 2A, P.O. Box 1672, 1211 Geneva, Switzerland
andreas.hirblinger@graduateinstitute.ch

Received 18 May 2021; accepted 5 October 2021

Abstract

The growing use of artificial intelligence (AI) in peace processes challenges the ways in which mediators conventionally generate and use knowledge to facilitate a convergence of conflict party positions. Mediators are increasingly unable to maintain their role as information brokers without the support of automated information gathering and analysis systems. However, the effectiveness of such AI-tools is likewise hampered by their limited autonomy across peacemaking contexts, by the paradigmatic nature of knowledge that is used to design AI-systems, and the influence of subjective factors that are difficult to measure. These challenges may be overcome by hybrid human-machine systems in support of peacemaking. This article formulates a research agenda for hybrid peacemaking intelligence that is concerned with the increasing interdependence of humans and machines. It suggests studying how agency is distributed in mediator-machine networks, the hermeneutics of machine-supported participatory approaches, and how hybrid arrangements change power relations in peacemaking.

¹ Andreas T. Hirblinger is an Ambizione research fellow at the Centre on Conflict, Development and Peacebuilding (CCDP), Graduate Institute for International and Development Studies, Geneva. He is also a visiting researcher at the Peace Research Institute Oslo (PRIO). His current research explores the effects of the increasing use of digital technologies by conflict parties, conflict stakeholders and those who aim to make or build peace. His past projects investigated the uses of Artificial Intelligence (AI) in peace mediation, the limits of data-driven approaches, and the emerging practices of digital inclusion. He holds a PhD from the University of Cambridge.

Keywords

peace mediation – artificial intelligence – conflict analysis – knowledge production – inclusion

Artificial Intelligence (AI) is a growing topic of concern for the practice and research of peace mediation. Commonly described as a ‘human-centered’ field, in which technologies play a subordinate role, contemporary third-party efforts to support the resolution of armed conflict are affected by AI in various ways. Intelligent machines influence how conflict stakeholders relate to each other, for instance, through algorithmically mediated social media platforms, and are proactively used by conflict parties to manipulate and spread information. Mediators increasingly recognize the potential of AI systems for analysis of the large amounts of data produced in the course of peace processes, not least due to the trend to enhance participation in peacemaking through digital means. Yet, the topic of AI has been scarcely explored in research on peace mediation, with the exception of a few noteworthy policy- and practice-oriented contributions (UN Innovation Cell 2018; Höhne 2019; Jenny et al. 2018; Wählich 2020). In fact, intelligence in general, not only the artificial type, has remained a fringe topic in peacemaking research.² This is surprising, given the relevance of intelligence for a diversity of other topics in mediation research that are concerned with the production of information, knowledge and expertise for peace mediation (Rauchhaus 2006; Convergne 2016).

This article reflects on how the increasing use of AI in armed conflicts and peace processes should change our understanding of, and research on, knowledge production for peacemaking. Innovations in AI hold the promise that the production of knowledge necessary for peace mediation – for instance about the conflict parties’ positions, perception or beliefs, or options for conflict resolution – may be automated with the help of intelligent machines, with varying degrees of autonomy and human supervision.³ I suggest that AI

2 A notable exception is Nathan’s (2014) exploration of the ‘intelligence requirement’ for UN mediation efforts, which focuses largely on aspects of gathering and analyzing information for mediators. Nathan views peacemaking intelligence similarly to military intelligence, focusing on the use of information available in the public domain. His discussion may at first seem far-fetched from the topic of this article. However, as I will demonstrate, all types of intelligence are closely dependent on the availability of relevant information and knowledge.

3 The scope of this article is limited to a discussion of AI relevant for the processing of information with the aim of producing peacemaking-relevant knowledge. This means that other applications of AI, such as for robotics, will not be covered in this article.

should be thought of as forming part of hybrid peacemaking systems, combining the strengths of human and machine intelligence, in efforts to produce knowledge that can facilitate convergence between conflict parties' positions. This article discusses how such hybrid peacemaking intelligence may enable new forms of reflexivity that encourage conflict parties to seek a shared understanding of the conflict. Yet, we must also be concerned with how agency is distributed in such systems, and their powerful effects on peacemaking efforts and outcomes.

The field of AI is concerned with how intelligent machines "can compute how to act effectively and safely in a wide variety of novel situations," as Russell et al. (2016: 29) put it. In the words of Bryson and Theodorou (2019: 306), intelligent machines have agency in the sense that they can take "actions" based on "perceptions" with the help of computation. It is important to note that understandings of AI vary. Some define it in terms of its ability to "think" and "act" like humans, while others prefer abstract criteria of rationality, measuring a machine's ability to make correct inferences and "do the right thing" (Russell et al. 2016: 31–37), such as drawing sound conclusions, making suitable recommendations, or correctly identifying objects. However, there is little disagreement about the fact that AI systems obtain their intelligence through algorithms, that is, sets of instructions that define a pathway of steps between inputs (what the machine perceives), and outputs (the action that the machine generates) (Basil & Bowen 2020: 295). Conventionally, AI researchers differentiate between 'symbolic' or 'expert' systems that generate algorithmic functions on the basis of existing knowledge, often provided by experts, and 'connectionist' systems that operate on the basis of training data, which generate functions from patterns observed in the data. Today, these two approaches are commonly combined in what has been called informed machine learning (ML) systems (von Rueden et al. 2020).

Importantly, an algorithmic approach also underpins many mediation efforts. When supporting conflict resolution, mediators commonly resort to reason and logic to persuade parties to agree on a solution (Wallensteen & Svensson 2014). Such reasoning processes not only require intelligence in terms of cognitive skills, but in terms of knowledge about the conflict and the parties as well, usually generated through variants of conflict analysis, employed in the search for a convergence of conflict party positions. Therefore, I suggest thinking of conflict resolution as a process of rationalization, through which the causes of conflict are identified and then linked to solutions to conflict. Importantly this entails both working with the "objective" causes of conflict, such as poverty or military aggression, and the "subjective" causes of conflict, such as the conflict parties' narratives and stances on a given matter. The tasks of

mediators are to make sense of these causes and find a pathway of steps that lead to their resolution.

Given the tremendous growth of data that is produced in the context of armed conflicts and peace processes, the appeal of AI systems lies in the possibility of automating such reasoning processes, or parts of them. However, as I will demonstrate, this increasingly requires the interdependent deployment of human and machine intelligence. For instance, AI algorithms employed in support of peacemaking are envisaged to help predict conflict outbreaks; assist with conflict analysis by gathering and assessing social media, event or geospatial data; support digital inclusion efforts that help determine conflict stakeholders' interests, needs and positions; or generate recommendations for peace agreements (Cederman & Weidmann 2017; Höhne 2019; Wählich 2020; Rutherford 2021; Arana-Catania, Van Lier, & Procter 2021). Such applications contribute at various moments to understanding the conflict and its possible solutions. While doing so, they straddle the line between "broad" tasks, such as learning, reasoning and perception, and "narrow" tasks, such as playing chess, diagnosing diseases, or writing poetry (Russell et al. 2016, 29–30).

However, AI-generated peacemaking knowledge matters little if it is not accepted and used by the humans involved in the peace process, including mediators, support actors, the broad array of conflict stakeholders, and parties to the conflict. Furthermore, for AI systems to have a productive role in dialogue efforts, they must be able to work with the discrepancies that stem from human intelligence. Peace processes are not only commonly characterized by conflict between the vested positions of the parties, but also by differences in the rationalizations – for instance, in terms of narratives, opinions, or arguments – that underpin their positions. This article will demonstrate that when discussing the utility of AI for peace mediation, it is unwise to separate machine intelligence from human intelligence that is involved in designing, maintaining and utilizing it. Rather than focusing on AI as a stand-alone autonomous system, I argue that it is more fitting to shed light on AI as part of human-machine arrangements that have hybrid intelligence: socio-technical systems that shape the production of peacemaking knowledge. Further, I argue that the emerging field of research on AI in peacebuilding should not only focus on data-technological aspects of the devices, methods and models used, but also on the distributed agency of humans and machines in efforts of knowledge production that can support the quest for political settlements.

The article focuses broadly on intelligent socio-technical systems that support the generation of peacemaking knowledge, that is, knowledge that mediators may use to navigate the conflict context, build mediation strategies, and

advance recommendations for conflict resolution. I will start by contrasting two discursive positions that dominate the discussion of AI in peacemaking, both of which hinder the exploration of hybrid intelligence systems for peacemaking: ‘human hubris’ grounded in an anthropocentric notion of peacemaking, and ‘machine hubris’ grounded in a technocentric notion of peacemaking. In response, I explore several challenges related to the generation of peacemaking knowledge in today’s digitalized peace processes, limiting the ability of humans and machines to gather information and reason in contexts that are heterogenous, ill-defined, shifting and ontologically insecure. I suggest that these challenges lead to the emergence of hybrid systems in which human and machine intelligence operate interdependently. In the second section, I suggest three ways through which we can move beyond the current debate, by developing a conception of hybrid peacemaking intelligence that integrates human and machine intellects, fostering a hermeneutical approach that considers the challenges of human-machine, reflexive knowledge production in the search for political settlements, and developing a greater sensitivity for the power relations that characterize hybrid peacemaking systems.

