

Human–AI Collaboration in International Mediation: Reimagining Trust, Accountability and Power in High-Intensity Conflicts

Akeem Aderayo SARAKI

Department of Social Justice and Security Studies,
Faculty of Management and Social Sciences,
Kwara State University, Malete.

P.M.B. 1530, Ilorin, Kwara State, Nigeria.

akeem.saraki@kwasu.edu.ng; kilani.jilani@gmail.com

+234806 747 6439

ORCID Id: <https://orcid.org/0000-0002-2184-6596>

Abstract

The modernization of diplomacy and peacebuilding activities through the rapid adoption of Artificial Intelligence (AI) is changing the international mediation and conflict resolution activities. Although tools based on AI like predictive analytics, automated translations, sentiment analysis, and decision-support tools have potential gains in efficiency and accuracy, their impact on the mediator and their legitimacy, trust, and accountability remain under-researched. This paper looks at the role of human-AI collaboration in redefining the mediation process of high intensity conflicts. Based on the qualitative case studies, interviews with experts, and participatory scenario-building workshops, the research examines the ethical, political and institutional implications of the AI implementation into international mediation. The results imply that the introduction of AI does not displace human mediators and transforms the authority, repatriates the power, and disorganizes the traditional concept of neutrality and consent. The article proposes a hybrid mediation system based on human judgment, openness, and participatory regulation of AI systems. The paper adds to the interdisciplinary arguments in the fields of peace studies, international relations and technology ethics by suggesting normative and operational principles of responsible AI-assisted mediation.

Keywords: International mediation, Artificial Intelligence, preventive diplomacy, trust, accountability, participatory peacebuilding, conflict resolution.

1. Introduction

Traditionally, international mediation has been based on human judgment, interpersonal credibility, and political legitimacy of third-party actors that can convince parties to conflict to discuss. Requested by states, regional agents, or international institutions, mediation in high-intensity conflicts relies on the capacity of the mediator to build trust and handle the perception of neutrality and develop a credible channel towards the de-escalation or ceasefire agreements. These are never very linear or procedural. Instead, they are played out in complicated political worlds, which are informed by fear, strategic ambiguity, asymmetric power, and competing victimhood and justice discourses (Svensson, 2020; Brummans, Higham, & Cooren, 2022; Banda, 2025). Therefore, mediation was traditionally perceived as a type of diplomacy as well as a kind of relationship and morality that is based on the human ability to process signals, discern emotions, and bargain meaning in stressful situations.

Over the past few years, though, this traditionally human-centred arena has started to add Artificial Intelligence (AI) technologies in a manner that becomes more and more consequential to the very process of peace diplomacy. Conflict prediction and forecasting, automated language translation, big data-based early warning, and decision support systems to assist mediators in interpreting large amounts of information and finding opportunities to start a negotiation are now used with AI systems (Dafoe,

2018; Fjelde and Hultman, 2021; Zhou, et. al., 2025). The increasing supply of digital traces, such as satellite data and social media, mobile data, humanitarian news, etc., has brought about new possibilities of computational methods of conflict analysis. Consequently, AI-driven applications are slowly reversing the informational premises of mediation by providing novel versions of so-called algorithmic knowledge providing speed, accuracy, and strategic benefits.

This technological revolution is usually described in a positive manner. AI is demonstrated as having the ability to enhance situational awareness, the early warning systems, and assisting mediators in predicting the tendencies of escalation before violence becomes catastrophic. When field access is limited or there are security concerns that make them unsafe to be in a field, algorithmic tools can also provide new avenues of tracking events and detecting local dynamics. The linguistic barriers that often complicate the process of negotiations can also be overcome with the usage of the AI-assisted translation systems, especially in the negotiations between the armed groups, the cross-boundary actors, or the ethnolinguistic communities with different ethnocultural backgrounds. Likewise, decision-support systems will be able to interpolate the complex streams of information and offer mediators the structured alternatives, which may enhance the effectiveness of negotiation planning and designing ceasefire monitoring systems (Dafoe, 2018; Shahzad, et. al. (2024).

Nevertheless, even with this increased interest, there is still little empirical and conceptual understanding regarding the impact of AI integration on the more fundamental social pillars of mediation, in particular, trust, accountability, and power relations. Mediation is not simply a mere information processing and optimisation task, but is perceptions of legitimacy, impartiality, and ethical credibility that influence mediation significantly. The mediators usually rely on the reputation of being neutral and having the capability of assuring confidentiality, fairness, and procedural justice. Once AI systems are part of the mediation, they do not simply add a technical capacity but can even modify the relational dynamics in which mediators establish confidence and deal with competing expectations. This is especially important in high intensity conflicts in which parties tend to distrust each other or feel that they are being manipulated, misinformed or subject to outside influence and in which there is already a lack of trust (Zhou, et. al., 2025).

With the advent of AI, then, there is an immediate question of who governs the creation and meaning of algorithmic knowledge, and how it is implemented in diplomacy. AI systems are not apolitical tools that exist beyond politics. They are influenced by data quality, training assumptions, institutional priorities and ingrained values, which may impact on what they perceive as threats, who they categorize as spoilers and the patterns they believe to be of interest. Researchers have become more concerned that algorithmic tools can replicate or reproduce the existing biases, particularly when datasets are biased, skewed, or mirror past dominant narratives (Ateeq, et. al., 2024). The implications of biases in a conflict setting can be severe, such as mischaracterising community actors, marginalising a type of violence or prioritizing state-centred views that miscount insecurities that happen every day.

The other key issue is accountability. In traditional mediation, lines of responsibility can be easily defined: mediators are making decisions, proposals, and hold themselves politically or institutionally accountable to their choices and their results. Nonetheless, AI-enhanced mediation opens the difference-responsibility scenario where mediators will use AI algorithm suggestions without comprehending the way they were formulated. This may make it difficult to justify decisions and undermine transparency in the negotiation. When conflict parties believe that decisions are made under the influence of some inaccessible system that they cannot question, trust will not enhance but on the contrary, it will fall apart. Furthermore, in settings where AI outputs are considered to be authoritative on account of their scientific form, there is a danger that mediators will give too strong emphasis on the computational evaluations at the cost of contextual information and human empathy-qualities that still cannot be ignored in delicate negotiations involving trauma, loss and moral indignation (Shahzad, et. al. (2024).

Raised questions of power are also related to the use of AI. High-intensity conflicts usually have an imbalanced access to resources, information and international recognition. The complex of AI-wave systems, particularly those developed and managed by technologically enhanced states/corporates/external agencies can create further

informational asymmetry among actors. This will potentially change the power balance at the negotiation table, and some parties will be able to use the power of algorithms, left outside the data-driven contextualization of the conflict. Consequently, AI will unwillingly promote the current hierarchies instead of inclusive peace processes. This issue is even more acute when mediation takes place in the situations when the digital surveillance, cyber operations and information warfare already form the realities of the conflict.

It is under this backdrop that this paper explores the influence of human-AI cooperation in transforming international mediation practices in high intensity conflict settings. It dwells on three questions that are linked to each other. To begin with, how does the integration of AI change the role of the mediator and the nature of negotiations, specifically, the role of the mediator as a trusted third party? Second, what does the rise of AI-led decision-making mean to the trust and responsibility of conflict parties in the situation when there is a lack of transparency of algorithms or when there is disagreement on the matter? Third, in what way can participatory and ethical constructs help in the responsible application of AI in mediation that the technological facilities reinforce and not diminish legitimacy, inclusion, and justice?

