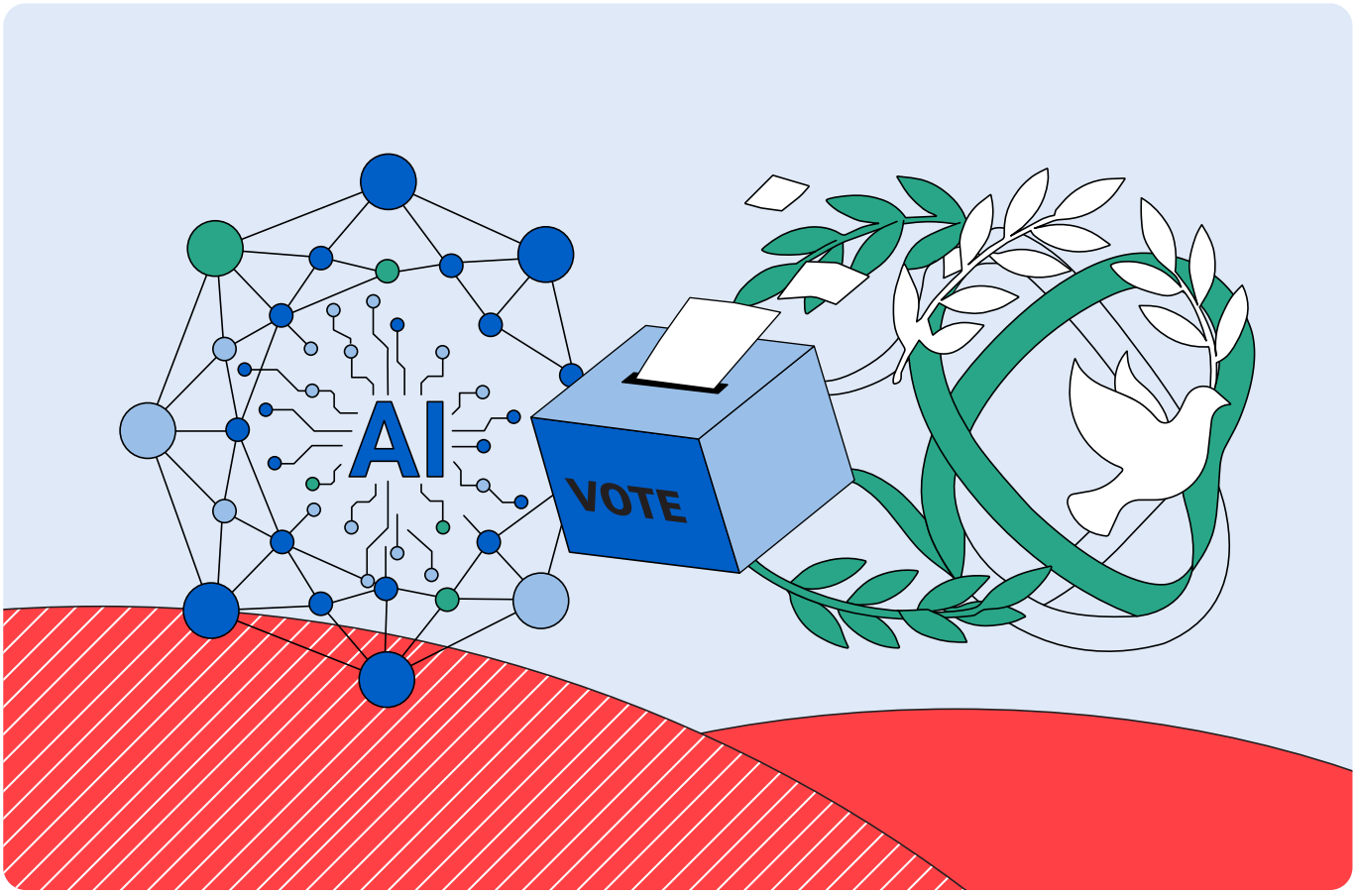
FBA Research report

AI and the New Shape of Democratic Power and Peace Processes

May 2026

Editors: Catalina Perdomo and Catalina Uribe Burcher



Folke Bernadotte Academy
Swedish agency for peace, security and development

FBA

This report is published by the Folke Bernadotte Academy (FBA). It was produced as part of FBA's work on artificial intelligence, democracy and peace processes, and brings together contributions from several authors and experts. The main chapters were written by Slava Jankin, Catalina Perdomo, Catalina Uribe Burcher and Martin Wählisch, with additional case-study contributions from practitioners and researchers working on AI, democratic governance and peacebuilding. The report examines how AI is reshaping democratic resilience and peace processes, with a focus on both the risks and opportunities this presents for institutions working in conflict prevention, democracy support and peacebuilding.

How to refer to this brief:

Jankin et al. 2026.
AI and the New Shape of Democratic Power and Peace Processes.
FBA Report, May.
Stockholm: Folke Bernadotte Academy.

DOI: https://doi.org/10.61880/SZFU2833

Content

Acknowledgements

by Catalina Perdomo and Catalina Uribe Burcher

Addressing complex topics with global implications requires the expertise of people working in many different areas. As the editors and co-authors of this report, we would like to extend our sincere appreciation to everyone who has contributed to its conception, development and production.

We would like to begin by thanking the chapter authors Slava Jankin and Martin Wählisch, whose expertise and insights form the core of this report. We are also grateful to Prathm Juneja for the case study of India and Andrii Bohdanenko of Osavul for the example of Moldova in the section on AI and elections. We would also like to thank Natalia González for the Chilean example in the section on public service delivery and Allan Cheboi and the Build Up team for the examples from Kenya, Somalia and Sudan in the section on citizen engagement and peace dialogues. We are grateful to Lisa Schirch for her contribution from Israel–Palestine to the section on AI in digital dialogues, Fadi Zaraket for the example of using AI for natural language processing and automated reasoning in the monitoring of digital spaces in Arabic and John Jaeger for his contribution from Ukraine and Yemen to the discussion of AI and ceasefire monitoring.

At the Folke Bernadotte Academy, we would like to thank Lisa Orrenius and Pontus Ohrstedt for their strategic guidance and oversight and Mathilda Lindgren for her invaluable support in the production and review processes. We are also grateful to Leire Bustamante Törnqvist for the steadfast logistical support that made the completion of this report possible. We are grateful to Essen International and Comactiva for their work copy-editing and laying out the final report.

Finally, we would like to thank our external reviewers, Branka Panic and Magnus Lundgren. Their constructive feedback and thoughtful reflections have greatly strengthened the final publication.

Foreword

This report comes at a defining moment in contemporary global politics. The international political landscape is undergoing a profound transformation, one marked by increasing geopolitical competition, democratic backsliding, and rapid technological change. The relative stability that characterized the politics of the period following the end of the Cold War has given way to renewed uncertainty. Today, global and regional powers compete for influence, military expenditures are rising, and nationalist and authoritarian tendencies are gaining ground – even within established democracies. At the same time, the legitimacy and effectiveness of longstanding and highly respected peacebuilding institutions and instruments are increasingly questioned. These developments are unfolding alongside the emergence of artificial intelligence (AI) as a transformative technology with far-reaching social and political implications.