Artificial Intelligence in Peace Mediation: Between Machine and Human Hubris

This section first discusses how the mediation community predominantly views the relationship between mediators⁴ and intelligent machines.⁵ This reflection is important, because technology’s impact on society is not singularly the result of its functionality, but also of how it is discursively constructed by those who design and use it. This has been well encapsulated in Jasanoff’s (2004) notion of co-construction, which is useful in shedding light on how discursive and material factors jointly determine the agency of technology. Jacobsen and Monsees summarized the value of the co-productionist angle

4 I use the term “mediator” as a shorthand for mediators, their teams and mediation support organizations that help facilitate processes.

5 The reflections presented in this section have been gathered through informal conversations and the participation in workshops and exchanges with staff of leading mediation and mediation support organizations, in the course of the ‘Mediating Machines?’ research project, conducted in 2020. It must be noted, however, that knowledge on the topic differs vastly within the mediation community. The views presented here are less telling of the nuanced knowledge that some experts possess, but more of the broader views of actors with anecdotal or lay knowledge on AI. I have added references from academic and popular sources to demonstrate that the views summarized in this section can be plausibly generated from the public discourse on AI.

as bringing together an interest in the social production *of* technology and the social production *by* technology (Jacobsen & Monsees 2019: 26). For the study of peacemaking, this pertains to how artificial and human intelligence are viewed by the mediation community and with what effects, but also extends to the outputs of such intelligent systems, that is, the knowledge that they produce and what that knowledge is used for. For AI systems in peace mediation to be employed and studied effectively, the first step is thus to demystify its abilities, opportunities and limits.

The Human Hubris: Anthropocentric Peacemaking

Peace mediation and high-level diplomatic talks commonly have the appeal of an ancient, old-fashioned craft. Parts of the Iranian nuclear deal were reportedly negotiated in a wood-paneled room at the United Nations in Geneva, filled with heavy tables and leather furniture. According to one observer, all non-essential technical equipment was removed to avoid information leaking to the outside world and travelling through the channels of the early 21st century social media ecosystem. Behind such anecdotes lies the persistent notion that peace mediation is human-centered (Eleiba 2020). In this view, good mediation outcomes rely on the facilitation of the process by eminent persons, in most cases senior or elderly men – preferably current or former statesmen or high-level officials – who assert their authority through charisma and wisdom. It is therefore unsurprising that among many peacemakers, the idea of employing AI in mediation settings commonly produces aversion. AI tends to be associated with deep-fakes and autonomous weapon systems (Höne 2019: 9), or recommender algorithms that produce ‘echo chambers’ that lead to increased political polarization – things opposed to making peace. Computer scientific talk about ‘algorithms,’ ‘machine learning,’ and ‘neural networks’ commonly merges into an enigma that is not only difficult to decipher but remains detached from the everyday work of mediation teams.

This negative stance towards the utility of AI is supported by the view that peace processes should remain an *art*, and not turned into a *science* (Standfield 2020). In this view, agreements cannot be achieved by creating objective, scientific knowledge or engineered, mechanical solutions. This notion is engrained in mediation practice through the pop-cultural representation of peace processes. Almost all Google image searches for ‘peace accord’ and a country name will reveal pictures of senior male party representatives and mediators talking to each other, shaking hands, or signing an agreement.⁶ While better

⁶ The author has done such searches for Colombia, Kenya, Mozambique, Northern Ireland, the Philippines, South Sudan, Sri Lanka, Sudan, Syria, and Yemen.

knowledge and information may be seen as important for producing such outcomes, their sources are not typically associated with machines, but with humans. Technology tends to be relegated to the margins, such as for the photo op. However, such positions build on an exaggerated demarcation between humans and machines, sustained by an anthropocentric view of peacemaking that is more an expression of professional cultures than an assessment of the behind-the-scenes, actual peacemaking practice.

New Knowledge Frontiers

The anthropocentric view of peacemaking is outdated. The increasing availability of digital technologies has not only changed how conflict parties wage war, but also how they engage and relate to each other in peace processes, creating new challenges for mediators to perceive conflict dynamics. Social media, for instance, is not only used by protest movements, but also by conflict parties and stakeholders in support of their interests and agendas, including to influence public opinion, for political mobilization, and to gain tactical advantage on the battlefield (Rohwerder 2015). Peacemaking efforts also increasingly produce large amounts of data, for instance, when mediation teams employ interactive websites, online surveys or messaging applications to enable participation (Hirblinger 2020). As a result, mediators usually find themselves in an 'ocean of information' (Varela 2021).

Mediators are commonly information brokers, building trust and reducing uncertainty through the provision of information to conflict parties (Kydd 2003; Savun 2008). Yet, this intermediary role of mediators is increasingly undermined by the digitization of armed conflicts and peace processes. The vast amounts of data generated by conflict parties and stakeholders cannot be analyzed by human hands and eyes – partly because of its size, and partly because it is generated on social media platforms that influence how data can be retrieved, and to what ends (Batinca & Treleaven 2015). This makes it more difficult for mediators to stay on top of all information flows. Rather than enjoying a privileged position, today's mediators are embedded in complex human-machine networks, through which information about a given conflict is generated and exchanged. But when mediators depend on machine-generated and processed data, they share influence over the information environment with many other humans and machines.

The digitization of peace processes therefore leads to a new knowledge frontier that creates new epistemological uncertainties for mediators. Paradoxically, at first, the spread of information technologies creates the perception that mediators can be closer to what happens 'on the ground.' However, the vast amount of information available to mediators actually has a distancing

effect – it becomes increasingly difficult to obtain an overall, objective picture and to understand patterns and trends. To stay connected to the peace processes, as the proponents of the adaptation of AI for peacemaking have argued, it therefore appears necessary to integrate intelligent machines into the data analysis effort, particularly to support the analysis of large amounts of data (Höne 2019).

The Machine Hubris: Technocentric Peacemaking

On the other side of the spectrum, some observers tend to encourage the idea that innovations in AI could lead to a future in which robots not only wage war against each other, but also make peace (Lamb 2017). If conflicts are increasingly automated, why should peacemaking not follow suit? If robots wage war on social media, why should they not also resolve conflicts?⁷ Such thought experiments may be inspired by the call for a global data-driven and ML-powered system that could make predicting wars as straightforward as forecasting the weather (Guo, Gleditsch & Wilson 2018), or the idea that entire peace agreements could be written with the support of ML tools (Rutherford 2021), just as they now successfully generate press releases, novels and poetry (Lombardo 2019). This would require forms of broad AI that would be equivalent, or even superior, to human intelligence and capable of replacing human mediators – a narrative that is ironically reproduced by Google Search's recommender algorithm.⁸ Such notions rightly trigger concerns about how AI systems may undermine fundamental human rights, for instance, related to political participation, nondiscrimination and equality, by producing knowledge and influencing human behavior in top-down, extractive and opaque manners, leaving little opportunity for human oversight and control (Helbing et al. 2019; Livingston & Risse 2019).

Context-dependent Peacemaking

However, it is important to delineate the limits of what intelligent machines are able to achieve in an autonomous fashion. AI systems tend to learn with varying degrees of autonomy from humans, ranging from full human control exercised in expert systems, to supervised and unsupervised machine learning systems. They also tend to operate with varying degrees of dependence on

⁷ Participant's remarks during a workshop on social media in peace mediation, held in Geneva, June 2018.

⁸ In February 2021, the author conducted a Google search with 'will AI'. The recommender system suggested to complete the search *inter alia* with 'take over the world,' 'replace humans,' 'take my job,' and several variants.

humans. However, as Bradshaw and colleagues have argued, it is important to note that AI systems never have full, “unbounded” autonomy, unless “the task and context are sufficiently constrained.” As they suggest, we should think of degrees of autonomy that result from the machine’s integration into a larger work system composed of interdependent agents, in which humans always maintain a degree of supervision (Bradshaw et al. 2013). Notably, the degree to which agents, whether machine or human, can successfully operate autonomously is context-specific (Ibid.). As it happens, peacemaking strategies and dynamics differ widely across armed conflict contexts, and evolve historically (Rocha 2019). They may also shift over time within a single mediation as the context evolves (Hellmüller 2021). Yet, when contexts change, AI systems will lose their autonomy or fail to perform their tasks, for instance, because they rely on algorithms that have been built with training data that is no longer appropriate.