The article takes a qualitative method of approaching these questions based on an interdisciplinary approach that would cut across peace and conflict studies, international relations, and technology ethics. Taking these fields into discussion, the study is no longer confined to the simplistic accounts of technological advancement but rather it focuses on the AI as a socio-political intervention that redefines the ways the peace diplomacy is being practiced, perceived and argued. By so doing it helps to build a new set of debates on the future of mediation by proposing that the key challenge of AI-assisted diplomacy lay in the maintenance and improvement of the relational basis on which sustainable peace settlements fundamentally rest; trust, accountability, and legitimacy (Dafoe, 2018; Ateeq, et. al., 2024).

2. Literature Review

2.1 International Mediation and High-Intensity Conflict

The concept of international mediation has always been perceived as one of the most distinguished means of peaceful management and ending of armed conflicts. Fundamentally, classic mediation theory understands mediation as a voluntary third-party process into a dispute that involves the role of mediator who facilitates communication, alleviates misperceptions, offers settlement alternatives and assists the parties towards a negotiated settlement. Three connected pillars of mediation such impartiality, consent, and leverage are generally pointed out in the canonical literature to be necessary conditions of mediation effectiveness (Bercovitch & Gartner, 2006). Impartiality is the perceived impartiality and justice of the mediator, and this allows credibility and lessen distrust among the parties involved in the dispute. Consent emphasizes the fact that mediation is usually based on the readiness of the parties to communicate and welcome the intervention of the third party. Leverage

describes the ability of the mediator to influence incentives and costs by persuading, using resources, diplomatic power or coercive pressure, to push the actors to compromise (Bercovitch & Gartner, 2006).

Nevertheless, as much as these principles give a very handy framework of understanding mediation in most settings, their application in a high-intensity conflict situation is much more complicated. Violence, population displacement on a mass scale, destruction of government apparatus, and intense social polarisation are all characteristic of high-intensity conflicts. Three overlapping realities, including asymmetric power relations, fragmentation of armed actors, and uncertainty of information, often undermine the basic assumptions developed in the classic mediation theory in such settings (Zartman, 2001). Such conditions do not only limit the manoeuvre of the mediator but also make it difficult to see the channels through which negotiated settlements will be born.

Asymmetry in power relations between parties is one of the greatest challenges. In most instances of high stakes war, either opponent might have an excessively stronger military advantage, international backing, or have better possession of land and institutions. Such imbalance tends to diminish the motivation to compromise, especially when the powerful players are also sure that they can win by using the sword. In this situation, mediation can be symbolic actions instead of real negotiations because the powerful side of the conflict may consider the discussions a strategic pause, whereas the weak one may take mediation as the final move to evade destruction (Zartman, 2001). The situation is even more accentuated when the war objectives of parties are constructed in existential terms, i.e. survival, identity, sovereignty or ideological purity, since the cost of compromise is perceived to be too high.

Meanwhile, high-intensity conflicts are frequently associated with numerous armed groups as opposed to coherent and united parties. Fragmentation may manifest itself in splinter groups, changing alliances, warlordism or the presence of transnational fighters. Such multiplicity of actors makes mediation difficult since a deal made with one side will not bind the others and spoilers will derail the mediation effort by new outbreaks of violence. The mediator then is forced to tackle a moving target, namely who are the legitimate representatives, internal cohesion and what actors must be part of an agreement to be meaningful. This issue of representation is acute when armed forces do not have obvious hierarchies or when leaders cannot plausibly assure their fighters to abide by the results of the negotiations (Zartman, 2001). Under such circumstances, mediation is not only a bargaining process between the opponents, but also a conflict system management practice, as the mediators will have to deal with in-group conflicts and inter-group antipathies at the same time.

Information uncertainty is another characteristic of high-intensity conflict. The violence, propaganda, and mistrust corrupt the information streams, it is hard to create credible reports of what is happening on the battlefield, what civilians suffer, or what the opponents want. There are usually partial or disputed facts that are held by the mediators and parties may play with narratives to advance

their own benefits. Such uncertainty may sabotage negotiation as it increases suspicion, creates an element of bad faith, and weakens verification processes. In situations where there is limited or unreliable information, the mediators might be unable to suggest attainable settlement agreement since they are not aware of the actual capabilities, red lines, or internal restrictions of the parties. Under these circumstances, mediation has more to do with high-intensity conflict than merely providing a platform to talk. It is a relentless process of trying to limit the risks of escalation, minimise misperceptions and stabilise the interactions to the extent that bargaining can take place. In this sense, mediation effectiveness becomes concentrated on building trust. Trust-building is not only an idealistic dream but a practical necessity to allow parties to make political risks, consider compromises, and experiment with cooperative gestures. However, trust is incredibly weak in intense conflicts that leave violence, traumas, social hatred, and fear of betrayal. Mediators hence use the gradual steps of building confidence that includes humanitarian corridors, prisoner transfers, monitoring ceasefires and third-party guarantees. These measures are able to bring small cooperation which opens the scope of conversation overtime.

Narrative framing is closely related to the idea of trust-building. In violent situations, parties tend to create a justification of violence and delegitimise rivals and narratives. Such stories can solidify identities and turn bargaining to a moral challenge. Mediators are then responsible not only to work with interests, but also with meanings - to assist parties to move beyond absolutist frames (complete victory, or no compromise), to pragmatic frames that acknowledge that both sides are vulnerable, and that they can coexist. To this extent, mediation is more a communicative than a political process: it aims to redefine how the parties perceive conflict, how they can reach outcomes that are acceptable, and project an alternative of violence into the future. The reframing of stories by the mediator can provide spaces of compromise in situations where the material conditions are still unfavourable.

In short, classic mediation theory is an appropriate baseline, and more adaptive ones are needed in high-intensity conflicts. Mediators must play in power gaps, disjointed actors and uncertainty at high costs of building trust and reframing. These complications imply that success of mediation in high-intensity conflict lies less in a single agreement and more in creating a channel between disorder to structured bargaining a process that is influenced by strategic leverages, legitimacy and careful use of narratives (Bercovitch & Gartner, 2006; Zartman, 2001).

2.2 Artificial Intelligence in Peace and Security

The emergence of artificial intelligence (AI) has represented a radical change of modern-day peace and security practice, redefining the way threats are perceived, the way violence can be predicted, and the way a decision can be made in a wartime and peacetime environment. AI is now pervasive in military strategy, intelligence collection, and government apparatus, generating what some researchers term a migration towards the data-driven and algorithmically mediated security designs. According

to recent studies, one of the domains where AI is becoming increasingly relevant is the early warning system, surveillance in the battlefield, and strategic decision-making, implying that AI is not so much a technological instrument but a phenomenon that changes the very logic of security governance (Dafae, 2018; Scharre, 2023).