Throughout history, states have viewed technological innovation as an important mechanism for enhancing their security. Despite their investment in increasingly advanced military technology, the outcome has not always been greater peace or political stability. While strong defence capabilities remain essential, experience shows that they are insufficient on their own to create lasting security. Arms races and technological competition can generate uncertainty and insecurity, as demonstrated during the Cold War, and can lead to devastating human and environmental consequences, as the events at Hiroshima and Nagasaki at the end of the Second World War have made plain. Sustainable peace requires more than deterrence; it depends on resilient institutions, inclusive political and dialogue processes, and long-term investments in conflict prevention and peacebuilding.

Military responses and long-term peacebuilding efforts should not be seen as competing priorities. Rather, they should be understood as complementary components of a comprehensive approach to peace and security. As with so many other complex modern problems, emerging technologies offer both opportunities and risks to contemporary peacebuilding. While unregulated technologies can present a significant threat to political stability, harnessing technological innovation in a responsible fashion offers the potential to prevent conflict, support democratic governance, broker peace, and sustain peace over time.

One technological innovation that has had a significant impact on contemporary global politics is the emergence of AI. In just a few years, AI has increasingly shaped geopolitical dynamics and introduced new complexities into international relations through intensified technological competition. It has also raised challenges related to information integrity and trust, both of which are central to democracy and peacebuilding efforts. However, AI also offers real opportunities for peace. This report examines how the integration of AI can support conflict prevention and peacebuilding efforts. It highlights emerging applications in democratic support and peace processes – two core areas of the work of the Folke Bernadotte Academy (FBA). At a time when attention often gravitates toward military responses, and when international support for long-term peace and democracy initiatives faces increasing pressure, exploring new tools and approaches is both necessary and timely.

AI is not a solution in itself. It is a tool whose impact will depend on how it is designed, governed, and applied. Used responsibly, it has the potential to enhance analysis, decision-making, inclusion, and effectiveness in peace efforts. Used poorly, it risks reinforcing inequalities, biases, and mistrust. For policy makers and peace professionals, understanding both the potential of AI and its limitations is no longer optional: it is essential.

For FBA, strengthening the global capacity for peace, security, and development remains at the core of our mandate. We hope this report contributes to an informed dialogue among policymakers and practitioners about the ways emerging technologies can be leveraged to support more effective conflict prevention and peacebuilding. At a time of increasing global uncertainty, investing in innovative, responsible, and forward-looking approaches to peace and democracy has never been more important.

Per Olsson Fridh
Director General, Folke Bernadotte Academy

Glossary of Terms

- **Algorithmic bias:** Systematic and repeatable errors in AI systems that produce unfair outcomes, often reflecting biases present in training data, model design or deployment.
- **Algorithmic transparency:** The degree to which the logic, data sources and decision-making processes of an algorithm or AI system can be understood and scrutinised.
- **Artificial Intelligence (AI):** The ability of machines or systems to perform tasks that normally require human intelligence, such as reasoning, learning, perception and language understanding.
- **Augmented reality:** A technology that, to enhance perception and interaction, overlays digital information, such as images, data or text, onto the real world.
- **Big data:** Extremely large datasets that can be processed computationally to identify patterns, trends and associations, especially relating to human behaviour.
- **Blockchain:** A distributed digital ledger that records transactions transparently across a network of computers, ensuring data integrity and traceability.
- **C2PA specification:** An open technical standard from the Coalition for Content Provenance and Authenticity that allows creators to attach secure metadata to digital content indicating how it was made and edited.
- **Chatbot:** A computer programme that uses AI or natural language processing to simulate human conversation through text or voice interaction.
- **Content provenance:** Information establishing the origin, creation process and modification history of digital material that is then used to verify authenticity.
- **Cryptographic provenance signals:** Tamper-resistant digital markers that prove the origin and modification history of a piece of content.
- **Data asymmetry:** An imbalance in the access to or control over data between actors that can, in turn, create power differentials and accountability gaps.
- **Data governance:** The policies, standards and procedures that ensure data are managed responsibly, securely and ethically across their lifecycle.
- **Deepfake:** Synthetic media, such as audio, video or images, created or altered with AI, to convincingly depict false events or identities.
- **Digital divide:** The gap between individuals, communities or countries that have access to digital technologies and those that do not.
- **Disinformation:** False information that is intentionally created or spread to deceive or harm.
- **E-governance:** The use of information and communication technologies by governments to improve transparency, efficiency and citizen participation in public administration.
- **Explainability:** The extent to which an AI system's reasoning, predictions or decisions can be understood and justified by humans.
- **Generative AI:** A class of AI systems capable of generating new content (e.g., text, images, audio or video) based on patterns learned from data.
- **Hallucination:** A phenomenon in which an AI system produces incorrect, fabricated or unsupported outputs and presents them as factual.
- **Hybrid threats:** Coordinated and adaptive actions combining conventional and non-conventional tools, such as disinformation, cyberattacks and illicit financing, to undermine democratic institutions.

- **Information disorder:** The spectrum of information problems (e.g., misinformation, disinformation and malinformation) that distort public understanding and harm democratic discourse.
- **Interoperable integration:** Ensuring that systems from different agencies can securely communicate and exchange data through standardised interfaces.
- **Large Language Model (LLM):** An AI model trained on vast text datasets to generate human-like language and perform tasks such as summarisation or translation.
- **Machine learning (ML):** A subset of AI in which systems automatically learn from data to improve their performance without being explicitly programmed.
- **Malinformation:** Genuine information shared with harmful intent, often by taking it out of context or exposing private details.
- **Misinformation:** False or misleading information shared without intent to harm, but that can still cause damage if believed or disseminated.
- **Model card:** A standardised documentation template that provides details about an AI model's purpose, performance, limitations and ethical considerations.
- **Natural language processing (NLP):** A branch of AI focused on enabling computers to understand, interpret and generate human language.
- **Open-source intelligence (OSINT):** Information gathered from publicly available sources, including media, online platforms and satellite data, used for analysis and situational awareness.
- **Predictive analytics:** The use of data, statistical algorithms and machine learning to identify patterns and forecast probable outcomes.
- **Responsible AI (RA):** The development and use of AI in ways that are ethical, transparent, accountable and aligned with human rights and democratic values.
- **Sentiment analysis:** The use of AI to determine the emotional tone or attitude expressed in text or speech, often applied to assess public opinion.
- **Synthetic-media detectors:** Tools designed to identify artificially generated or manipulated digital content such as deepfakes.
- **Transparency record:** A publicly accessible document describing an AI system's purpose, data sources, performance and safeguards to ensure accountability.
- **Vendor lock-in:** A situation in which an organisation becomes so dependent on a particular supplier's technology or service that switching providers is prohibitively difficult or costly.
- **Virtual reality:** A fully immersive, computer-generated environment that allows users to experience and interact with simulated surroundings, sometimes used for training or peace education.