Moreover, peacemaking contexts not only vary across cases, but are also difficult to ascertain, posing further challenges for the effective use of AI in mediation. As Eronen (2016) put it, peacemaking is commonly confronted with “wicked problems,” that is, problems in which there is little agreement among the conflict parties, not only about how to resolve the conflict, but about what caused the conflict in the first place. Yet, without a clear problem definition, machines (and similarly, humans) struggle to develop algorithms that can generate outputs based on selected input data. Employing logic, “requires knowledge of the world that is *certain*” (Russell et al. 2016: 35), a condition that is difficult to meet in the context of armed conflicts and peace processes.

The context-specific, shifting and wicked nature of armed conflicts makes the achievement of a peace agreement a considerably more difficult task than stopping a car at a red light. It requires a ‘broad’ or ‘general’ intelligence that can reason in context (Shevlin et al. 2019). AI systems operate through models, which formalize problems and objectives, to generate an output based on input data. However, where problems and objectives are hard to define, and input data changes across contexts, they may produce results that are either poor, inappropriate, or both.⁹ This means that AI systems employed in peacemaking will remain heavily dependent on human agents, who narrowly define the context in which they operate, and the problems that they should solve. It is not surprising that the AI applications most prominently experimented with in support of peace mediation focus on relatively narrow knowledge building tasks, carried out in clearly defined contexts in which inputs and outputs can be easily stated, such as to extract and analyze the opinions, positions and

9 I would like to thank the anonymous reviewer for suggesting I clarify this point.

sentiments expressed by conflict parties and stakeholders in large amounts of text (Kolovou 2019). For example, the United Nations, in collaboration with academic and private sector partners, have experimented with systems that analyze a large number of responses from mass online focus groups to map public opinion (ITU 2019; Warrell 2020), or that help analyze large amounts of notes from dialogue efforts according to topics and conflict party positions (Arana-Catania, Van Lier & Procter 2021).

Paradigmatic Peacemaking Knowledge

The successful development and deployment of AI systems for broader tasks is dependent on the availability of sufficient context-specific information across the range of armed conflicts and peace processes. However, suitable expert knowledge and data are not easy to come by. To start with, there is limited formalized knowledge about peace processes and peace mediation that could be fed into expert knowledge-based systems. Mediators are commonly bound by confidentiality clauses and treat their expertise like guild knowledge; their power stems just as much from what they know as from what they choose not to disclose. Consequently, it proves difficult to build peacemaking algorithms based on practitioners' knowledge.

Scientific knowledge could be a viable alternative, but peace research is unlikely to fill this gap. In contrast to other knowledge domains such as law, peacemaking is characterized by little formalized, definitive knowledge. Scientific theories about what causes armed conflict and what ends them, which aim to make general claims about the correlation between various factors, tend to be not only contested (Call 2012), but also influenced by scientific paradigms that determine ontological perspectives and epistemological approaches (Kivimaki 2016). At the same time, the nature of armed conflict is continuously evolving and dynamic (von Einsiedel, et al. 2017). If algorithms were to be based on established theories, they would produce results that correspond to narrow scholarly views that are rather ill-suited for making sense of the context-dependent realities of conflict.

For the broader AI tasks relevant for peacemaking, such as predicting conflict onset and dynamics, as well as factors that stand in the way of conflict resolution, algorithmic models must be built from a more heterogeneous array of data, so that machines can generate inferences based on patterns between various data types – such as public sentiments, levels of violence, and peace process milestones. For instance, machines tasked to assess the potential effects of proposed agreement provisions on conflict dynamics or on the durability of a peace agreement, would require historical data on those two variables. Much relevant training data can be generated from already structured datasets, such

as datasets about conflict events and battle-related deaths (Eck 2012), conflict management events, third-party interventions and preventative actions (Wallensteen & Svensson 2014), and the content and implementation of peace agreements (Bell & Badanjak 2019). In addition, there exist many other data sources about possible factors that influence conflict occurrence, such as levels of youth unemployment, elite rivalry, or government corruption (Caplan & Hoeffler 2017).

However, a central challenge with such historical data is that, by definition, it only contains information about past efforts, thus making it ill-suited for generating innovative solutions to the changing nature of armed conflict; such historical data limits all possible inferences to the universe of past cases. For instance, a machine could generate correlations between specific power-sharing provisions and long-term reduction in armed violence, and may be able to make recommendations for the power-sharing provision best suited for a given context. However, the algorithm will be unable to 'think' *out of the box*, beyond the repertoire of options contained in the historical training data. Similar to symbolic AI systems, connectionist systems based on supervised learning are also shaped by scientific trends, hence reflecting what researchers, and to some degree decision makers and bureaucrats, think matters in peacemaking at a given point in time.

For instance, in the wake of United Nations Security Council Resolution 1325, quantitative data about gender equality has increasingly been collected, but before the resolution was adopted in 2000, such data were rare (UN Women 2014). If a machine was to determine the effects of armed conflicts on women's well-being in 1988, it likely would have been less able to answer the question than today, despite the fact that gender mattered in armed conflict as much then as it does now. Training data is therefore not only 'structured' in the sense that it has been coded, but also that it has been selectively produced. The capabilities of ML-built AI is thus largely dependent on the intellectual and academic trends in peace mediation and the stances of those with authority over the processes of knowledge production.

A stronger independence from scientific trends could be achieved by unsupervised ML-systems that inductively establish correlations between different events through the reading of unstructured information. For instance, self-learning methods running on neural networks neither require formalized theories, nor structured training data (Cederman & Weidmann 2017: 475). However, when employed for the task of civil war prediction, unsupervised prediction models have an accuracy of less than 50%, which drastically reduces their relevance for practice. They also struggle to identify low-probability events (armed conflicts are not likely), define the unit of analysis (intra-state conflicts often

have transnational dimensions), and account for larger changes in the system (armed conflicts are influenced by geopolitics). Given the inherent complexity of armed conflicts, these models are thus more suitable for “pattern prediction” that documents non-linear relationships in the data “rather than the precise empirical forecasting of events” (Cederman & Weidmann 2017: 475). Yet, these patterns must be interpreted by human agents to provide a suitable basis for decision making and action. In other words, unsupervised learning systems may be able to overcome the challenges of paradigmatic peacebuilding, because they do not rely on expert knowledge. Yet, their effectiveness is strongly dependent on human interpretation.

Subjective Dimensions of Peacemaking

Mediators may at times use scientific methods, but they are not scientists. And while peacemaking requires addressing tangible problems, it often does so by working with the views, narratives and opinions that underpin the conflict parties' positions, in an effort to reframe and change them (Mason & Cross 2007; Harper 2006), and to create a new ontological basis for a political settlement. Therefore, when building knowledge about the conflict and options for its resolution, mediators will not be satisfied with the objective picture that emerges from data about tangible 'hard' factors and events, but will have to be concerned with the subjective viewpoints of the conflict parties.

Let us assume we want a machine to learn the relationship between two single events, such as the assassination of an incumbent head of state and the outbreak of violence. Learning across cases, we will likely see that there exists no law-like relationship between these two events, but that other factors play a role, including 'hard' objective factors, such as the political system, but also a large number of 'soft' subjective factors that are hermeneutically related. For instance, is the information about the assassination shared and how is it made public? What narratives and perceptions emerge in relation to the news of the event? Making sense of the relevance of such seemingly objective events is particularly difficult because the ontological landscapes of peacemaking contexts are not stable, but characterized by what has been referred to as 'ontological insecurity,' as views of the parties to the conflict and the parties themselves change during the negotiation process (Rumelili & Çelik 2017). In other words, the dynamics of peace processes are determined not only by a set of factors that can be objectively established, but also by how these factors are viewed by those who are involved in the conflict.

However, efforts to machine-read subjective judgments and perceptions and assess how they matter in conflict face considerable challenges. While political leaders may provide narratives to mobilize their supporters, their

personal motives for entering into conflict or agreement commonly remain hidden behind the multiple online ‘transcripts’ (Soriano & Sreekumar 2012). As Nathan argues, “peacemakers cannot simply infer a party’s intentions from its public pronouncements or objective factors” (Nathan 2014: 225). Much vital information exchange between conflict parties and mediators takes place informally and off the record, such as secretive conversations between ruling elites. This means that many relevant aspects of a conflict that need to be considered for understanding its causes and developing options for conflict resolution remain opaque to machines, unless mediators, conflict parties and stakeholders volunteer to provide information about them, which mediators can interpret in context.

From Mediator-Machine Impasse to Mediator-Machine Hermeneutics

The above discussion demonstrates that increasingly digitized peace processes pose challenges to both humans and machines that aim to produce knowledge in support of a settlement of conflict – such as about the causes and dynamics of conflict, about future events, and about options for conflict resolution. Knowledge is defined as true, justified beliefs about the world, which in the context of AI can be understood as “validated information about relations between entities in certain contexts” (von Rueden et al. 2020). When it comes to the general task of making sense of armed conflicts and resolving them, humans and machines grapple with determining what the relevant entities of knowledge are, agreeing on the relations between them, and observing both entities and relations. Therefore, it is necessary to reconsider how knowledge relevant for peacemaking efforts can be generated. I argue that such an undertaking entails at least three principal steps and point to the incipient efforts to implement these in theory and in practice.