Early warning systems based on AI have been on the rise in the sphere of conflict prevention due to the fact they can single out the risk patterns faster and with greater granularity compared to other existing monitoring systems. Through the processing of large volumes of data, e.g., social media activity, satellite images, data on mobile phones, or economic data, AI systems can detect anomalies, which could be signs of imminent violence. The tools prove to be especially appealing in unstable situations when state institutions possess insufficient ability to keep a track of the threats in real time. AI-based forecasting models can assist the decision-makers with probabilistic forecasts of the location of the violence, the kind of violence that may appear, and the populations that may be the most vulnerable. Such forecasts in theory enable an earlier intervention including diplomatic intervention, preventive deployments or humanitarian preparedness.

Another aspect of security where AI is becoming increasingly significant is its operational aspects. AI helps in surveillance, identifying targets and logistics optimisation in the battlefield. Military systems are being designed to include autonomous and semi-autonomous systems that bring in questions of speed and delegation of lethal decision-making. The researchers believe that because AI can process sensor data at scale, it will be able to improve situational awareness and minimize some types of operational uncertainty (Scharre, 2023). These benefits, however, are not devoid of threats: the very same systems that are making the process of deliberation slower can also become ways of making it faster and the dependency on automated pattern recognition can lead to misclassification that may prove fatal.

Outside the military sector, AI has been used more in peacebuilding and conflict transformation. Instead of emphasizing the use of coercive security, researchers and practitioners have investigated how AI can reinforce peace processes by enhancing the analysis and monitoring processes. Cases in point are that AI has been applied to chart hate speech, identify online incitement, and track the propagation of extremist discourse. Such applications are important as modern wars are becoming digital, with social media capable of transferring polarisation and violence at a very fast pace. Through the study of language patterns and engagement trends, the AI systems can detect the earliest signs of mobilisation and hostility thus providing a chance of preventive communication and community intervention opportunities (Kreutz et al., 2022).

The other sphere of innovation is the application of AI to predict violence and aid in planning humanitarian response. Machine learning models can distinguish the conditions that are correlated with the escalation of violence by combining the data about conflict events with environmental and socio-economic variables. They have been used to forecast riots, communal conflicts and election-related violence, potentially providing useful information to the peacekeepers, non-governmental

organizations and the local government. Crisis mapping can also be assisted by AI through the analysis of satellite data to identify the destroyed villages, displaced human populations or damaged infrastructure in order to target the humanitarian effort faster.

The analysis of negotiations, peace agreements, and political discourse have started to be helped with the help of AI tools in peace negotiations. Using natural language processing on historical peace agreements, AI can find similar patterns that relate to lasting settlements, e.g. power-sharing agreements, verification measures and inclusiveness. We can also rely on AI to guide mediation and negotiators to evaluate the way the public messaging changes during the negotiation cycles so that it becomes possible to adopt a more responsive engagement approach (Kreutz et al., 2022). The proposed developments indicate that AI can be used to ground the practice of mediation and peacebuilding, which can enhance the ability of the third parties to deal with the complex conflict situation.

Nonetheless, AI adoption in peace and security is not a controversial issue. The accumulating literature on critical studies cautions that in the context of AI systems, these tools are not objectively but socio-technical systems that are influenced by power dynamics, institutional agendas, and history-based disparities and inequalities. Among them is the issue of algorithmic bias, in which AI systems recreate and enhance discriminatory tendencies of training data or institutional procedures (Benjamin, 2019). Biased AI can be disastrous in conflict situations, such as the false identification of either individuals or groups as harmful, the strengthening of ethnic profiling, or the targeting of marginalized groups disproportionately.

The most similar one is the criticism of data colonialism, which points out that the extraction of data and the development of AI tend to reflect global inequalities. Most AI systems that are used in the Global South are the systems that are imported by outside forces, trained on data gathered without significant local consent, and deployed with little responsibility to the impacted communities. It makes it questionable in terms of ethical and political issues of sovereignty, ownership, and justice in digital governance. Unless AI systems are done in a way that is transparent and involving locals, they are likely to increase mistrust and recreate domination trends instead of promoting peace.

Ethical opacity is another significant issue that can also be referred to as the black box problem. Complex machine learning systems are notorious in giving outputs that cannot be explained even by its designers. The impossibility of providing explanations for algorithmic decisions is a deep ethical dilemma in the security context when the outcomes of the decision can be surveillance, detention, or lethal force. As long as AI-driven systems are unable to give understandable justifications, responsibility is lost, and the victims can have no effective way to appeal or complain (Benjamin, 2019; Dafae, 2018).

Concisely, the presence of AI in peace and security can be described by its potential and threat. Although AI may contribute to early warning, better monitoring, and evidence-based peacebuilding, it may also facilitate the strengthening of prejudices, further surveillance, and lack of responsibility. These tensions indicate that cautious

governance regimes are required that will make AI benefit human security and not harm it (Dafoe, 2018; Kreutz et al., 2022; Scharre, 2023).

2.3 Trust, Accountability, and Power in Digital Governance

With the incorporation of AI and digital technologies into peace and security systems, the issue of trust, accountability, and power has taken the focus of discussions regarding digital governance. Digital governance can be defined as institutional structures, rules, and control systems, in which societies govern digital infrastructures, data practices and algorithmic decision-making. Digital governance, in terms of peace and security, influences the application of AI tools in surveillance, risk forecasting, resource distribution, and planning its operations. Not only is the technical performance of these systems important but also the question of whether it is believed, responsible, and consistent with democratic principles.

The problem of trust is especially critical since the legitimacy, and the cooperation of the population are crucial to security governance. The credibility of the AI systems is not something that can be established without institutional credibility, transparency and a perception of fairness. The scholars predict that trust could be enhanced in the case of the deployment and design of AI systems that are transparent, explainable, and firmly controlled by the institution (Floridi et al., 2018). Transparency can be defined as the openness of the information regarding data being gathered, the way algorithms work, and the way their outputs are utilized. The process of explainability implies that AI decisions are interpretable and explainable to humans. Oversight is the independent monitoring agencies that have the capacity of auditing systems and setting standards and dealing with abuse.

Trust is particularly weak in peace and security environments since AI systems tend to overlap with force of state. Conflict may leave communities with a prior history of abuse, corruption or exclusion that unquestionably undermines trust of security institutions. The possibility of implementing AI-controlled surveillance or predictive policing in these settings can only make people more suspicious, at least not to create trust in communities that believe AI technologies are tools of control and not protection. Trust thus involves technical transparency in addition to significant involvement of communities to influence the way technologies are utilized.

The main issue with digital governance is that accountability is divided between networks of actors through the decision-making system of AI. The conventional models of accountability presuppose that there are identifiable individuals or institutions that can be responsible in making decisions. Nevertheless, in governance based on AI, decisions can be generated in complex ways involving human operators, software creators, data suppliers and institutional decision-makers. This produces so-called accountability gaps, in which the responsibility is distributed and hard to trace (Bovens, 2007). In cases of harm, i.e. wrongful targeting, discriminatory profiling, or unfair refusal of services, it is

unclear who is responsible: it can be the human operator, the institution, the programmer or the algorithm itself.

This decentralization of responsibility is not a technical matter alone; it is very political. Accountability in security governance is a vital requirement since life-threatening decisions can be adopted by right of coercion. When AI systems are applied to defence arrests or limit movement or assign protection resources, then aggrieved persons must have a way of appealing against the decisions and seeking redress. In the absence of these mechanisms, AI can become an instrument of irresponsible authority that will undermine human rights and law and order. Bovens (2007) points out that accountability is a pillar of democratic governance, which needs to have transparent accountability, transparency and penalties that can be imposed on malpractice.