Acronyms

AI:	Artificial Intelligence
AR:	Augmented Reality
ATRS:	Ad Transparency Reporting System
C2PA:	Coalition for Content Provenance and Authenticity
CDEI:	Centre for Data Ethics and Innovation (United Kingdom)
DSIT:	Department for Science, Innovation and Technology (United Kingdom)
EMB:	Election Management Body
EU:	European Union
FBA:	Folke Bernadotte Academy
LLM:	Large Language Model
NATO:	North Atlantic Treaty Organization
NIST:	National Institute of Standards and Technology (United States)
NLP:	Natural Language Processing
OSINT:	Open-Source Intelligence
RA:	Responsible AI
RACI:	Responsible, Accountable, Consulted and Informed (matrix)
RIA:	Information System Authority (Estonia)

Executive Summary

This report examines how artificial intelligence (AI) is reshaping democratic governance and peace processes. It outlines the impact AI has had to date and attempts to explain what this transformation means for institutions working in the field of conflict prevention, democracy support and peacebuilding. As AI systems increasingly influence information flows, analytical capabilities, public administration and citizen engagement, this technology is coming to play a structural role in contemporary power dynamics.

While AI offers significant opportunities to enhance public service delivery, democratic participation and peace process support, it also introduces the serious threats of information manipulation, opacity, and bias along with weakened accountability structures. These threats are particularly acute in fragile and conflict-affected contexts, where institutions are under strain and trust in public authority is often limited. This report seeks to respond to this challenge by analysing the influence AI is already having on democratic resilience and peace processes and by identifying some practical ways of ensuring its responsible, rights-based and context-sensitive use.

The influence of AI is examined across three core domains of democratic practice: elections, public service delivery, and citizen engagement. In electoral contexts, AI poses significant new challenges for electoral integrity and public trust because it has lowered the cost and increased the scale of disinformation, impersonation and targeted influence. At the same time, AI has also been shown to support election administration, information management and threat detection when combined with governance frameworks and human oversight. In public service delivery, while AI offers potential efficiency gains and improved access, it also risks entrenching inequality and eroding accountability if deployed without transparency, contestability and clear institutional responsibility. In citizen engagement and dialogue work, a component of both democracy and peace process support, AI-powered tools have the potential to enable large-scale participation, deliberation and inclusion processes, even in fragile political settings, but only when designed to avoid manipulation, exclusion and algorithmic bias.

The report also examines the use of AI in peace processes, including conflict analysis, mediation support, ceasefire monitoring and post-conflict peacebuilding. In the preparatory phase of a peace process, AI has the potential to deepen conflict analysis by rapidly processing large volumes of unstructured data – news, social media, historical records, and diplomatic communications – to map conflict dynamics, identify key actors, trace grievance narratives, and flag potential spoilers. In mediation processes, AI can enhance cross-cultural communication by enabling real-time translation, sentiment analysis, and discourse mapping. AI can also broaden and strengthen ceasefire monitoring by enabling systematic observation where human access is limited. During post-conflict peacebuilding, AI can help to oversee and assess compliance data and detect emerging risks of renewed violence. Despite all these potential positive applications, AI should still be used with caution. AI should only be used to assist political judgement or trust-building, not to replace it. In peace contexts AI should be applied with ethical safeguards and contextual care to prevent skewed outcomes and undue stakeholder influence.

Drawing on global case studies and policy practice, this report concludes that AI should be treated as an enabling tool rather than a solution in itself. Across all domains, the key factor determining the positive impact of AI is not the technology itself, but the institutional capacity to govern it. In the future, the potential for AI to have a positive impact on peacebuilding will only happen by strengthening institutional capacity, embedding transparency and accountability in AI deployment, and ensuring that democratic and conflict sensitivity principles guide both technological design and political use.

01 Introduction

by Catalina Perdomo and Catalina Uribe Burcher

The world is experiencing rising political tensions and serious security challenges. The need to manage existing conflicts and prevent more from breaking out is greater than ever. Peaceful solutions and resilient democracies play a crucial role in conflict prevention and resolution. For the FBA, supporting peace processes and democratic governance are core mechanisms necessary for ensuring long-term conflict transformation.

"For the FBA, supporting peace processes and democratic governance are core mechanisms necessary for ensuring long-term conflict transformation."

AI is broadly defined as technology capable of performing complex tasks traditionally associated with human intelligence, such as visual perception, speech recognition, decision-making and language translation.[1] Recent advances in natural language processing, generative problem-solving models and multimodal systems have led to the emergence of sophisticated AI chatbots, realistic image and video generation, and enhanced decision-making tools. These developments have had an immediate effect on democracy building and peace processes. This report provides a structured overview of the role AI applications can play in conflict prevention and peacebuilding efforts, with a specific focus on their contribution to democratic resilience and peace process support.

As Stephen Hawking stated more than a decade ago, "the rise of powerful AI will be either the best, or the worst thing, ever to happen to humanity. We do not yet know which."[2] Since the advent of the internet, humanity has struggled to cope with the ever greater flow of information.[3] While AI represents one of the foremost technologies specifically designed to process and manage this vast volume of data, its widespread uptake has led to a new challenge: how to ensure that AI uses the information available to it – from social media outlets, satellite imagery, automated surveillance systems and countless other digital sources – in safe and ethical ways. Even though the unprecedented availability of data, coupled with the ability of AI to process this information on a large scale, has the potential to help humankind solve global problems, there is also the possibility that it will greatly disrupt human society.

To cite Hawking again, "alongside the benefits, AI will also bring dangers, like powerful autonomous weapons, or new ways for the few to oppress the many."[4] This is precisely what we are witnessing in several countries today, with AI fuelling attacks on democratic institutions and powering lethal drones in armed conflicts.[5]

Despite these dangers, AI is here to stay. Its application in all manner of contexts is growing rapidly. According to Stanford University's 2025 AI Index Report, the percentage of AI-enabled medical devices approved by the American Food and Drug Administration increased by 3,617 per cent between 2015 to 2023 – a 37-fold rise in less than a decade.[6] Given the rapid infiltration of AI in everyday life,[7] it seems fair to assume that this technology will only grow its influence. It will be the responsibility of humanity to ensure that AI is harnessed for good rather than evil.

Governments, civil society and the private sector now have a unique opportunity: to develop AI as a tool for understanding and solving many of the world's most complex problems.[8] It seems shortsighted to view AI solely through the lens of its destructive potential. With its ability to drive future innovation, AI is being hailed by some as the new electricity, as the power underpinning a new technological revolution.[9]

One of the most enduring challenges confronting humanity is the effort to live peacefully together. For centuries, humans have sought to solve and prevent wars, to ensure peaceful coexistence and governance, and to resolve conflicts without bloodshed. Central to this endeavour is the question of how conflict can be prevented through democratic resilience and strengthening and how societies can resolve conflicts through dialogue instead of weapons. Of course, different tools and technologies have been identified as useful in the effort to maintain and achieve peace. Given the rapid uptake of AI across so many different contexts, the idea that it might be a helpful tool in this endeavour has started to be raised.

"the research into AI demonstrates that it presents an opportunity to develop more effective and context-sensitive responses to conflict."

While research exploring the connections between AI, peacebuilding and conflict prevention is relatively new, especially when compared to the body of scholarly work investigating AI in military defence, armed conflict and war,[10] the field is growing rapidly. The existing research examines familiar themes from the prevention and peacebuilding literature as well as newer, more critical perspectives.[11] Understood as a complementary tool within the broader concept of "PeaceTech", the research into AI demonstrates that it presents an opportunity to develop more effective and context-sensitive responses to conflict. What does the research say about the opportunities, challenges and implications of using AI for peace?