Beyond Humans vs. Machines: The Emergence of Hybrid Peacemaking Intelligence

First, we must move away from conceptions of human and machine intelligence as separate, unrelated systems of knowledge production and use, and towards a notion of hybrid peacemaking intelligence that captures how humans and intelligent machines increasingly operate interdependently when generating peacemaking-relevant knowledge. This does not mean that intelligent machines have intentionality, but it requires acknowledging that they have agency, in that they have the “capacity to perform activities in a particular environment in line with a set of goals/objectives that influence and shape

the extent and nature of their participation,” as Engen, Pickering and Walland (2016) have put it. While it goes without saying that a machine’s goals, objectives and pathways to achieve them have been set by its human designers, the machine asserts agency vis-à-vis other humans within the human-machine network in which it is employed. And as I will demonstrate, machines do not only become increasingly necessary for the execution of some knowledge-related tasks, but they also influence human reasoning and decision-making processes that are at the core of peacemaking.

The discussion above should have made it clear that AI is not created, and does not exist, in separation from human intelligence. While AI is commonly represented in contrast to human intelligence, the two are in fact closely intertwined. It goes without saying that the models and algorithms that are at the core of intelligent machine operations are built by humans, controlled by humans and will succeed or fail to generate relevant outputs because of human skills and craft. Among AI developers, this has led to the insight that “human judgment plays a role throughout a seemingly objective system of logical decisions” (IBM 2014).

However, human mediators will likewise become more reliant on the collaboration with intelligent machines. The integration of AI into everyday technologies, including tools used by mediators, such as online maps and recommender systems, has given rise to the view that “human intelligence and AI are complementary, symbiotic, inseparable” (Kibby 2020). International organizations such as the UN have taken note that intelligent machines are used in close collaboration with human personnel across a range of tasks relevant for peacemaking, including to analyze social media and radio content for ‘fake news’ and rumors (UN Innovation Cell 2018), or to predict electoral violence (OICT 2018). Importantly, intelligent machines increasingly contribute to the execution of tasks that could not be achieved without them. For instance, to deal with the large amounts of data produced on social media or through digital inclusion efforts, human mediators form interdependent relationships with AI systems that help, among other things, with the retrieval, analysis, storage, or visualization of information. In such cases, mediation efforts rely on forms of intelligence that can be called ‘hybrid,’ distinguished by the fact that neither human intelligence nor machine intelligence could solve the task at hand independently. The collaboration between human and machine intelligence therefore becomes a necessity (Akata et al. 2020).

This does not mean that machines and humans stand on an equal footing. However, it does recognize that the augmentation of human intelligence with machine intelligence will become increasingly inevitable for knowledge production in peace mediation contexts that are influenced by digitalization. The growing dependence of mediators on intelligent machines in hybrid systems

also means that both humans and machines assert an influence over the final product, that is, the knowledge that results from such collaborations. Indeed, the examples above suggest that human intelligence is not only increasingly entwined with machine intelligence in conflict prevention and resolution efforts, but that machines influence how humans make sense of the world, for instance, by focusing the analysis on phenomena that can be read by ML tools, such as sentiments.

We should therefore study how humans and machines produce knowledge for peace processes in an interdependent manner. Proponents of Science and Technologies Studies (STS) have for some time developed research perspectives that engage with the integration of human and non-human agencies in 'socio-technical systems,' starting with the basic premise that objects have agency, too – relationally, with humans (Hoijsink & Leese 2019). While this insight originally pertained to non-intelligent objects, such as bridges or water pumps, STS scholars have more recently focused on AI-powered machines, such as autonomous weapons systems (Leese 2019). Relational human-machine agency can also be observed in knowledge generating systems, such as those used in health care, where it has been observed that intelligent machines increasingly influence human thought processes and decision making (Engen, Pickering & Walland 2016). In peacemaking, however, human-machine interactions in such hybrid systems have yet to be studied. Shedding light on the emergence of mutual dependencies between mediators and machines should be at the core of the research agenda on hybrid peacemaking intelligence.

In practice, initial efforts to develop and pilot test hybrid arrangements are already under way and demonstrate the intricate relationships between humans and machines in the production of peacemaking-relevant knowledge. For example, the UN Innovation Cell in collaboration with the UN Office of the Special Envoy of the Secretary General for Yemen (OSESGY), has tested AI for the automated analysis of content generated by large-scale online focus groups, to enable a real-time dialogue between the mediator and several hundred participants that could inform the Envoy's mediation approach on topics such as a nationwide ceasefire or the humanitarian situation (Warrell 2020). To analyze public opinion on such topics, the mediation team curated a set of questions asked to the online focus group. AI-powered text-mining tools were then used to analyze and structure the large number of responses, and identify the views of the population (OSESGY 2020; ITU 2019; Warrell 2020).

Documentation of a prototype for such 'large scale synchronous dialogue processes' suggests a process in which mediators, participants and intelligent machines all contribute by generating and analyzing information: the dialogue moderators pre-select the topics of discussion, participants provide responses

and evaluate responses of others through a structured voting exercise, and algorithms are employed for a variety of tasks, including the processing of natural language and the computation of statistically reliable results based on a limited amount of input data. At the end of the cycle, these results are reviewed by the dialogue moderators to inform future dialogue cycles. The close intertwining of human and machine intelligence in such circular processes, involving the construction of queries and the collection, analysis, evaluation and review of data, enables an approach that promises to process a large amount of information in a time that is considerably shorter than that of conventional inclusion methods (Bilich et al. 2019).¹⁰

A further example of hybrid intelligence in support of peace mediation are efforts to employ text-mining methods to facilitate the OSESGY's information extraction, knowledge management and analysis of conflict party positions (Arana-Catania, Van Lier & Procter 2021). In this case, a team of data scientists and legal anthropologists, in collaboration with the mediation team, employed several topic-modelling techniques for large quantities of notes taken during the mediation's meetings with the conflict parties. The documentation of this effort reveals that the outputs – that is, visualizations of issues and conflict party positions on these issues – were conditioned by a great number of interdependent information processing steps, involving both humans and intelligent machines. For instance, the original notes used for the exercises were manually produced by human notetakers, the mediation team identified the main topics across which parts of the topic modelling should be executed, the data scientists chose existing off-the-shelf ML methods, set the parameters for statistical analysis and experimented with different types of queries, and the project team jointly interpreted the machine-generated outputs. The outcomes of this exercise are a product of hybrid intelligence because they could not have been generated by the mediation team alone, and they were co-produced through an interdependent and re-iterative sequence of human and machine actions. In this case, the agency of ML systems lies in offering a set of off-the-shelf tools that influence how the text data is analyzed, such as the BERT Natural Language Processing tool, through which the distances between conflict party positions are assessed. This tool does not have an intentional impact on the mediation, but it has agency, in the sense of carrying out the data analysis in a specific way and in line with a set of goals, in this case, privileging a particular linguistic understanding of what a conflict party position is.

While intelligent machines play an indispensable role in such efforts, human capacities to evaluate data outputs and build models that can fit the mediation

10 I would like to thank Felix Kufus for pointing me to the documentation of this prototype.

contexts will remain indispensable for such systems to work effectively. For example, text-mining methods may classify opinions with up to 90% accuracy, but they have greater difficulty analyzing their semantic content beyond a pre-defined typology of sentiments or emotional polarity (Hemmatian & Sohrabi 2019). This is even more the case with semantically more ambitious applications, such as topic-modelling, which depend on human interpretation and judgment, for instance, to decide on the number of topics that the machine should detect, provide seed words, or set statistical parameters for the text analysis (Curiskis et al. 2020; Zhao et al. 2015; Arana-Catania, Van Lier & Procter 2021). Text-mining methods also struggle with analyzing dialects, non-standard expressions, and semantically complex text. In the case of the Yemen focus groups, a considerable challenge was to develop and refine natural language processing (NLP) modules that could adequately capture the Arabic dialects spoken in the region. Therefore, human analysts will continue to play a key role in verifying machine-generated outputs and improving the detection algorithm, for instance, by continuously adding training data. While the major role of intelligent machines is to automate the detection of entities and the statistical computation of results, and thus enabling to scale the analysis effort, both human and machine intelligence are needed to generate the analytical output.

Overall, mediators and mediation support actors will retain considerable agency in hybrid intelligence peacemaking systems, comparable to other fields of use, where data inputs are generated from human information and data outputs are meant to inform human decision making and action. Because data is provided and consumed by humans, human experts will continue to act as 'middlemen' between the human data sources, human knowledge end users, and the machines that help generate knowledge by way of collecting and analyzing the data. For example, efforts to design 'human-machine' partnerships have for some time been advanced in the medical field, with the aim to identify and classify diseases more accurately (Patel et al. 2019). Comparable integrated systems that keep the 'human-in-the loop' (Wang 2019) are likely to also be the standard in peace mediation, not only to produce viable results, but also to guarantee ethical standards that are in line with human rights (Jobin, Ienca & Vayena 2019).