Meanwhile, power is reinvented in digital governance by commanding data infrastructures, computational capacity and technical skills. Artificial intelligence systems are reliant on data streams, online platforms, and analytical instruments, which are usually owned by states, individual businesses, or foreign entities. The individuals who have dominance of these infrastructures are provided with the power to determine what is risky, it is their behaviour that is being observed and what kind of interventions they should prioritize. In this regard, AI does not merely facilitate governance--it forms governance by being visible, measurable and actionable.

These power changes have important ramifications on peace and security. This capacity to gather and process data can empower some actors and exclude other actors in conflict situations. To take the most recent example, AI-maintained surveillance can be used by state institutions to centralize the power, whereas communities cannot access the same resources to record the injustice or act in the interest of justice. Equally, parties outside the system can influence the security agendas using information-based systems of operation that favour quantifiable threat over experience. Such dynamics support the thesis that control over AI is not only the control over technology but the control over political power and authority.

According to Floridi et al. (2018), ethical governance of AI should be based on human values, such as fairness, accountability, respect for autonomy, and so on. This needs institutional changes to integrate AI into legal and normative regulations, such that the systems are auditable, explainable, and coherent with human rights practices. It further involves acknowledging that trust cannot be created by engineering with a technical design; it rests on political legitimacy, inclusion and fair sharing of power.

Finally, trust, accountability, and power are three aspects of digital governance that cannot be separated in the field of peace and security. Unaccountable trust can bring about blind trust on systems which might be detrimental. There is no point in having accountability but lack of transparency since no one can be accountable over an opaque decision. The potential misuse of power is that it may make the ruling domineering. To implement AI in any meaningful way in terms of peace and security, it should be regulated in such a way that human security is enhanced, rights are not violated, and the technological innovation

should not surpass democratic control (Bovens, 2007; Floridi et al., 2018; Benjamin, 2019).

3. Theoretical Framework

The current study is rooted in an interdisciplinary theoretical framework which unites Relational Mediation Theory, Sociotechnical Systems Theory and Participatory Peacebuilding. These two viewpoints together would present a solid conceptual framework through which the mediation processes can be viewed not only in terms of the actions of humans and the role of institutions, but also as the increasing role of digital and artificial intelligence (AI) systems in governing and resolving conflict scenarios. The framework views mediation more as a process than as an intervention, where trust, legitimacy, and authority are continuously discussed by the communication process, institutional structure, and the construction and use of technological means (Lederach, 1997; Latour, 2005).

To begin with, the paper relies on Relational Mediation Theory that perceives mediation as a communicative and a relationship-based process. According to this perception, mediation is not just a tool of obtaining agreements or a resolution to conflicts but a systematic process in which parties restore their recognition, mend their broken relationship and create new common meaning that enable peaceful coexistence (Bush & Folger, 2005; Winslade & Monk, 2000). The relational mediation methods stress the value of dialogue, empathy, active listening, and mutual understanding and express the view that sustainable conflict transformation involves alterations in the way parties to a conflict view each other and narratives about their conflicts (Lederach, 2003). As a result, the key to mediation effectiveness is not only trust between conflicting parties but also trust in the neutrality of the mediator, his/her competence, and integrity. This is notably critical in situations whereby social rifts, past resentments, and friction between states and society make the negotiated settlement a challenge (Bercovitch and Jackson, 2009).

Relational mediation theory is the approach applicable to the given study since the integration of AI-enhanced devices into the mediation (the automated risk forecasting, sentiment analysis, online reporting systems, or decision-support systems) may transform communicative dynamics and alter the nature of the relationships between parties. When mediation is essentially a form of trust interaction, the role of AI is either cemented or undermined based on how it is framed, who is in control and whether communities view it as equitable and open (Tyler, 2006; Mayer, Davis & Schoorman, 1995). As an illustration, AI tools would enhance confidence regarding mediation through the enhancement of early warnings, documentation and the management of evidence. However, they can also cause suspicion when parties feel that technology is biased, spying, or serving the interests of the elites. It is the relational lens that allows the study to assess AI as a technical intervention and as a variable that changes communication, credibility, and interpersonal recognition in the context of mediation.

Secondly, the paper applies Sociotechnical Systems Theory, which disputes the premise that technology is neutral, objective, or not dependent on the social context.

Sociotechnical views hold that technological systems are never constructed outside institutional contexts, political agendas, organisational practices and cultures; thus, they are a product of values and power relations in the contexts of their creation and implementation (Bijker, Hughes & Pinch, 1987; Orlikowski, 1992). This practice is in tandem with the academic research in science and technology studies (STS) that has shown that algorithms, digital platforms, and AI models can augment inequality when they encode preexisting prejudice, reproduce exclusionary classifications, or prioritize some results above other ones (Winner, 1980; Noble, 2018). In this perspective, AI should not be seen as a lone tool, but rather as a larger sociotechnical system where design decisions, sources of data, institutional incentives and governance structures determine how the system itself acts and whose interests it fulfils.

To use the sociotechnical systems theory to mediation, it is important to note that AI-mediated mediation activity cannot be assessed only based on the technical performance metrics like speed, accuracy, or prediction ability. Rather, the question that is important is what institutional logic and political agendas are being embedded in technology, and how do they affect mediation outcomes? (Jasanoff, 2004; Gillespie, 2014). As an example, an AI early-warning tool can be technically efficient in warning of the risks of conflict but still have detrimental effects when implemented in ways that are not applicable to all communities, enables securitised reactions, or is used as a rationale to impose coercive actions. In the same spirit, information-based mediation systems can discriminate against group members who do not have access to digital technologies, which further exacerbates disparities in participation and voice (Eubanks, 2018). The sociotechnical theory therefore substantiates the analytical basis in taking AI as an institutional actor whose impacts are mediated by governance institutions, political interests as well as unequal access to technological resources.

Third, there is the Participatory Peacebuilding, which fortifies the framework with the inclusion, local ownership, legitimacy, and consent being the key principles towards sustainable peace interventions. The participatory approach believes that communities that have been affected most by conflict have a better chance to succeed in their efforts to build peace when they are actively included in problem definition, intervening, and evaluating the outcomes (Paffenholz, 2014; Mac Ginty, 2011). This view also criticizes the existence of top-down, technocratic frameworks of peacebuilding that consider local actors as passive recipients as opposed to acting as agents of knowledge, capacity, and moral authority. Participation is not only a procedural matter, but it is a normative and a practical prerequisite to establish legitimacy and guarantee that interventions are also relevant to the local realities (Lederach, 1997; Richmond, 2011).

Participatory peacebuilding is especially applicable in the context of AI-mediated mediation since technological implementation tends to take place in externally motivated programmes, donor-supported digital technologies, or state-led digital governance projects that might not be able to engage with local communities properly. Unless there is

meaningful involvement, AI tools will have the potential of becoming a tool of exclusion, surveillance or elite control instead of a tool of conflict transformation (Mac Ginty & Richmond, 2013). Participatory peacebuilding requires the involvement of the affected communities in the process of developing what data is gathered, the way risks are perceived, the nature of privacy preservation, and accountability. It further emphasizes the fact that consent is mandatory: communities need to be informed about AI systems, but they must be able to accept, deny, or redesign them according to their experience in life and shared priorities (Paffenholz, 2014; UNDP, 2019).