PeaceTech is broadly understood as the use of digital innovation to support peace-building such as, for example, the use of SMS text messaging to survey a population with little internet connectivity.[12] AI can be used to strengthen digital resilience, support humanitarian operations, enhance access to reliable information, and enable more inclusive and de-centralized approaches to conflict prevention and resolution.[13]

More critical perspectives have also emerged from the research into AI and peacebuilding. These aim to point out that AI's inclusion into the global international system presents significant challenges for peacebuilding efforts. For instance, the way these technologies have been developed has given them inherent limitations and certain in-built biases.[14] Generative AI, a type of AI that is designed to create content like images and texts or to replicate natural language forms for interaction with humans, such as chatbots, are trained on vast amounts of data drawn from numerous online sources. These technologies have access to far more information than any individual could realistically process and may therefore exhibit fewer individual biases. However, because much of this data originates disproportionately from the "West", it tends to more strongly reflect Western perspectives and realities. Some AI companies have sought to mitigate such biases by introducing corrective mechanisms or mandates within their systems, although with mixed results to date.[15]

Other research has pointed out that, even though AI may help mitigate informational uncertainties for those with the resources to leverage its potential, for those without such resources, their exclusion is likely to deepen. The growing use of AI in military contexts, for example, such as in drone operations, has gradually increased the reliance on automated systems. This shift often prioritizes technical efficiency over ethical considerations, making it harder to acknowledge the broader political and security issues that might require a reassessment of the way AI technologies are deployed.[16] This is why the use of AI in peacebuilding and conflict prevention poses a significant ethical challenge that demands a human-centred and accountable approach.[17]

This report seeks to expand the field of vision surrounding the use of AI in conflict prevention and peacebuilding. It is also intended to serve as a starting point for the FBA to identify the challenges, opportunities and risks associated with using AI in its conflict prevention and peacebuilding work. As mentioned above, this report aims to provide a structured overview of the use of AI applications in conflict prevention and peacebuilding efforts, with a specific focus on their role in supporting, specifically, democracy and peace processes. Addressing these two themes within a single, integrated report achieves two important goals: one, it investigates two areas of work that are central to the FBA's mission and two, it enables the identification of possible linkages between the use of AI to support democracy and peace processes that may be overlooked if the areas are addressed separately. This integrated approach makes it possible, for example, to identify AI technologies that are relevant to both fields and point out the risks that AI presents to both democracy assistance and peace process support alike.

This report explores AI usage within three core domains of democratic practice – elections, public service delivery and citizen engagement – and examines its use in peace processes by focusing on dialogue and mediation, social media monitoring and ceasefire monitoring. It also presents comparative examples from around the world involving both technologically advanced and resource-constrained contexts, with case studies presented from Chile, India, Israel, Kenya, Moldova, Palestine, Somalia, Sudan and Ukraine.

In addition, this publication aims to serve as a reference point for Swedish and international policy discussions, to inform the FBA's programmatic engagement in the period 2025–2027 and beyond, and to support evidence-based dialogue on the implications of AI for democracy and peace process support. This report is a collective effort authored by multiple contributors, so some variation will be evident in analytical approach and writing style. This reflects the authors' diverse perspectives and work experiences. While the text should not be read as a fully homogeneous narrative, the contributions are united by a common purpose and shared analytical ambition.

This publication represents a first step in FBA's broader process of considering how to approach AI within its conflict prevention and peacebuilding mandate. It also seeks to complement the Swedish government's efforts to identify measures that ensure the safe and reliable use of AI, in line with AI regulations initiated by the European Union.[18] As part of its evidence base, this report also draws on material from two workshops, held in Stockholm in 2025, that involved the participation of academics, experts, civil society organizations, and the private sector.

Chapter 2 in this report explores AI within the three core domains of democratic practice: elections, public service delivery and citizen engagement. Chapter 3 examines AI applications in peace processes, with a focus on conflict analysis, dialogue and mediation, ceasefire monitoring and post-conflict peacebuilding. With this evidence in mind, Chapter 4 presents the main conclusions and makes a number of recommendations, aimed at both democracy and peacebuilding practitioners.

02 Democratic Resilience in the Age of Advanced AI

by Slava Jankin

Overview

by Catalina Uribe Burcher

AI has the potential to affect – both positively and negatively – democratic political systems around the world. This power has become an obvious subject of discussion, with government policymakers and concerned practitioners seeking to understand its long-term impact. To understand the AI trend it is essential, first, to take stock of the current state of democracy and technology more broadly and the ways in which these two concepts interact. As a working definition, democracy is understood here as a system of governance grounded in popular control over public decision-making and political equality among citizens. It encompasses both electoral processes and the broader rule of law, including the human rights and accountability structures that sustain them.[19]

One of the key trends within political research circles is the growing number of threats to democracy, in particular the notion of democratic backsliding. International IDEA's 2024 report reviewing the global state of democracy shows that, through its measurement of different democratic indices, democratic practice is relatively strong. The year 2024, for example, was an election "super-cycle", when around three billion people across the world had the opportunity to vote, more than at any other time in history.[20] At the same time, there has been a visible decline in the quality of elections in many countries. This is reflected in decreasing voter confidence, which has fallen in more than one fifth of the countries covered by International IDEA's indices, and in declining voter turnout.[21] The global average for electoral participation dropped from 65 to 55 per cent over the past 15 years, while protests and riots linked to elections have become more frequent.[22]

According to the same indices, electoral disputes have also become more frequent. Between mid-2020 and mid-2024, one in five elections was challenged in at least one legal proceeding, one in five saw the losing presidential candidate or political party publicly reject the outcome, and one in ten was boycotted by opposition parties.[23] Overall, countries experiencing a decline in their democratic performance outnumbered those who demonstrated progress, with only one in four countries improving their democratic performance while four out of every nine were revealed to be democratically worse off.[24]

The growing instability of democracy in some contexts has taken place at the same time as key changes have occurred within the field of technology, particularly the introduction of AI. That each change has influenced the other seems likely, with hybrid threats posing a considerable risk to political stability. Among the most pressing challenges are cyberattacks and the political polarisation fuelled in part by hate speech, misinformation and disinformation. For instance, in 2024 a deepfake audio recording impersonating Dmytro Kuleba, the former Ukrainian foreign minister,

was used to target the Chair of the United States Senate Foreign Relations Committee, Benjamin Cardin, in a fake phone call.[25] Similarly, in 2025, an audio deepfake of the US Secretary of State Marco Rubio was used to contact foreign ministers and other politicians.[26] The fact that both recordings successfully convinced the targets (at least temporarily) that they were dealing with the real Dmytro Kuleba and Marco Rubio, demonstrates how good AI systems have become at deceiving their audiences, especially in their use of audio tools to impersonate public figures.

Because most social media platforms are driven by algorithms that privilege engaging content, the use of bots and other tools to manipulate attention and sentiment often result in the amplification of extreme material.[27] This louder tone, in turn, incentivises the online harassment of many different vulnerable groups. Digital violence against women in politics, for instance, is becoming increasingly common.[28] In this sense, technology can be viewed as contributing to the erosion of a country's social fabric as well as to the manipulation of political engagement and decision making.[29]

The experience of Myanmar and the role that hate speech on Facebook played in the Rohingya exodus of 2016 and 2017 illustrates the risks associated with limited monitoring and moderation of online content.[30] In a more recent example, intelligence officials in the United States reported that Russia and Iran had used AI to generate content aimed at destabilising the 2024 presidential election.[31] With the advances in technology that have taken place since the mid-2010s, the latter example shows how disruptive actors are now able to produce more sophisticated and better-targeted material at greater speeds and larger scales than ever before.