However, in contexts where large amounts of data must be processed at ever increasing speeds and quantity, while mediation teams act in both inclusive and nimble manners, the close intertwining of human and machine intelligence will become inevitable. Future research on these dynamics should therefore shed light on the organizational arrangements and the everyday practices that shape how knowledge is generated in such hybrid systems. Such

efforts may benefit from a dedicated field perspective that can draw on STS methods, such as following relational networks through which data is translated into authoritative and actionable knowledge (Jasanoff 2017), studying the dynamics of co-production through which machines become normalized and accepted as peers to human analysts (Jacobsen & Monsees 2019), or digital ethnography approaches documenting how AI research labs become sites of peacemaking (Pink et al. 2016; Hjorth et al. 2017). Hybrid systems will also be shaped by efforts to regulate the use of AI in peacemaking through mediation-specific guidelines and standards, developed by international organizations such as the UN. Our research should therefore also engage with the implications of these normative frameworks on the design and functioning of hybrid systems, particularly when it comes to human oversight and control. Yet, we may also want to reflect on the types of knowledge that hybrid systems produce, and the power dynamics by which these systems are characterized – two aspects that I turn to in the last sections of this article.

Beyond Neutrality and Bias: Hybrid Intelligence and the Hermeneutics of Peacemaking

Once we acknowledge the increasing relevance of hybrid intelligence systems for peacemaking, we may want to explore further what type of knowledge they can and should produce in support of peacemaking. To this end, I suggest studying how AI can be employed in ways that encourage reflexive knowledge supportive of dialogue efforts. As argued above, peace mediation requires engaging ‘objectively’ with ‘hard’ factors – but also dealing with the narratives, perceptions, and positions of the conflict parties. Therefore, a pivotal question pertains to how hybrid peacemaking approaches could go beyond the current concerns with algorithmic bias, because these inevitably look at intelligence and its outputs from the vantage point that the creation of objective knowledge is desired and possible. In practice, however, mediators remain confronted with the conflict parties’ partial knowledge and subjective views, and are commonly vulnerable to allegations of partiality. We may therefore want to inquire how to move beyond the objectivity-bias dichotomy by developing reflexive approaches to knowledge production that allow conflict parties and stakeholders to engage dialogically with the functions and data outputs of hybrid systems.

As Nathan (2014: 209) argues, mediators require “deep understanding of the disputant parties’ internal calculations about the conflict and its resolution.” A skill commonly necessary for such a deep understanding is to navigate between the diverging truth claims about the conflict. To exemplify, senior mediators stress the need to “be aware of the different explanations for why

the conflict erupted in the first place” (Brahimi & Ahmed 2008: 5). However, such interpretations are likely to shift in the course of the mediation process. It is also widely accepted that mediators commonly aim to seek convergence between the parties’ positions. In bargaining theoretical models, the parties are viewed as rational actors with pre-defined interests, but such interests are very often fluid and changing, rather than static. Outside the narrow confines of peace mediation research, numerous scholars point to the fact that how conflict parties think, talk or write about the conflict affects their behavior in peace processes (Ramsbotham, Miall & Woodhouse 2011). Yet, armed conflicts are also commonly characterized by challenges to knowing anything with certainty, which results in difficulties agreeing on a common understanding of the elements that constitute a conflict and how they relate to each other. Seeking a common interpretative framework and working towards a mutual understanding of the rationalizations that underpin the various positions is therefore an important mediation task.

This begs the question of how mediators usually position themselves vis-à-vis the narratives, opinions or arguments that underpin the conflict parties’ rationalizations. Ample mediation research has focused on the mediator’s neutrality or bias towards the conflict parties’ positions and views on the conflict. This pertains to the vested interests of mediators, for instance, their allegiance to influential foreign powers, but also to their social and cultural proximity to the conflict parties and ability to understand the parties’ viewpoints, as is the case with ‘insider mediators’ (Svensson & Lindgren 2013). While some scholars have stressed that mediators should be ‘credible’ and ‘truth seeking’ (Kydd 2003), others have pointed to the necessity of mediators reflexively engaging with various truth claims and interpretations of conflict, for instance, by displaying a ‘sense of humour’ (Bercovitch 1984, cited in Bercovitch, Anagnoson & Wille 1991: 15), and fostering the conflict parties’ imagination to envision alternatives (Lederach 2005). This suggests that one important mediation approach may be to take a detached position that can reflexively engage with and evaluate the various truth claims.

However, algorithms are not commonly viewed as being strong proponents of reflexivity. ML-tools for instance tend to be commonly scrutinized for potential biases, resulting, for instance, from ‘skewed’ data used to train the AI algorithm (Mehrabi et al. 2019). Nonetheless, intelligent machines can also support reflexive interpretation, for instance, when they assess the biases of other systems, such as those resulting from the training data set (Barton 2019). Moreover, AI systems can guide or correct human behavior through the detection of human biases, for instance by assessing if a proposition in a text is broadly accepted as true (Recasens, Danescu-Niculescu-Mizil & Jurafsky 2013),

detecting the stances of an author towards a target (Küçük & Can 2020), or identifying sarcastic, misleading and propaganda-laden statements in media texts (Rashkin et al. 2017). This means that AI may increasingly take the role of a cognitive aid, through which mediators can reflect on their own, and the conflict parties' positions, assumptions and ways of reasoning, therefore helping with the production of reflexive knowledge.

Researchers should therefore explore how hybrid intelligence systems may enable *hermeneutical* peacemaking practices that can support conflict parties in jointly navigating their own and others' truth claims about the conflict. A hermeneutical view on conflict resolution suggests that conflicts break down because of incommensurable views and narratives on the conflict, and therefore, a lack of 'shared reality.' Consequently, peace mediation requires working towards a 'common language,' that is, shared interpretations of the world that speak of the conflict parties' success in viewing the conflict through the eyes of the other (Väyrynen 2005). Following such an approach, hybrid peacemaking systems would be less concerned with establishing an objective view on the conflict, but rather use their capability to analyze large amounts of data to document, visualize, and engage with the conflict parties' varying interpretations of the world. Such reflexive hybrid systems would not strive to establish objective knowledge, but to achieve a degree of congruence among the parties' rationalizations of the conflict that forms the basis of intersubjectively shared peacebuilding knowledge which can support dialogue processes.

Hybrid systems can enable a hermeneutical peacemaking approach in at least two ways. First, by working on the conflict parties' truth claims about the world, we can assess how statements, narratives or discourses create a specific meaningful reality for them. This question can be tackled with support of the broad spectrum of ML-driven text mining and analysis tools, including the variants of sentiment, opinion and argument analysis (Kolovou 2019). Efforts to mine the conflict parties' or stakeholders' position from data collected in the course of dialogue efforts or through online focus groups constitute a valuable first step in this direction. Contrasting these various subjective claims with each other and in relation to objective factors, such as events, could make a meaningful contribution to the development of a joint inter-subjective perspective on conflict dynamics. Such measures would also help to counter the increasing fragmentation of conflict party views that result from the algorithmic mediation of online and social media content.

Second, hybrid systems could help shed light on the methods of knowledge acquisition and the sources of knowledge that define conflict parties' horizons. Through which means have the conflict parties come to believe a specific version of reality? How are truth claims constructed and shared in digital

networks? And how, for instance, do social media discourses intersect with the views and stance of political elites? Answering such questions requires the ‘data subjects,’ that is, those stakeholders who have produced the mined data, to be involved in its interpretation themselves (Nafus 2018). Such a hermeneutical approach would mean that representatives of conflict parties or stakeholders are actively involved in the design and use of the machine component of the system, or are at least able to retrace its operations. This would guarantee that the hybrid systems work in a ‘double hermeneutical’ manner (Giddens 2013: 374), which strives not only towards a convergence of truth claims, but also a shared understanding of how such claims are methodologically generated.

Hybrid intelligence systems committed to a hermeneutical approach would operate through a reiterative process of intellectual tasks, along the spectrum of joint model development, data collection, and analysis. This could entail the development of models and training data grounded in the conflict-affected context, for instance, drawing on experiences with the development of ‘community-based indicators’ (Firchow 2018) for the measurement of peace-building outcomes. Importantly, a hermeneutical approach would require that machine outputs are not only used for the mediation’s analysis, planning and decision making, but in larger dialogue processes as well, in which the conflict stakeholders can reflexively engage with the co-produced knowledge. Arguably, this would remain a somewhat conflictual process, where no total agreement about valid information can be achieved – just as that is an unlikely goal for the settlement of conflict (Bell & Pospisil 2017). Nevertheless, the more convergence that can be achieved, the more hybrid peacemaking systems will be able to help parties move forward towards non-violent settlements that are sustained through an intersubjective understanding of the conflict.