Combined, these three different schools of thought will offer a unified theoretical approach towards the explanation of mediation as a human-AI assemblage, in which the power, legitimacy and results are co-created both by the technological apparatus and the human participants. Relational mediation theory predicts trust-based communicative underpinnings of mediation, sociotechnical systems theory provides the explanation of AI being embedded in power and institutions, and participatory peacebuilding demands the inclusion and local ownership as the condition of legitimacy and sustainability (Bush & Folger, 2005; Orlikowski, 1992; Mac Ginty, 2011). The integrated framework thus goes beyond naive beliefs that AI is going to automatically improve the efficiency of mediation. Alternatively, it conceptualises AI-mediated mediation as a problematic governance practice where trust, power, and legitimacy are negotiated, distributed, and created or destroyed through the interaction of people, institutions, and technologies (Latour, 2005; Jasanoff, 2004).

4. Methodology

4.1 Research Design

The research design chosen in this study, the qualitative multi-method research design, was used to produce rich and context sensitive data on how digital technologies and specifically, the emerging AI-enabled tools are defining modern mediation practice. The choice of a qualitative approach was deemed as one of the most suitable ones since the study is dealing with socially constructed issues like trust, legitimacy, power relations, and accountability in the context of mediation, which cannot be sufficiently described by numerical measures alone (Creswell & Poth, 2018). Through a mix of various qualitative approaches, the research enhanced its ability to investigate the realities that existed in the operation of digital mediation, as well as the perceptions held by the actors within the peace processes.

The study was designed based on three complementary elements. To begin with, the three mediation contexts where digital tools were actively used were analysed by means of the case study analysis. The case studies were chosen purposely to ensure variation of the type of mediation and institutional context, and case studies that include United Nations-supported mediation missions, regional peace efforts, and Track II or informal dialogue processes were chosen. Case studies constitute a beneficial methodology of analyzing the modern phenomena in a real-world setting and in detail, especially in cases when the demarcation between the phenomenon

itself and the context is unclear (Yin, 2018). In this way, the study could follow the introduction of digital tools, the types of influence that they had on the mediation process, as well as the underlying institutional and political conditions that influenced the use of digital tools.

Second, the semi-structured interviews with about 25-30 most important stakeholders were held in mediation and peacebuilding practice. The participants of the interview were mediators, diplomats, data analysts, and civil society actors, which is increasingly diverse in terms of the number of actors who are involved in the contemporary processes of peace. Semi-structured interviewing ensured consistency in the main subjects and provided a possibility to be flexible in exploring experiences and interpretations of the participants in a deeper manner (Kvale & Brinkmann, 2015). This approach came in especially handy in soliciting professional opinions on sensitive topics like the dangers of algorithmic bias, the politics of accessing data, and the possibility of digital tools to transform the mediator neutrality and legitimacy.

Third, the analysis incorporated participatory scenario workshops that are meant to re-create future mediation settings where AI tools assume an enhanced role. Through these workshops, an interactive environment was developed where stakeholders got to experiment on hypothetical mediation environments, consider opportunities and risks, and express the issue surrounding transparency, accountability, and human agency. Scenario-oriented approaches are those common methods aimed at exploring the unknown and forecasting the impacts that emerging technologies might impose on the complex governing setting (Van der Heijden, 2005). The study would be conducted by including participatory workshops to transcend the retrospective analysis to record the future looking perceptions and expectations regarding AI mediated peace processes.

4.2 Data Analysis

Thematic analysis was used to analyse all the qualitative data with codes being informed by the conceptual focus of the study on trust, accountability, power as well as legitimacy. Thematic analysis was chosen because it offers a methodological template to detect, categorise and analyse tendencies in qualitative data (Braun & Clarke, 2006). The coding was done in an iterative manner starting with the broad descriptive codes and then to more analytical ones that reflected a relationship between technology use and mediation results.

To improve methodological rigour, the research used triangulation in case studies, interviews, and scenario workshops. This allowed triangulation of results by juxtaposing patterns and discrepancies between various sources of evidence enhancing credibility and minimizing the possibility of being dependent on one narrative or point of view (Denzin, 2012). Triangulation was also very useful in the context of more profound interpretation, indicating how institutional settings influence the process of adopting digital tools, as well as attitudes to their legitimacy on the part of the stakeholders.

4.3 Ethical Considerations

The focus of the study was on the implementation of ethical safeguards because of the sensitivity of mediation settings and susceptibility of communities impacted by conflicts. The study was conducted in accordance with accepted ethical standards, such as the informed consent, confidentiality, and protection of the participants. The participants of the workshop gave informed consent and were assured of anonymisation to reduce the possible professional or political dangers of disclosure (Orb, Eisenhower & Wynaden, 2001). The research also embraced conflict-sensitive research practices, as the researcher acknowledged that when dealing with actors in the peace process, one has to be sensitive to power imbalances, trauma exposure and implications of involvement.

Besides, AI systems and digital tools were analyzed critically against bias, risk of surveillance, and data protection. Since the structural inequalities could be reproduced by the existing algorithmic systems based on the biased training data or obscure patterns of the decision-making, the study did not consider technology as neutral but rather based on the politics and institutional conditions (Noble, 2018). Additional considerations were given to the ethics of data collection, storage, and retrieval with an explicit focus on situations when digital mediation instruments can predispose sensitive data or pose a threat to participants in vulnerable environments.

5. Findings

The introduction of artificial intelligence into the sphere of mediation practice has created new possibilities to enhance conflict analysis and support the negotiation process, yet the changes in authority, trust, and responsibility have become complicated. Instead of being mere neutrals, which can benefit the work of mediation, AI systems tend to restructure the way mediation is designed, what knowledge is valued, and to which parties in conflict are perceived to be more legitimate. In a wide range of settings, AI usage has thus become not just a technical phenomenon, but a governance problem, a fact that affects the social relations that mediation efficacy is grounded on. The reconfiguration of the mediator authority is one of the most important effects of the AI-assisted mediation. Historically, the legitimacy of mediators has been related to relational expertise: the capacity to establish rapport and exhibit empathy, interpret poorly understood nonverbal communication, and maneuver political sensitivities based on experience and interpersonal credibility. But with the integration of AI tools into the mediation processes, the power distance is transferred into technological knowledge increasingly. Situational awareness can be improved with the help of AI as it is able to process a great amount of data (incident reports, social media content, satellite images, or historical conflict data) effectively and quickly to detect patterns and predict dangers. Their abilities may be useful in assisting mediators to predict the causes of escalation, map network of actors, and design negotiation strategies more accurately. However, mediators are equally likely to fall into a new dependency where they depend more on analysts, data scientists and algorithmic outputs to influence decision-making.

Practically, such dependence establishes new power structures of influence in mediation teams. The people who are in control of how data are interpreted, how analytical models are designed, or how algorithmic indicators are chosen can start to have disproportionate power over the priorities of mediation. Consequently, technical authority can replace relational competence and contextual familiarity indirectly. This does not imply that the mediators become irrelevant, but it rather implies that the mediators may no longer be the only source of strategic judgment but a translator of the technical to the negotiation judgment. This may over time alter the perceived credibility of knowledge in mediation process, privileging the area of quantifiable indicators and predictive products above lived experience, cultural subtlety, and community-based knowledge of conflict.