Beyond specific incidents of intentional disruption, AI has also had a more general impact on public trust and sentiment surrounding democratic elections. A Pew survey from 2024 found that 39 per cent of Americans believed that during the upcoming presidential campaign AI would be used mainly for harmful purposes while 57 per cent of adults from both major parties expressed serious concern about AI's potential to influence the election and spread false or misleading information about it.[32] Another telling case occurred during the 2024 Indian elections, when AI was used to create many deepfakes of party candidates and politicians, which then led to extensive online trolling and disinformation (see section on AI and elections below).

"AI, a technology that works even faster and generates even more results, risks intensifying these power imbalances and marginalising these already technologically marginalised groups still further."

Another challenge that technology poses to democratic practice relates to the question of accessibility. There is robust evidence showing that women, rural communities and persons with disabilities already face "data asymmetries": a comparative disadvantage in accessing and benefiting from technology and technological benefits.[33] AI, a technology that works even faster and generates even more results, risks intensifying these power imbalances and marginalising these already technologically marginalised groups still further.

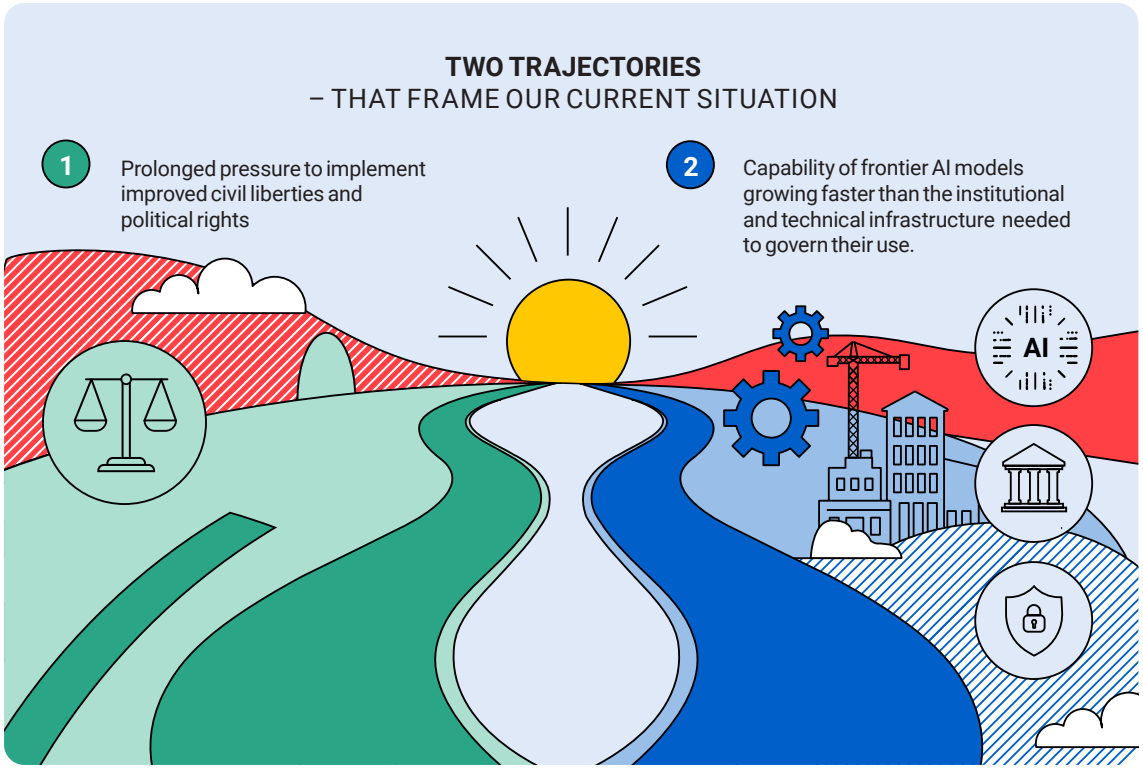
Another concern that the research has identified in relation to AI has to do with the invasion of privacy arising from the potential misuse of surveillance and censorship systems within various domains, from law enforcement (e.g., predictive policing and border control management) to social media monitoring of political dissent. The use of such observational systems raises multiple questions about regulation, control and accountability. The growing use of AI by security forces to operate unmanned vehicles, such as drones, is also prompting debate about the democratic oversight of AI usage in armed conflicts. The threat to democratic accountability emerges because with AI, human actors can be removed from the chain of responsibility in a variety of public institutions, not least the military and security sectors.[34] The anxiety surrounding the full automation that AI promises is only heightened by the awareness of AI's potential for hallucinations and the doubt this casts on its reliability. Although some AI systems have reduced their hallucinogenic tendencies, such errors remain a persistent problem, especially when decisions based on AI outputs may affect large segments of the population, such as in the case of public decision making.

Despite these and other persistent challenges, digitalisation and AI also offer democracies significant opportunities. For instance, there is growing interest in applying AI and machine learning to improve public service delivery. These tools are also being explored to facilitate dialogue between citizens and the state,[35] an approach that holds particular promise for mediation and constitution-building processes. Another promising area concerns transparency and access to information. For instance, some election management bodies have begun using AI to process large datasets and generate rapid data analysis, while others employ it to engage voters and conduct education campaigns.[36] During the 2024 elections in India, for instance, AI was used positively by parties and candidates to reach out to remote communities, through real-time content translation and other campaign activities that would otherwise have been too costly or cumbersome to undertake manually. These and other benefits of AI will be highlighted throughout this chapter.

Considering the risks and opportunities presented by AI requires recognizing that digital tools are now ubiquitous. The debate is therefore no longer about whether they should exist, but about how they can be designed and used to minimise their associated risks – such as hate speech, misinformation and disinformation – while safeguarding online freedom of expression and digital rights and, ideally, strengthening social cohesion.[37]

This chapter examines how advanced AI affects three democratic domains – elections, public service delivery, and citizen engagement – and what public institutions can do to reduce risk while maintaining democratic control. It focuses on practical governance levers: information integrity, institutional capacity, market incentives, and the security context.

Two trajectories frame our current situation. First, many countries have experienced prolonged pressure to implement improved civil liberties and political rights.[38] Second, the capability of frontier AI models has grown faster than the institutional and technical infrastructure needed to govern their use.[39]



"By specifying public-interest requirements, leveraging procurement, governing deployment, monitoring real-world effects and revising systems over time in line with democratic goals, public institutions ought to be able to retain the agency necessary to shape how AI develops and is used."