Beyond Eminence: Investigating Power in Hybrid Peacemaking

Finally, we also need to shed light on the political dynamics of hybrid peacemaking and the power relations that characterize human-machine networks that produce peacemaking knowledge. While humans will stay ‘in the loop’ of hybrid intelligence systems, we should be concerned with who these humans are, what power they exercise, and how this affects peacemaking efforts. For instance, they may belong to an exclusive cycle of expert mediators and computer scientists, which may have negative effects on the process. Studies of the use of AI for political decision-making and public policy suggest that AI-generated outputs enjoy limited legitimacy if they are produced without the active involvement of the citizens that they are meant to govern (Starke & Lünich 2020). We also know that peace mediation produces more durable results when processes are inclusive and participatory, thus enabling

the transformation of relationships between a diverse set of antagonists (Hirblinger & Landau 2020). However, participation does not mean that the agents involved in knowledge making efforts enjoy the same degrees of power and authority. Therefore, future research should focus on the politics of knowledge production in hybrid intelligence peacemaking systems, through which those affected by conflict get involved in determining what legitimate knowledge entails.

I suggest thinking about power as a distributed effect of human-machine networks, determined by a range of humans, from the computer scientists involved in the development of AI, to mediation support actors who roll out and maintain AI systems, to mediators and their teams that utilize them, and the conflict stakeholders and broader population who engage with them during data collection and who are affected by how the data is used. The introduction of hybrid intelligence to peacemaking changes who has the authority to make knowledge claims, and to what degree such claims attain acceptance among conflict parties and stakeholders. Humans influence machine intelligence in the conceptual phase, when deciding about what the tools are supposed to do and what methods to use; in the training phase, through the coding of training data; when collaborating with it, for instance, when selecting and preparing the data that should be analyzed; and finally, when utilizing the results. Along this spectrum of activities, various humans interact with the machines – the AI designer, the AI user in a mediation team, and the conflict stakeholder invited to a data analysis workshop, for instance; they have varying powers to influence the knowledge production process. This also means that as humans design AI systems and collaborate with them, the technologies may indeed redistribute agency among those humans who engage with it – rewarding those who can create, manage and use them. More research should therefore be invested into what Rammert (2012) has referred to as distributed agency in human-machine systems.

Our research should thus be concerned with how power is distributed in hybrid systems, how it is inscribed into models and data, and how they correspond with particular peacemaking approaches. For example, what digital and computational literacy do conflict stakeholders require to meaningfully interact with AI systems and hold them accountable? Do AI systems differentiate between the gender, ethnic or religious affiliation of the user when coding social media data, for instance, based on the language the user uses, the geotag of a Tweet, or the information provided in user profiles? Such choices may determine to which group particular grievances or positions are attributed. Moreover, various kinds of AI-based analytical methods will create knowledge that will serve one type of conflict resolution effort, but not another. Is the

system concerned with analyzing sentiments to inform efforts that can help address popular grievances or security concerns, or is it focused on analyzing the rational opinions and arguments presented by political elites on social media? What new exclusions emerge from the use of such methods, as they fail to capture the particular ways in which some conflict stakeholders make sense of conflict?

The power relations in mediator-machine networks are also shaped partly by what kind of AI methods and systems are employed. For example, systems can be open source or proprietary, which enables or hinders the reproduction of analytical processes and comparison of methods by other data scientists (Sonnenburg et al. 2007). However, even when open-source systems are used, a further concern pertains to the 'black boxed' nature of many AI methods and options to enable varying levels of human oversight and scrutiny over the inference process. The transparency of AI systems is a declared objective of mediation actors such as the UN (Azoulay n.d.), but in practice it continues to constitute a considerable challenge. Non-transparent systems provide little opportunities for oversight and may result in the concentration of power in the hands of those who design the systems and approve of their functions and outputs. However, efforts to strengthen the transparency of AI systems through interpretability methods are also often not only driven by an interest in transparency *per se*, but by other ends, such as increasing public trust (Krishnan 2020) – and this affects power relations. On the other hand, interpretability is also often less of a priority than the usability and effectiveness of applications (Lo Piano 2020). This means that power is not only exercised in the design of AI systems but also in efforts to explain and interpret their results. We must therefore also be concerned with who will be able to conduct such checks, and to what ends they are conducted. Those who translate between AI and human intellects will become powerful interlocutors in efforts to create conflict resolution knowledge. While transparent methods and translation efforts will likely increase the democratization of peacemaking efforts, they will also shift knowledge hierarchies and require considerable digital literacy among the conflict stakeholders, if they are to be inclusive (Hirblinger 2020).

The increasing employment of hybrid intelligence will also lead to the opening of the 'black boxes' of mediators and conflict parties. Where mediators operate as part of hybrid systems, mediators' reasoning and decision-making processes may become more transparent and reproducible, because they will have to co-own knowledge with the AI. This will lead to increased formalization of mediator knowledge, which might not be desirable for many mediators, given that informal knowledge and information exchange remain important to

peacemaking practice. Where conflict parties and stakeholders are integrated in efforts to gather and analyze information about their views and positions, they may be hesitant to unveil how they really think about the conflict. Yet, these challenges are not qualitatively different from traditional mediation contexts. The most valuable sources for conventional data gathering and analysis for mediation purposes usually stem from within conflict parties or their immediate environment, and require the building of trusted relationships (Nathan 2014: 223). While hybrid systems call for increased collaboration and exchange among the various human parties to the system – mediators, support actors, conflict parties and stakeholders – these constituents may also hope to retain the utmost possible influence over the interpretation process. Such powerful dynamics merit further study.

Conclusion and Outlook

Conflict resolution can be thought of as a process of rationalization that produces peacemaking knowledge, in which the causes of conflict are identified and logically linked to solutions to conflict. The digitalization of armed conflicts and peace mediation contexts fundamentally changes the requirements for the production of such knowledge. The scale of data and the algorithmic mediation of information flows means that knowledge production for peace mediation will increasingly rely on hybrid intelligence systems, in which mediators utilize AI. Yet, resolving conflict will never be a matter of simply pressing a button. We must not be distracted with the unlikely scenario of an artificial superintelligence that could automate peace mediation. Rather, we should be concerned with how hybrid systems tackle much more narrow data analysis tasks that can help produce knowledge in support of conflict settlement. Normative standards and a pragmatic concern with the sustainability of peace processes further necessitate that such knowledge production is conducted in an inclusive manner, involving conflict party representatives, conflict stakeholders, insider mediators and local communities, not only in the collection of data, but increasingly, also in the development of models and data analysis.

This article suggested a research agenda on hybrid peacemaking intelligence, concerned with both the study and development of hybrid systems suitable for peace mediation. This agenda explores how knowledge is generated in hybrid systems, and particularly how mediators and their teams increasingly rely on intelligent machines for knowledge production. We should aim to better

study and describe such systems and their interactions, analyze their effects on peace mediation and process outcomes, and be concerned with efforts to regulate them through normative frameworks. Moreover, I suggest remaining skeptical about the possibility of creating objective, unbiased knowledge. Instead, we may want to inquire how hybrid systems can enhance the reflexivity of conflict parties and stakeholders through operationalizing a hermeneutical approach that helps to scrutinize AI-supported methods and data and builds an intersubjective knowledge base for dialogue. Finally, I suggest that we should be concerned with the distributed agency and power relations in hybrid peacemaking networks. While these may be participatory, inequalities and exclusions may be inscribed in AI models, methods and data, with considerable effects on the dynamics and outcomes of AI-supported peacemaking. It is likely the case that arriving at a joint understanding of what the conflict is about and how it can be resolved will remain the exception. Peace mediation is a messy and contested process, and so will be efforts to produce meaningful and actionable peacemaking knowledge with the help of AI. Despite and because of this, we may want to explore how hybrid peacemaking systems can work towards achieving a partial convergence between conflict parties and stakeholders – or study why they fail to do so.

References

- Akata, Zeynep, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, et al. (2020). "A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence," *Computer* 53 (8): 18–28.
- Arana-Catania, M., F. A. Van Lier and Rob Procter (2021). "Machine Learning for Mediation in Armed Conflicts." *ArXiv:2108.11942* [Cs], August. At: <http://arxiv.org/abs/2108.11942>.
- Azoulay, Audrey (n.d.) "Towards an Ethics of Artificial Intelligence." New York: United Nations. At: <https://www.un.org/en/chronicle/article/towards-ethics-artificial-intelligence>.
- Basil, John, and Joseph Bowen (2020). "AI as a Moral Right Holder," in Markus Dirk Dubber, Frank Pasquale and Sunit Das, editors, *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press.
- Batrinca, Bogdan and Philip C. Treleaven (2015). "Social Media Analytics: A Survey of Techniques, Tools and Platforms." *AI & Society* 30, 1: 89–116.
- Bell, Christine and Sanja Badanjak (2019). "Introducing PA-X: A New Peace Agreement Database and Dataset." *Journal of Peace Research* 56, 3: 452–66.