Tightly related to these shifts in power is the issue of the trust and appearance between sides of the conflict. Mediation, in essence, is a process which requires trust: the parties should feel confident that the mediator is impartial, that his/her interests are heard and that the process will not put the parties into new perils. In some instances, AI may reinforce this trust into giving the impression of objectivity. As an illustration, the parties to a dispute might perceive AI-supported evaluations to be more objective than a human one, especially when a mediator is likely to support one side or have an ulterior motive. As artificial intelligence devices are applied to confirm assertions, clear up dates or provide evidence in systematic forms, some parties might believe that the mediation process is more equitable and believable.

But such a perception is not universal. In most environments, AI-based mediation may send out suspicion as opposed to confidence. Other parties view AI involvement as a measure of foreign surveillance, particularly when the tools are created or financed by third parties. Some might consider it to be some technocratic usurpation by local wisdom with remote forms of expertise. When the context is defined by colonial backgrounds, geopolitical mistrust or the effects of states being oppressed the deployment of sophisticated technologies can only serve to fuel the fear that mediation is being employed to collect information as opposed to the process of peace. Due to this, trust is therefore highly dependent upon transparency, i.e. do the mediators describe transparently how AI insights are obtained, what types of data they are being used, and how their outputs lead to a decision. The risk of introducing AI tools as a black box where the reasoning process cannot be disputed is that it leads to the loss of the validity of the mediation process regardless of their high accuracy of the analysis.

In addition to trust, greater issues are also concerned with accountability gaps in mediation outcomes in the case of AI integration. Mediation is a process of making strategic decisions: what issues to focus on, who to consult, whether to insist on compromise or not, when to postpone negotiations. Where such decisions are guided by algorithmic suggestions - like predictive notices of escalation, or risk rating of actors- then accountability is lost. In the case of a successful mediation strategy, the human mediator can get credit of the result. However, in its failure, it may not be clear whether it was because of human

judgment, inaccurate data, biased algorithms, or poor interpretation of AI outputs. This has not been merely a hypothetical ambiguity. In mediation under real-life setting failure can have tremendous implications, such as returned violence, loss of trust or political retaliation. In the absence of well-defined accountability frameworks, AI may cause a diffusion of responsibility where none of the actors could be held entirely responsible when making decisions that were influenced by the technological suggestions.

The concern of accountability is further aggravated by the inclination to consider algorithmic products as authoritative. Although AI systems are characterized as such, decision-support tools, the prediction of the systems can have a psychological effect, whereby mediators would act on advice to evade blame. This may result in some kind of algorithmic difference, when humans have a sense that the advice offered by AIs must be followed because it seems to be based on evidence, systematic, and justifiable. This may eventually diminish the room in ethical judgement and contextual discretion- two attributes that are critical in the mediation whereby inflexible solutions are never effective in a wide cultural and political setting.

AI also implementation in mediation frequently reveals participative inadequacies that denote the reproduction of the power imbalance. Often, local actors and impacted communities, in most instances, are hardly consulted in decisions regarding the use or otherwise of AI, the kind of data to be gathered, and the application of the results. This isolation strengthens trends of outside agencies bringing technological intercessions without significant ownership back home. It may also facilitate data mining habits, where local experience and data are mined to develop models without sufficient consent, benefit-sharing, and responsibility. Where communities do not feel that they are being considered as partners but as objects which are being studied, AI is linked to exploitation instead of support.

Lastly, AI in mediation frequently reveals the lack of participation that recreates the lack of power. Local actors and communities to which AI may affect are often not consulted during the decision-making process to whether to use AI or not, the data to be collected or not, or how the outputs should be utilized. This omission supports trends of introduction of technological interventions by external institutions without significant local ownership. It may also empower data harvesting activities, where local experiences and data are mined to construct models with an insufficient level of consent, sharing benefits, or responsibility. AI is linked to exploitation instead of assisting communities when communities believe they are being researched, and not as partners.

These participatory gaps are important in that, in mediation, it is not just about agreement-making but also developing legitimacy and common ownership of peace processes. Unless introduced with involvement of communities, AI tools may only increase mistrust, intensify beliefs of elite domination, and disenfranchise local knowledge systems that are commonly instrumental in the establishment of sustainable peace. Thus, although AI can increase the technical capacity of the mediator, its value over time will be determined by the manner in which it is incorporated, in a way that does not affect relational

legitimacy, builds trust by creating transparency, have clear accountability, and inclusive participation in both design and implementation.

6. Discussion of Findings

This research has shown that the introduction of artificial intelligence (AI) into the mediation practice will not only be an efficiency-saving innovation, but a paradigm shift change that will alter the role of authority, trust, and accountability in peace processes. Instead of supporting decision-making in a neutral way, AI systems will be engrained in social relations, institutional frameworks and legitimacy processes that define mediation outcomes. This discussion explains these findings based on three theoretical frameworks that are complementary to each other, namely Relational Mediation Theory, Sociotechnical Systems Theory, and Participatory Peacebuilding, to demonstrate how AI can impact effectiveness by changing interpersonal and systemic conditions under which conflict resolution occurs.

According to the Relational Mediation Theory, mediation legitimacy is based on the capacity of the mediator to develop trust, empathy, and interpersonal credibility between the conflict parties. The mediation authority as conventionally established is by relational competence: perceiving of emotional expression, awareness of power sensitivities, assertion of dignity, and maintenance of dialogue amid confrontation. The results show that AI-mediated mediation creates a significant change in this basis of power. Although AI tools improve situational awareness through processing massive amounts of data, including social media trends, previous conflict patterns, and risk sensors, this technical ability can also be used to refreeze the role of the mediator, shifting them away to become a technical one, as opposed to a relational one. This corroborates the relational mediation theory as it depicts that legitimacy is not necessarily enhanced by information superiority but rather legitimacy is developed socially by interactional trust and perceived fairness. By relying on algorithmic insights, in which the mediators appear to be more detached and distant to lived realities, the conflict parties can become less convinced that the mediator is emotionally receptive and responsive to the realities that hold the dialogue together. In this regard, AI will not substitute relational legitimacy but can push it aside in case the technical authority takes pre-eminence over human interaction.

Meanwhile, the results also indicate that AI may be used to supplement relational mediation in certain circumstances. The application of AI in a transparent way to validate the claims, organize the evidence, or explain the disputed dates may bolster the perceptions of fairness and minimize the suspicion of biased mediators. Such a result can be explained by the focus of relational mediation theory on building trust because AI can contribute to increased trust in the case when it is assumed that it is an unbiased tool that helps to facilitate a balanced situation. Nevertheless, the results warn that trust still depends on the context: where people are historically distrusted, geopolitics is worried, or people think they are being spied on, AI intervention can be perceived to be intrusive and authoritative instead of helpful. Therefore, the relational

mediation theory can support the explanation of the reasons why AI can make it or break it in the same system depending on the circumstances. It is not that technological accuracy is important but that the mediator is able to maintain relational credibility by remaining open, culturally sensitive and providing ethical reassurance.