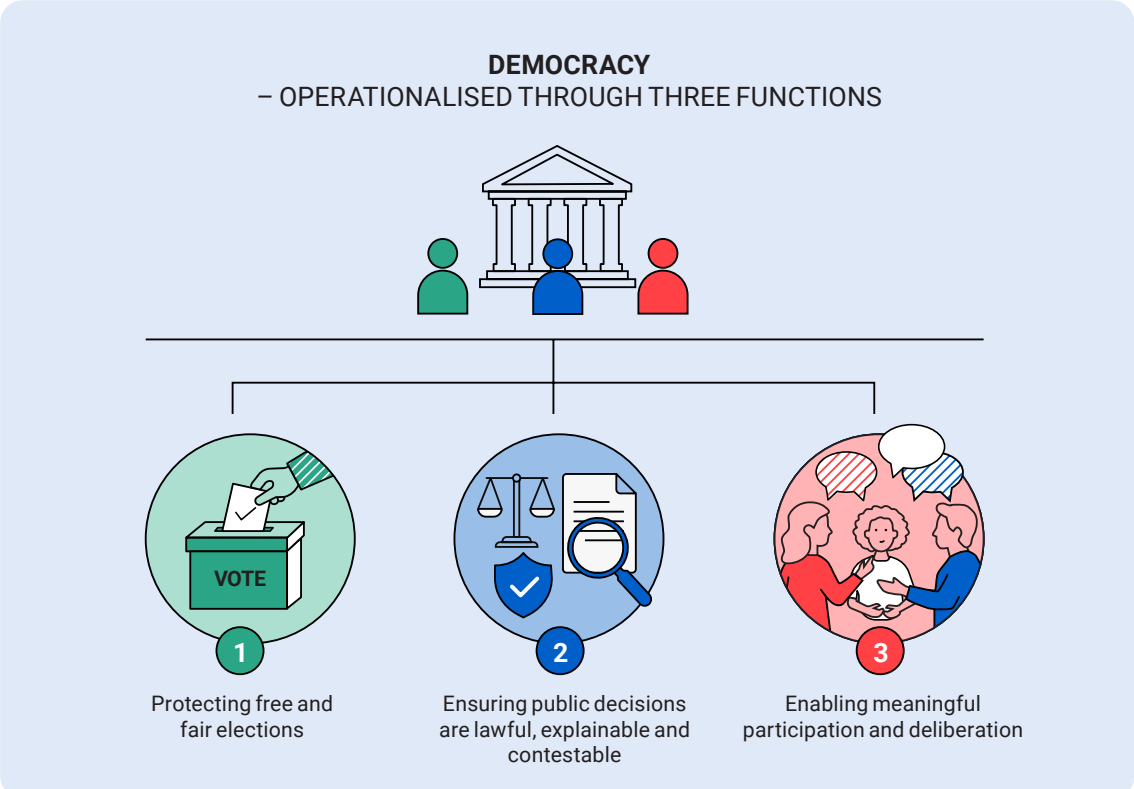
Taken together, these trends create powerful political and technological pressures on democratic governance. Yet these trends are not inevitable. By specifying public-interest requirements, leveraging procurement, governing deployment, monitoring real-world effects and revising systems over time in line with democratic goals, public institutions ought to be able to retain the agency necessary to shape how AI develops and is used.[40]

This public ability, or institutional capacity, depends on the presence of a small set of basic tools. To manage AI, public institutions should ensure their relationship with vendors gives them named decision rights, usable data and logs, evaluation and assurance routines, and enforceable contractual levers. When organisations lack any of these elements, their ability to govern AI fails, regardless of good intentions or technical sophistication.

If organisations need to have certain capacities in order to manage AI properly, a diagnostic survey can be a useful tool for determining its current capacity. A proper capacity diagnostic can identify whether an organisation has named roles with clear decision rights, established procedures for handling incidents and appeals, maintained queryable logs of decisions and outcomes, deployed appropriate verification and evaluation tools, secured contracts with documentation and audit rights, obtained a legal mandate with a sufficient budget, and planned concrete change actions, including executive sponsorship and role redesign.[41] If an organisation is thinking to build this capacity, it should first confirm that AI is the appropriate solution for their problem. This means testing whether the data are representative, the potential harms are reversible, if simpler alternatives exist, and the costs of monitoring and redress are proportionate to the expected benefits.[42] This appropriateness check prevents organisations from automating poor decisions and anchors proportionality in real constraints.

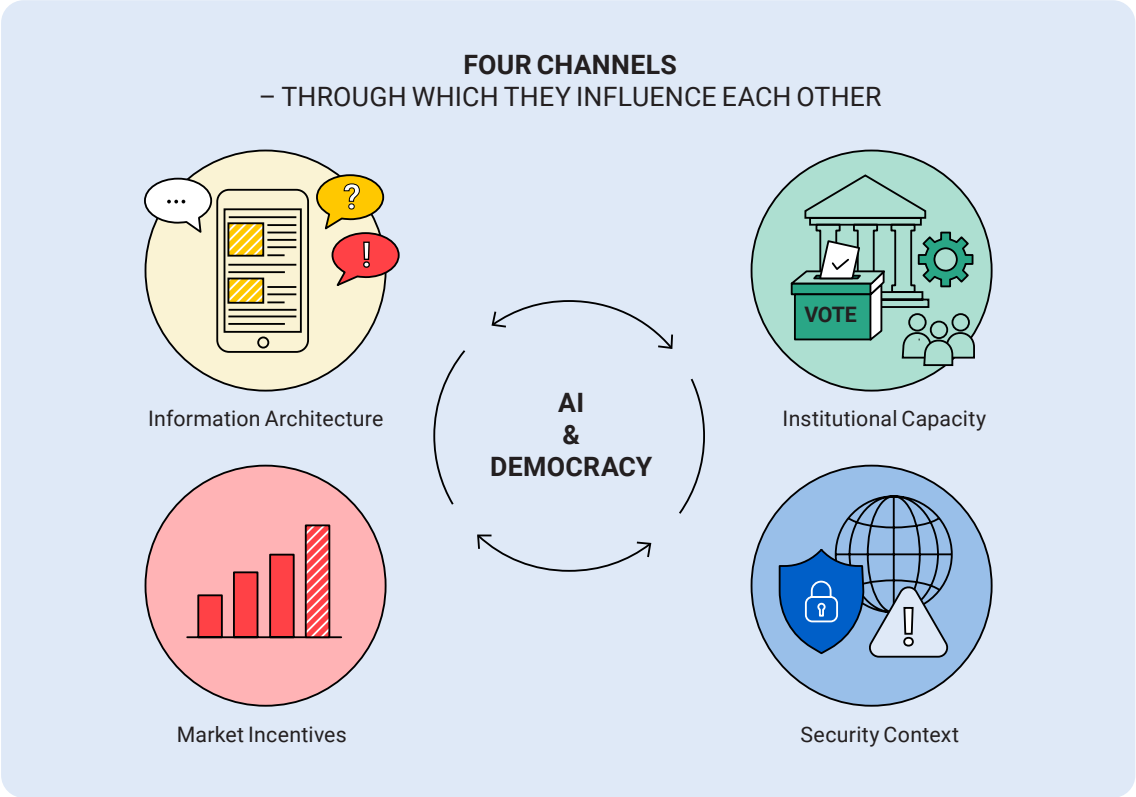
2.1. Analytical Framework

This chapter is written for practitioners responsible for democratic functions: election management bodies (EMBs), public agencies delivering citizen-facing services, and organisations running citizen-engagement processes. For the purposes of this chapter, "democracy" is operationalised through three functions: (1) protecting free and fair elections; (2) ensuring public decisions are lawful, explainable and contestable; and (3) enabling meaningful participation and deliberation.



The framework outlined here is an organising device (a way of synthesising the literature cited in this chapter) intended to display the main pathways, or "channels", through which AI changes these democratic functions and through which democratic functions shape AI. The chapter concludes by translating these channels into a short, phased checklist.