- Bell, Christine and Jan Pospisil (2017). "Navigating Inclusion in Transitions from Conflict: The Formalised Political Unsettlement." *Journal of International Development* 29, 5: 576–93.
- Bercovitch, Jacob (1984). *Social Conflict and Third Parties: Strategies of Conflict Resolution*. Boulder, CO: Westview.
- Bercovitch, Jacob, Theodore Anagnoson and Donnette L. Wille (1991). "Some Conceptual Issues and Empirical Trends in the Study of Successful Mediation in International Relations," *Journal of Peace Research* 28, 1: 7–17.
- Bilich, Jordan, Michael Varga, Daanish Masood, and Andrew Konya (2019). "Faster Peace via Inclusivity: An Efficient Paradigm to Understand Populations in Conflict Zones." Presented at AI for Social Good workshop at NeurIPS, Vancouver, Canada. At: https://aiforsocialgood.github.io/neurips2019/accepted/track1/pdfs/105_aig_neurips2019.pdf.
- Bradshaw, Jeffrey M., Robert R. Hoffman, David D. Woods, and Matthew Johnson (2013). "The Seven Deadly Myths of 'Autonomous Systems.'" *IEEE Intelligent Systems* 28, 3: 54–61.
- Brahimi, Lakhdar and Salman Ahmed (2008). "In Pursuit of Sustainable Peace: The Seven Deadly Sins of Mediation." New York: NYU Center for International Cooperation. At: <https://ciaotest.cc.columbia.edu/wps/cic/0001438/0001438.pdf>.
- Bryson, Joanna J. and Andreas Theodorou (2019). "How Society Can Maintain Human-Centric Artificial Intelligence," in Marja Toivonen and Eveliina Saari, editors, *Human-Centered Digitalization and Services*. Translational Systems Sciences. Singapore: Springer.
- Call, Charles T. (2012). *Why Peace Fails: The Causes and Prevention of Civil War Recurrence*. Washington DC: Georgetown University Press.
- Caplan, Richard and Anke Hoeffler (2017). "Why Peace Endures: An Analysis of Post-Conflict Stabilisation." *European Journal of International Security* 2, 2: 133–52.
- Cederman, Lars-Erik and Nils B. Weidmann (2017). "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science* 355, 6324: 474–76.
- Convergne, Elodie (2016). "Learning to Mediate? The Mediation Support Unit and the Production of Expertise by the UN." *Journal of Intervention and Statebuilding* 10, 2: 181–99.
- Curiskis, Stephan A., Barry Drake, Thomas R. Osborn, and Paul J. Kennedy (2020). "An Evaluation of Document Clustering and Topic Modelling in Two Online Social Networks: Twitter and Reddit." *Information Processing & Management* 57, 2.
- Eck, Kristine (2012). "In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets." *Cooperation and Conflict* 47, 1: 124–41.
- Eleiba, Ahmed (2020). "Digital Technologies: Game Changer for Peace?" *The KOFF Peacebuilding Magazine*. swisspeace. At: <https://www.swisspeace.ch/apropos/wp-content/uploads/2020/11/november-2020-4.pdf>.

- Engen, Vegard, J. Brian Pickering and Paul Walland (2016). "Machine Agency in Human-Machine Networks: Impacts and Trust Implications," in Masaaki Kurosu, editor, *Human-Computer Interaction. Novel User Experiences*. Cham: Springer International Publishing.
- Eronen, Oskari (2016). "Organising Artisans for Peace: CMI on a Learning Curve," in Emery Brusset, Cedric De Coning, and Bryn W. Hughes, editors, *Complexity Thinking for Peacebuilding Practice and Evaluation*. London: Palgrave MacMillan.
- Firchow, Pamina (2018). *Reclaiming Everyday Peace: Local Voices in Measurement and Evaluation after War*. New York: Cambridge University Press.
- Giddens, Anthony (2013). *The Constitution of Society: Outline of the Theory of Structuration*. New York: John Wiley & Sons.
- Guo, Weisi, Kristian Gleditsch and Alan Wilson (2018). "Retool AI to Forecast and Limit Wars." *Nature* 562: 331–33.
- Harper, Christopher (2006). "Mediator as Peacemaker: The Case for Activist Transformative-Narrative Mediation." *Journal of Dispute Resolution* 2, 10.
- Helbing, Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari and Andrej Zwitter (2019). "Will Democracy Survive Big Data and Artificial Intelligence?" in Dirk Helbing, editor, *Towards Digital Enlightenment: Essays on the Dark and Light Sides of the Digital Revolution*. Cham: Springer International Publishing.
- Hellmüller, Sara (2021). "The Challenge of Forging Consent to UN Mediation in Internationalized Civil Wars: The Case of Syria." *International Negotiation* 27, 1.
- Hemmatian, Fatemeh and Mohammad Karim Sohrabi (2019). "A Survey on Classification Techniques for Opinion Mining and Sentiment Analysis." *Artificial Intelligence Review* 52, 3: 1495–1545.
- Hirblinger, Andreas T. (2020). "Digital Inclusion in Mediated Peace Processes: How Technology Can Enhance Participation." At: <https://www.usip.org/publications/2020/09/digital-inclusion-mediated-peace-processes-how-technology-can-enhance>.
- Hirblinger, Andreas T. and Dana M Landau (2020). "Daring to Differ? Strategies of Inclusion in Peacemaking." *Security Dialogue* 51, 4: 305–22.
- Hjorth, Larissa, Heather A. Horst, Anne Galloway and Genevieve Bell, editors (2017). *The Routledge Companion to Digital Ethnography*. New York: Routledge.
- Hojtink, Marijn and Matthias Leese, editors (2019). *Technology and Agency in International Relations: Emerging Technologies, Ethics and International Affairs*. London: Routledge.
- Höne, Katharina (2019). "Mediation and Artificial Intelligence: Notes on the Future of International Conflict Resolution." Geneva: DiploFoundation. At: https://www.diplomacy.edu/sites/default/files/Mediation_and_AI.pdf.

- IBM (2014). "Everyday Ethics for Artificial Intelligence." At: <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.
- ITU (2019). "United Nations Activities on Artificial Intelligence (AI)." At: https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2019-1-PDF-E.pdf.
- Jacobsen, Katia Lindskov and Linda Monsees (2019). "Co-Production: The Study of Productive Processes at the Level of Materiality and Discourse," in Marijn Hoijsink and Matthias Leese, editors, *Technology and Agency in International Relations*. London: Routledge.
- Jasanoff, Sheila, editor (2004). "The Idiom of Co-Production," in *States of Knowledge: The Co-Production of Science and Social Order*, International Library of Sociology. London: Routledge.
- Jasanoff, Sheila (2017). "Virtual, Visible, and Actionable: Data Assemblages and the Sightlines of Justice." *Big Data & Society* 4, 2.
- Jenny, Joëlle, Rosi Greenberg, Vincent Lowney and Guy Banim (2018). "Peacemaking and New Technologies. Dilemmas & Options for Mediators." Mediation Practice Series. Center for Humanitarian Dialogue. At: <https://www.hdcentre.org/wp-content/uploads/2018/12/MPS-8-Peacemaking-and-New-Technologies.pdf>.
- Jobin, Anna, Marcello Ienca and Effy Vayena (2019). "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, 9: 389–99.
- Kibby, Matthew (2020). "The Impact of AI on Human Intelligence: Can We Live Without It?" At: <https://www.ameinfo.com/industry/technology/the-impact-of-ai-on-human-intelligence>.
- Kivimaki, Timo (2016). *Paradigms Of Peace: A Pragmatist Introduction to the Contribution to Peace Of Paradigms Of Social Science*. London: Imperial College Press.
- Knox, Hannah and Dawn Nafus, editors (2018). *Ethnography for a Data-Saturated World: Materialising the Digital*. Manchester: Manchester University Press.
- Kolovou, Athanasia (2019). "Machine Learning Methods for Opinion Mining in Text: The Past and the Future," in George A. Tsihrintzis, Maria Virvou, Evangelos Sakkopoulos, and Lakhmi C. Jain, editors, *Machine Learning Paradigms: Learning and Analytics in Intelligent Systems*. Cham: Springer International Publishing.
- Krishnan, Maya (2020). "Against Interpretability: A Critical Examination of the Interpretability Problem in Machine Learning." *Philosophy & Technology* 33, 3: 487–502.
- Küçük, Dilek and Fazli Can (2020). "Stance Detection: A Survey." *ACM Computing Surveys* 53, 1: 12:1–12:37.
- Kydd, Andrew (2003). "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47, 4: 597–611.
- Lamb, Harriet (2017). "Can Robots Do a Better Job Of Building Peace?" *Huffington Post UK*. 20 September. At: https://www.huffingtonpost.co.uk/harriet-lamb/robots-bring-us-peace_b_18031158.html.