The results are also the best reason supporting the Sociotechnical Systems Theory that notes that technologies cannot be comprehended outside of the social ecosystems in which they are used. AI applications in mediation are not perceived as standalone tools; they are in a network of human decision-makers, institutional interests, databases, and organisational structures. The perceived reconfiguring of mediator authority denotes a sociotechnical change wherein new actors, such as data analysts, technical consultants, algorithm designers and platform managers, become pertinent to the priorities of mediation. This brings new knowledge hierarchies, disconnecting the relational knowledge and the familiarity that comes with a place to an indicator and predictive performance. According to the sociotechnical systems theory, it is a redistribution of agency: the agency is transferred to both humans and machines but not equally. The actors that manipulate the system - model selection, indicator design and interpretation - determine what is displayed as a problem and what is deemed as evidence in negotiation strategy.

Notably, the development of the accountability gaps is also explained by the sociotechnical systems theory. The results indicate that in situations whereby algorithmic suggestions are used in strategic decision-making, accountability is obscured between the mediator judgment, data constraints, and system biases. This is standard sociotechnical paradox: results are created because of hybrid interactions, but there is still an accountability structure destined to work based on human-only decisions. This spread of responsibility is especially hazardous in the case of mediation where failure may result in a new eruption of violence or even political pressure. The identification of the sociotechnical process in the study of the so-called algorithmic deference, the tendency of mediators who can accept AI recommendations to escape responsibility, is also another indication of how sociotechnical systems can create changes in behaviour. Though AI may be represented as a piece of advice, its results gain symbolic power as they seem to be in a system and are defensible. This supports the sociotechnical argument that technology influences the behaviour of human beings not only for a purpose but also by perceived legitimacy, institutional imperative and by motivators of risk management.

The results also support Participatory Peacebuilding as an important model of the assessment of the use of AI in mediation. Participatory peacebuilding posits that inclusion, local ownership and appreciation of community knowledge as a fundamental resource to conflict transformation are central to sustainable peace. The analysis indicates that the implementation of AI tends to reveal the lack of participation: the local actors are often not involved in making decisions related to the use of AI, the types of data that need to be gathered, and the effects that the results of this technology should have on the negotiation process. This isolation strengthens the status

quo of power imbalance, especially where the introduction of AI systems does not include any meaningful consent or benefit distribution with other institutions. Participatory peacebuilding can be used to understand why this exclusion leads to a lack of legitimacy: communities can take AI as being not supportive but extractive, surveilling, or dominated by the elite. Here AI is linked with domination, as opposed to dialogue, undermining the social support of peace agreements that are necessary to last.

Moreover, participatory peacebuilding brings out the epistemic injustice of AI-mediated mediation. Focusing on measurable pointers and predictive modeling, AI systems can sideline local knowledge systems that feel subtlety, history, and contextual relationships. The results also show that the effectiveness of mediation lies not only in enhanced technical analysis but in the perception of the conflict parties that are seen, respected and involved in the peace process development. This supports the participatory argument that it is the production of legitimacy through the creation of solutions which are not dictated to people but through shared ownership. The addition of AI would have a positive impact only when applied in participatory design, clear communication, and the means to enable the societies to question, challenge, and contextualise the outputs of the algorithms.

Altogether, the results prove AI-assisted mediation can be assumed to be a relational, sociotechnical, and participatory phenomenon. The theory of relational mediation describes how AI reinvigorates the principles of trust and mediator legitimacy, the theory of sociotechnical systems explains the redistribution of power and responsibility, and the theory of participatory peacebuilding explains the ultimate importance of inclusion and local ownership. Collectively, these frameworks indicate that the value of AI in mediation is not defined by the power of its analysis only, but by the governance decisions that inform the utilization of AI. Successful AI-mediation must thus have planned mitigation to maintain relational legitimacy, transparency and responsibility in the decision-making process, as well as integrating AI into participatory peacebuilding processes in a manner that empowers and does not marginalise the affected communities.

7. Conclusion

The world of international mediation is being hastily transformed by artificial intelligence, which can provide potent means, helping to reinforce early warning initiatives, better information processing, enhance scenario forecasting and promote more coordinated peace diplomacy. When properly used, AI can increase the ability of mediators to read more complex conflict contexts, detect conflict escalation patterns and create interventions that are timelier and context sensitive. Such opportunities are specifically useful in the contemporary world where the conflicts continue to become multidimensional, rapidly changing, and affected by digital fake news, transnational networks, and divided political power.

Nevertheless, there are also severe ethical and political risks, which cannot be overlooked in case of the increasing utilization of AI in mediation. The effect of algorithmic bias, making decisions without understanding

how they occurred, and inequalities in access to data as well as the possibility of surveillance or manipulation could undermine the validity of mediation interventions and increase distrust between parties to a conflict. Artificial intelligence can replicate existing power dynamics, advantages dominant discourses, or create the falsehood of neutrality by producing contentious political facts as objective data. Such risks may diminish confidence in the peace processes and undermine the credibility of the mediators in weak contexts where there are already limited trust and institutions are in contention.

This is why in the AI era effective peace diplomacy will not be simplistic technological optimism. Rather, it will need an ideological vision of hybrid governance that entails transparency, accountability, human rights, and substantive involvement in the core of AI-mediated mediation. The use of AI tools should be viewed as a decision support mechanism, not a decision maker so that the human mediator holds responsibility and can make ethical, cultural, and politically wise decisions.

Conclusively, the future of mediation is not in the phenomenon of automating mediation solutions, but in making human judgment stronger and more responsible by engaging in responsible human-AI co-operation. AI can become a positive collaborator in peacebuilding when informed by effective ethical principles and inclusive governance models to improve the mediation process without sacrificing the legitimacy and integrity of international conflict resolution.

8. Recommendations

The piece suggests that Hybrid Mediation Governance Framework should be adopted to facilitate the responsible use of artificial intelligence (AI) in mediation, specifically in situations where disputes are sensitive, identity-related, and politically sensitive. It is based on the idea that, although AI has the potential to enhance mediation with such capabilities as speed, pattern identification, and decision-support, it equally poses the threat of major ramifications when implemented without ethical protection, on-the-ground legitimacy, and institutional responsibility. Hence, instead of this system substituting the human decision-making, the suggested framework makes AI an assistant tool integrated into the human-centred governance framework, where the main focus is placed on fairness, trust, and contextual sensitivity. First, the framework is focused on human-in-the-loop decision making to make sure that AI is not decisive but only consultative in mediation results. Mediation in peace and conflict contexts is not a technical activity that involves the choice of the best solutions; rather, it is a relational and political activity that is framed by history, feelings, identity, and sense of justice. Therefore, AI-based suggestions can never be viewed as ultimate decisions. Rather, trained mediators should have the right to interpret, authenticate, and modify AI outputs, depending on localities, cultural prototypes, and the experience of the disputing parties. Such a strategy maintains the agency of human beings, safeguards legitimacy of mediation procedures, and it minimizes the chances of entrusting automated systems with sensitive forms of moral and political decisions.