Four channels can be said to reflect the influence that AI and democracy have on each other. Information architecture (the systems that generate, rank, and distribute content) affects who is heard, what is seen, and what is trusted. Institutional capacity determines the extent to which public bodies can choose appropriate uses of AI, contract responsibly, and maintain oversight over time. Market incentives shape the nature of what is built and the purpose for which it is optimised (for example, engagement at scale versus public value). The security context determines both threats (operational influences, intrusions) and defences (detection capability, response procedures, international norms).



To analyse the relationship between AI and democracy, the remainder of this chapter is divided into sections covering elections, citizen engagement, and public service delivery. Within each section, the risks AI presents (and the ways those risks can be mitigated) are linked to at least one of the four analytical channels, so that the chapter avoids becoming a disconnected list of concerns.

2.2. Elections and Electoral Integrity

Risks: How AI Threatens Free Electoral Processes

Generative models have transformed the economics of manipulation. Convincing multimedia fakes are now cheap and scalable. During the war in Ukraine, a deepfake video of President Zelensky urging surrender spread widely across the internet before it was debunked. The feasibility of such attacks determines the importance of a rapid response. Manipulative tactics driven by AI are now migrating into electoral contexts worldwide.[43] In democratic elections the core test is whether the manipulative potential of AI undermines freedom (citizens' ability to form and express political preferences without deception, coercion or intimidation) and fairness (the presence of a level playing field and an impartial, competent election administration).

"To date, AI-enabled election threats have manifested in four distinct areas: synthetic media and impersonation; targeted persuasion and harassment at scale; procedural sabotage of election workflows; and abuse of communications/platform infrastructure that disrupts information access and enforcement."

To date, AI-enabled election threats have manifested in four distinct areas: synthetic media and impersonation; targeted persuasion and harassment at scale; procedural sabotage of election workflows; and abuse of communications/platform infrastructure that disrupts information access and enforcement. Beyond crude media fakes, influence operations now use micro-targeting (directing tailored messages at narrowly defined groups), combined with pre-tested narratives, to increase message precision and effectiveness. Media ecosystem research has shown how AI tools can create an environment of asymmetric amplification (where platform ranking and sharing mechanisms can be influenced to push some messages much more frequently than comparable alternatives) and to deepen polarisation (where message content can be used to channel consumers more frequently into one-sided, mutually hostile narratives and thus weaken shared factual baselines).[44]

The abuse of communications infrastructure presents another risk to democracy. In January 2024, for example, AI-generated robocalls in New Hampshire spoofed a national leader's voice to suppress turnout. Regulators in turn imposed significant fines and restrictions on AI voice robocalls, while prosecutors charged the consultant responsible,[45] thus making the point that the enforcement response matters as much as the initial violation. These incidents also expose a recurring responsibility-chain problem. In AI-enabled election interference, accountability is often widely distributed: across campaigns or contractors (who commission content), model/tool providers (who supply capability), platforms and telecoms (who distribute it), and public authorities (who must create regulations and enforce them). When responsibility is diffuse, response can be slow and punishments can be contested. EMBs can reduce this risk by pre-defining their oversight actions (detection, verification, public correction, takedown/label decision, and legal referral), and by publishing these escalation pathways in advance so citizens and parties can see how infringements will be handled. When the European Commission passed the Digital Services Act enabling them to investigate TikTok over alleged election-related systemic-risk failures in Romania, it signalled that platform due diligence has become an election-integrity lever with real penalties.[46]

Opportunities: AI Enhancing Electoral Democracy

Despite these risks, AI also offers tools to strengthen electoral integrity. France's VIGINUM, the national agency established in 2021 to detect, analyse and counter foreign digital information manipulation and interference, demonstrates how democracies can build defensive capabilities by integrating three complementary approaches: open-source intelligence to collect and verify publicly available digital content; platform monitoring to track the spread, amplification and coordination of narratives across social media and online services; and network analysis to map relationships between accounts, content and infrastructure in order to identify coordinated manipulation and inauthentic behaviour. This approach offers a template for democratic information defence that respects civil liberties while identifying foreign interference.[47]

In terms of electoral administration, properly constrained AI assistants can improve voter services and reduce clerical burdens. When transparency and human oversight are built into design, AI systems can help with registration, ballot processing, and accessibility without introducing political bias. The gains are operational rather than political and the objective is efficiency and access rather than influence.[48]

Democracy Shaping Electoral AI

Democratic institutions are beginning to assert control over how AI is used in the context of elections (or electoral AI) through multiple channels. For example, the EU Political Advertising Regulation imposes transparency requirements on targeting and sponsorship. Meanwhile, technical standards such as the C2PA specification – an open technical standard from the Coalition for Content Provenance and Authenticity – allow creators to tag their digital content with secure metadata that describes how it was made and edited. This enables the use of cryptographic provenance signals: tamper-resistant digital markers that verify the origin and modification history of a piece of content. These markers, in turn, help to improve the systems and tools platforms use to detect manipulated or harmful content.[49]

"The most effective approach to managing AI combines legal transparency mandates with procurement requirements for provenance and detection capabilities to create a market incentive for responsible development."

The most effective approach to managing AI combines legal transparency mandates with procurement requirements for provenance and detection capabilities to create a market incentive for responsible development. Systematic record keeping also provides an accountability trail. For example, whenever an EMB requests an online platform or media outlet to remove content, it should keep a record of the incident and note the outcome. These records could follow a format similar to the Ad Transparency Reporting Systems (ATRS) that some platforms use to log, review and publicly report on political advertising. This helps build institutional memory about digital threats and official responses.[50]

Building the capacity to manage potential digital threats requires more than technical tools. EMBs need the literacy skills necessary to monitor synthetic-media detectors (tools used to identify artificially generated content), understand detector limitations, and coordinate a rapid response to infringements across regulators, platforms, and law enforcement. This starts with naming a senior responsible owner, developing and implementing clear and responsible plans with step-by-step guidelines showing how the organisation should respond to a specific type of crisis or attack and outlining ways to integrate AI scenarios into existing cyber exercises.[51]

Who is responsible and what criteria can be used to measure the success of any regulatory framework? Primary owners include the EMB, the courts, internet service providers, platforms, and independent monitors. Success should be measured through median time-to-takedown for verified synthetic or materially misleading content made public during pre-determined periods.

The example of AI and elections in India

India's 2024 national election is possibly the first true "AI election", with AI tools being used by political campaigns, disinformation producers, and EMBs. Political organizations used AI to create personalized campaign videos, consensually-generated candidate content, information chatbots, and more.[52] While the research suggests that AI-generated misinformation made up only a small proportion of the total viral election content circulating in India in 2024,[53] there were some prominent examples of disinformation and misinformation, including deepfaked videos of candidates and deep faked audio imposed over real video to suggest political events that never happened. AI-generated misinformation was also used to exacerbate existing religious tensions.[54]

However, AI was also used to help administer the 2024 elections in India. The nation-wide ERONET voter list management tool used AI to support the detection of duplicate registrations.[55] Various states have partnered with developers to build internal "Election GPT" chatbots that can answer election workers' questions.[56] While the benefits of using AI to improve two of the most time-consuming aspects of electoral management – roll maintenance and training – are substantial, AI hallucinations and the consequences this inaccuracy might cause, is still likely to harm electoral integrity.