- Lederach, John Paul (2005). *The Moral Imagination: The Art and Soul of Building Peace*. New York: Oxford University Press.
- Leese, Matthias (2019). "Configuring Warfare: Automation, Control, Agency," in Marijn Hoijtink and Matthias Leese, editors, *Technology and Agency in International Relations*. London: Routledge.
- Livingston, Steven and Mathias Risse (2019). "The Future Impact of Artificial Intelligence on Humans and Human Rights." *Ethics & International Affairs* 33, 2: 141–58.
- Lo Piano, Samuele (2020). "Ethical Principles in Machine Learning and Artificial Intelligence: Cases from the Field and Possible Ways Forward." *Humanities and Social Sciences Communications* 7, 1: 1–7.
- Lombardo, Giorgia (2019). "What Happens When AI Writes a Novel?" Medium. 3 December. At: <https://medium.com/demagsign/what-happens-when-ai-writes-a-novel-173d407af567>.
- Mason, Simon A. and Ryan Cross (2007). "Mediation and Facilitation in Peace Processes." Zurich: Swiss Federal Institute of Technology. At: <https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Mediation-and-Facilitation.pdf>.
- Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman and Aram Galstyan (2019). "A Survey on Bias and Fairness in Machine Learning." *ArXiv:1908.09635 [Cs]*, September. <http://arxiv.org/abs/1908.09635>.
- Nafus, Dawn (2018). "Working Ethnographically with Sensor Data," in Hannah Knox and Dawn Nafus, editors, *Ethnography for a Data-Saturated World*. Manchester: Manchester University Press.
- Nathan, Laurie (2014). "The Intelligence Requirement of International Mediation." *Intelligence and National Security* 29, 2: 208–26.
- OICT (2018). "What Is Artificial Intelligence (AI)?" New York: United Nations. At: <https://unite.un.org/sites/unite.un.org/files/emerging-tech-series-ai.pdf>.
- OSESGY (2020). "Cutting-Edge Tech in the Service of Inclusive Peace in Yemen." At: <https://osesgy.unmissions.org/cutting-edge-tech-service-inclusive-peace-yemen>.
- Patel, Bhavik N., Louis Rosenberg, Gregg Willcox, David Baltaxe, Mimi Lyons, Jeremy Irvin, Pranav Rajpurkar, et al. (2019). "Human – Machine Partnership with Artificial Intelligence for Chest Radiograph Diagnosis." *Npj Digital Medicine* 2, 1: 1–10.
- Pink, Sarah, Heather A. Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi, editors (2016). *Digital Ethnography: Principles and Practice*. Los Angeles: Sage.
- Rammert, Werner (2012). "Distributed Agency and Advanced Technology. Or: How to Analyze Constellations of Collective Inter-Agency." in Jan-Hendrik Passoth, Birgit Peuker, and Michael Schillmeier, editors, *Agency Without Actors? New Approaches to Collective Action* London: Routledge.

- Ramsbotham, Oliver, Hugh Miall and Tom Woodhouse (2011). *Contemporary Conflict Resolution*. Cambridge, UK: Polity.
- Rashkin, Hannah, Eunso Choi, Jin Yea Jang, Svitlana Volkova and Yejin Choi (2017). "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics.
- Rauchhaus, Robert W. (2006). "Asymmetric Information, Mediation, and Conflict Management." *World Politics* 58, 2: 207–41.
- Recasens, Marta, Cristian Danescu-Niculescu-Mizil and Dan Jurafsky (2013). "Linguistic Models for Analyzing and Detecting Biased Language." *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics. At: <https://www.aclweb.org/anthology/P13-1162>.
- Rocha, José Pascal da (2019). "The Changing Nature of International Mediation." *Global Policy* 10, S2: 101–7.
- Rohwerder, Brigitte (2015). "Social Media and Conflict Management in Post-Conflict and Fragile Contexts." *Helpdesk Research Report* 1184. GSDRC Applied Knowledge Services. At: https://www.academia.edu/10546379/Social_media_and_conflict_management_in_post-conflict_and_fragile_contexts.
- Rumelili, Bahar and Ayşe Betül Çelik (2017). "Ontological Insecurity in Asymmetric Conflicts: Reflections on Agonistic Peace in Turkey's Kurdish Issue." *Security Dialogue* 48, 4: 279–96.
- Russell, Stuart J., Peter Norvig, Ernest Davis, and Douglas Edwards (2016). *Artificial Intelligence: A Modern Approach*. Third edition. London: Pearson.
- Rutherford, Alex (2021). "What Machines Can Teach Us About How to Write Constitutions." *Medium*. 18 January. At: <https://medium.com/futuring-peace/what-machines-can-teach-us-about-how-to-write-constitutions-3d675e33533b>.
- Savun, Burcu (2008). "Information, Bias, and Mediation Success." *International Studies Quarterly* 52, 1: 25–47.
- Shevlin, Henry, Karina Vold, Matthew Crosby and Marta Halina (2019). "The Limits of Machine Intelligence." *EMBO Reports* 20, 10: e49177. At: <https://doi.org/10.15252/embr.201949177>.
- Sonnenburg, Soren, Mikio L Braun, Cheng Soon Ong, Samy Bengio, Leon Bottou, Geoffrey Holmes and Yann LeCun (2007). "The Need for Open Source Software in Machine Learning." *Journal of Machine Learning Research* 8: 2443–2466.
- Soriano, Cheryll Ruth and T. T. Sreekumar (2012). "Multiple Transcripts as Political Strategy: Social Media and Conflicting Identities of the Moro Liberation Movement in the Philippines." *Media, Culture & Society* 34, 8: 1028–39.

- Standfield, Catriona (2020). "Caught between Art and Science: The Women, Peace and Security Agenda in United Nations Mediation Narratives." *International Feminist Journal of Politics* 22, 5: 629–51.
- Starke, Christopher and Marco Lünich (2020). "Artificial Intelligence for Political Decision-Making in the European Union: Effects on Citizens' Perceptions of Input, Throughput, and Output Legitimacy." *Data & Policy* 2.
- Svensson, Isak and Mathilda Lindgren (2013). "Peace from the Inside: Exploring the Role of the Insider-Partial Mediator." *International Interactions* 39, 5: 698–722.
- Turner-Lee, Nicol, Paul Resnick, and Genie Barton (2019). "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms." *Brookings* (blog). 22 May. At: <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.
- UN Innovation Cell (2018). "Experimenting with Big Data and AI to Support Peace and Security: UN Global Pulse." New York: United Nations. At: <https://www.unglobalpulse.org/document/experimenting-with-big-data-and-ai-to-support-peace-and-security/>.
- UN Women (2014). "Women, Peace and Security Statistics: Where We Stand and How to Move Forward." *Fifth Global Forum on Gender Statistics*. Aguascalientes, Mexico. At: https://unstats.un.org/unsd/gender/mexico_nov2014/Session%207%20UNW%20paper.pdf.
- Varela, Miguel (2021). "Technology and Peace Mediation: A Shift in the Mediator's Profession." in Catherine Turner, editor, *Rethinking Peace Mediation: Challenges of Contemporary Peacemaking Practice*. Bristol: Bristol University Press.
- Väyrynen, Tarja (2005). "A Shared Understanding: Gadamer and International Conflict Resolution." *Journal of Peace Research* 42, 3: 347–55.
- von Einsiedel, Sebastian, Louise Bosetti, James Cockayne, Cale Salih, and Wilfred Wan. (2017). "Civil War Trends and the Changing Nature of Armed Conflict." Occasional Paper #10. New York: United Nations University Centre for Policy Research. At: https://collections.unu.edu/eserv/UNU:6156/Civil_war_trends_UPDATED.pdf.
- von Rueden, Laura, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, et al. (2020). "Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems." *ArXiv:1903.12394* [Cs, Stat], February. At: <http://arxiv.org/abs/1903.12394>.
- Wählich, Martin (2020). "Big Data, New Technologies, and Sustainable Peace: Challenges and Opportunities for the UN." *Journal of Peacebuilding & Development* 15, 1: 122–26.
- Wallensteen, Peter and Isak Svensson (2014). "Talking Peace: International Mediation in Armed Conflicts." *Journal of Peace Research* 51, 2: 315–27.

- Wang, Ge (2019). "Humans in the Loop: The Design of Interactive AI Systems." Human-Centered Artificial Intelligence. Stanford University. At: <https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems>.
- Warrell, Helen (2020). "UN Tries Mass Polling to Build Peace Deals." *Financial Times*, 20 February. At: <https://www.ft.com/content/9dag69d8-518e-11ea-8841-482eed0038b1>.
- Zhao, Weizhong, James J. Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding and Wen Zou (2015). "A Heuristic Approach to Determine an Appropriate Number of Topics in Topic Modeling." *BMC Bioinformatics* 16, 13: S8. At: <https://doi.org/10.1186/1471-2105-16-S13-S8>.