Second, the framework suggests the introduction of algorithmic transparency procedures that allow AI-based mediation processes to be coherent to the parties in conflict. The black box problem is one of the most significant risks of AI in conflict management because a decision is made to seem like something that transpired through an invisible or unquestionable mechanism. This kind of opaqueness may become even more suspicious during the mediator and may strengthen their perceptions of bias particularly amongst low-trust communities. Transparency procedures consequently demand that the mediators are able to articulate transparently what information the system consumes, how the results are formed and what constraints or doubts there are in the product of the model. Such does not imply disclosing proprietary code but rather making sure that AI assistance is conveyed in a language that can be understood by parties so that they can challenge, dispute, or refuse AI propositions without the fear that such will be imposed upon them by the so-called technology.

Third, the article suggests participatory AI design, which is based on the involvement of the local peace actors and the community stakeholders in the development and adaptation of the AI tools. The culture of mediation is different in various situations, and the local norms, the relations of power, and the social identities determine the conflict dynamics. The systems generated by AI that are not local input are prone to recreating the external assumptions that can be culturally inappropriate or even politically detrimental. A participatory design process means traditional leaders, civil society organisations, women groups, youth networks and grassroots mediators are involved in setting the values, indicators and decision rules, which would be integrated in AI tools. This enhances the cultural validity of AI-assisted mediation, and it is more likely that the communities will trust and embrace technology as a supportive tool and not a solution forced on them.

Fourth, this framework proposes that ethical oversight committees be formed to observe AI application in mediation and hold it accountable. Because AI systems may increase bias, misread the context, or lead to unequal results, control should be institutionalised and not considered an option. The boards of ethical review ought to consist of a multidisciplinary team of professionals, including peace and conflict studies, law, data science, human rights, sociology, and community leadership. They would be involved in approving the AI deployment procedures, hearing complaints, risking unfairness and harm, and holding the national law norms and the international human rights standards. These committees would also assist in the governance of data, privacy assurances and ensure mediation information is not abused by spies or political interference.

Lastly, the article mentions the necessity of mediator capacity building, especially by means of digital literacy and AI-awareness training. Even the most ethically oriented AI tools are not effective in the mediators who do not have the skills to interpret outputs, define bias, and convey the results in a responsible manner. Mediators should thus be provided with practical skills in the interpretation of simple data, awareness of ethical risks, handling of privacy, and the constraints of algorithmic

reasoning by training programmes. This capacity-building aspect is needed to make sure that the mediators do not lose their empowered professionals who can implement AI in a strategic way but still have the human judgement, empathy, and contextual understanding that effective peacebuilding needs to have. Overall, the Hybrid Mediation Governance Framework offers a balanced pathway for leveraging AI innovation while safeguarding the integrity, legitimacy, and justice orientation of mediation processes. It ensures that technology strengthens, rather than undermines, peacebuilding practice by anchoring AI within transparent, participatory, and ethically accountable human institutions.

References

- Ateeq, A., et al. (2024). The impact of AI as a mediator on effective communication: Enhancing interaction in the digital age. *Frontiers in Human Dynamics*, 6, 1467384. <https://doi.org/10.3389/fhumd.2024.1467384>
- Banda, J. H. K. (2025). The politics of mediation: Analyzing the influence of international actors on domestic conflict resolution and state sovereignty. *Journal of Political Science and International Relations Studies*, 2(1), 1–10. <https://doi.org/10.61784/jpsr3006>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Bercovitch, J., & Gartner, S. S. (2006). Is there method in the madness of mediation? Some lessons for mediators from quantitative studies of mediation. *International Interactions*, 32(4), 329–354. <https://doi.org/10.1080/03050620601048344>
- Bercovitch, J., & Jackson, R. (2009). *Conflict resolution in the twenty-first century: Principles, methods, and approaches*. University of Michigan Press.
- Bijker, W. E., Hughes, T. P., & Pinch, T. (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brummans, B. H., Higham, L., & Cooren, F. (2022). The work of conflict mediation: Actors, vectors, and communicative relationality. *Human Relations*, 75(4), 764–791. <https://doi.org/10.1177/0018726721994180>
- Bush, R. A. B., & Folger, J. P. (2005). *The promise of mediation: The transformative approach to conflict* (Rev. ed.). Jossey-Bass.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
- Dafoe, A. (2018). *AI governance: A research agenda*. Governance of AI Program, Future of Humanity Institute, University of Oxford.
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), 80–88. <https://doi.org/10.1177/1558689812437186>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Fjelde, H., & Hultman, L. (2021). Weakening the enemy: A disaggregated study of violence against civilians in conflict. *Journal of Peace Research*, 58(3), 427–441. <https://doi.org/10.1177/0022343320986718>
- Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.
- Jasanoff, S. (2004). *States of knowledge: The co-production of science and social order*. Routledge.
- Kreutz, J., Fjelde, H., & Öberg, M. (2022). Using machine learning to detect early warning signals of political violence. *International Studies Quarterly*, 66(2), sqac014. <https://doi.org/10.1093/isq/sqac014>
- Kvale, S., & Brinkmann, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). Sage.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Lederach, J. P. (1997). *Building peace: Sustainable reconciliation in divided societies*. United States Institute of Peace Press.
- Lederach, J. P. (2003). *The little book of conflict transformation*. Good Books.
- Mac Ginty, R. (2011). *International peacebuilding and local resistance: Hybrid forms of peace*. Palgrave Macmillan.
- Mac Ginty, R., & Richmond, O. P. (2013). The local turn in peace building: A critical agenda for peace. *Third World Quarterly*, 34(5), 763–783. <https://doi.org/10.1080/01436597.2013.800750>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Orb, A., Eisenhauer, L., & Wynaden, D. (2001). Ethics in qualitative research. *Journal of Nursing Scholarship*, 33(1), 93–96. <https://doi.org/10.1111/j.1547-5069.2001.00093.x>
- Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, 3(3), 398–427. <https://doi.org/10.1287/orsc.3.3.398>
- Paffenholz, T. (2014). Civil society and peacebuilding. *Journal of Peace Research*, 51(1), 69–82. <https://doi.org/10.1177/0022343313501888>
- Richmond, O. P. (2011). *A post-liberal peace*. Routledge.
- Scharre, P. (2023). *Four battlegrounds: Power in the age of artificial intelligence*. W. W. Norton & Company.
- Shahzad, M. F., et al. (2024). Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning. *Heliyon*, 10(8). <https://doi.org/10.1016/j.heliyon.2024.e29523>
- Svensson, I. (2020). Mediation of interstate conflicts and civil wars. *Oxford Research Encyclopedia of International Studies*. <https://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-528>
- Tyler, T. R. (2006). *Why people obey the law* (2nd ed.). Princeton University Press.
- UNDP. (2019). *Digital strategy 2019–2021*. United Nations Development Programme.
- van der Heijden, K. (2005). *Scenarios: The art of strategic conversation* (2nd ed.). Wiley.
- Winslade, J., & Monk, G. (2000). *Narrative mediation: A new approach to conflict resolution*. Jossey-Bass.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.
- Zartman, I. W. (2001). The timing of peace initiatives: Hurting stalemates and ripe moments. *The Global Review of Ethnopolitics*, 1(1), 8–18. <https://doi.org/10.1080/14718800108405087>
- Zhou, Q., et al. (2025). The mediation of trust on artificial intelligence anxiety and continuous adoption of artificial intelligence technology among primary nurses: A cross-sectional study. *BMC Nursing*, 24(1), 724. <https://doi.org/10.1186/s12912-025-03406-0>