The example of AI and elections in Moldova
During the 2024 presidential election in Moldova, Osavul – a technology company that offers AI-driven solutions for pre-emptive situational awareness for security missions – supported authorities to deploy its AI-powered solutions to monitor, analyse, and respond to complex information threats targeting the country's pro-European government. The platform's capabilities enabled real-time detection of disinformation campaigns coordinated by pro-Russian actors, which sought to discredit reforms, weaken trust in institutions, and polarize society.

Osavul applied two of its core tools: the Information Threat Detection System (Early Warning System) and the Topic Modelling and Sentiment Intelligence Engine. The Early Warning System provided alerts on emerging threats that, coupled with a preliminary analysis of malicious actors, allowed government institutions to respond quickly. The Topic Modelling and Sentiment Intelligence Engine clustered public discourse into thematic blocks, tracked emotional tone, and correlated sentiment with key political and geopolitical events. Together, these tools offered a Common Operating Picture of Moldova's information environment, helping authorities differentiate between legitimate criticism and coordinated manipulation, anticipate narrative escalations, and formulate timely strategic responses.

During the election, authorities using Osavul's system identified coordinated cross-platform activity, spikes in engagement coinciding with policy announcements, inauthentic behaviour in high-frequency accounts, and the narrative laundering of discourses from fringe portals into mainstream anti-EU rhetoric. The platform mapped entire disinformation networks, revealing how false narratives about censorship and alleged dictatorship were propagated. It also highlighted manipulation techniques, including rallies, symbolic actions, and video-based content.

2.3. Public Service Delivery and E-governance

Risks: When Automated Government Fails
In the Netherlands, between 2013 and 2019, an automated risk system wrongly accused thousands of families – many with migrant backgrounds – of financial misdoing and forced them to pay back their state benefits. The harm inflicted was considerable, including wider discriminatory outcomes. Known as the Dutch childcare benefits scandal, it remains the starkest warning of automated government gone wrong when combined with poor governance and weak redress mechanisms.[57] The fundamental problem in the Dutch case was not the technology but the absence of sufficient documentation, monitoring, and contestability. When automated screening operates in welfare and tax contexts without administrative safeguards, it can systematically misallocate harm to those least able to fight back.[58]

Relying too heavily on automated systems also comes with more subtle risks. If staff cannot override automated decisions, important human contextual judgement is lost. Service workers default to a "computer says no" attitude that in turn degrades both service quality and public trust. In the United Kingdom, public sector guidelines address this problem directly, clearly stating that AI is never a standalone solution. Teams must confirm that their problem genuinely warrants AI, that their data are fit for purpose, and that any potential harms are reversible before proceeding with any build or procurement process.[59]

Opportunities: AI-enabled Services

When properly designed and governed, AI can significantly improve public services. Estonia's Bürokratt initiative, for example, provides government website users with a single digital assistant that works across all government agencies. It provides multilingual access and organises its service delivery model around real-life moments, such as having a child, changing address or starting a company. Bürokratt is essentially a government-owned virtual assistant framework that allows citizens to interact with different public services through a single, AI-enabled interface. However, even service-oriented AI can concentrate benefits unevenly (for example, towards digitally confident groups) or be misused without clear access safeguards, auditability, and redress. However, the base intention with service AI is deliberately open and modular. The object is not to create a monolithic system but to promote interoperable integration – that is, ensuring that systems from different agencies can securely "talk to each other" and exchange data.[60] In governance terms, interoperability and modularity matter because they reduce single-vendor dependence, make it easier to apply consistent logging and audit controls across agencies, and support contestability by allowing decisions to be traced across service workflows.

More modest applications of AI to government services can also deliver value. Triage, summarisation, translation, and document extraction can substantially reduce administrative backlogs when deployed through measured pilots with appropriate guardrails and continuous outcome monitoring. The key is to use systems that use machines to handle the routine work so that people can make the important decisions.[61]

"The key is to use systems that use machines to handle the routine work so that people can make the important decisions."

Governing Service AI

Democratic oversight of service AI requires multiple interventions working in concert. Canada's Directive on Automated Decision-Making mandates Algorithmic Impact Assessments proportionate to risk level. Government agencies are obliged to consider the consequences the decision to use AI can have on people's rights, benefits, or legal status, especially what might happen if the systems malfunction or are biased. To use higher risk systems, the requirements for transparency and human review are escalated.[62] The UK's Algorithmic Transparency Recording Standard provides a complementary approach, offering a standardised template for publishing system records that enables both accountability and learning across government departments.[63]

Documentation standards help to surface limitations, intended uses, and performance trade-offs that might otherwise remain hidden. One standard is Model Cards, a short description of an AI model's intended uses, limitations, performance across different groups, and ethical implications.

Another is Datasheets for Datasets, a document used when training data that describes a dataset's origin, composition, collection process, recommended uses, and known risks or biases.[64] The Alan Turing Institute's ethics guidance provides another useful framework for process-based governance that focuses not only on outcomes but also on how AI systems are designed, developed, deployed and monitored. The guidelines are intended to help maintain proportionality and clarity about the impact an AI model has throughout the algorithmic system lifecycle.[65] Together, these instruments create multiple checkpoints that can prevent the kind of cascading failure seen in the Dutch case.

Who is responsible for government AI and what criteria can be used to measure its successful use? Front-line responsibility should be owned by the deploying agency (operational responsibility) and overseen by its supervising ministry and the government's independent audit function (assurance). Accountability should be supported by the data protection authority (rights compliance), with suppliers contractually bound to provide logs, documentation, and incident reporting.

The example of AI and public service delivery in Chile
In 2020, the Chilean Transparency Council investigated the use of AI in the Chilean public sector. The inquiry revealed that AI was widely employed in both strategic and support processes, but often without sufficient transparency or accountability. This opacity threatened public trust – an essential pillar of democracy – particularly as Chile's constitution mandates transparency in public decision-making.

To address these challenges the Chilean government, in line with the country's national AI policy, issued guidelines for the public use of AI in December 2023 (Circular No. 711). These guidelines emphasised human-centred AI, transparency, accountability and data protection.

Following a broad public consultation process, in August 2024 the Chilean Transparency Council issued Resolution No. 372, Latin America's first recommendation on algorithmic transparency in the public sector. Its aim was to regulate the functioning of AI systems in public service through general and specific guidelines. The general guidelines require public bodies to provide clear and accessible information about their AI systems and transparent contracts, with secrecy justifying some limited exceptions. The specific guidelines apply to systems that affect fundamental rights or social benefits and are organised around three pillars: proactive transparency, which requires agencies to publish details on the objectives, operations, data and outcomes of AI systems; accountability, which ensures that citizens can challenge opaque systems; and ethical, human-centred AI. In 2025, the Chilean Transparency Council published a Q&A guide to support agencies in putting these standards into practice.

These measures are intended to strengthen public trust, prevent unfair outcomes and protect fundamental rights, all while safeguarding the right to data privacy and confidentiality guaranteed by law. By embedding transparency into the use of AI, the Chilean Transparency Council seeks to align public sector innovation with democratic values and to set a regional precedent for responsible AI governance.

2.4. Citizen Engagement and Democratic Participation

"these dynamics can create polarisation spirals and asymmetric amplification, allowing extreme voices to dominate while moderate perspectives disappear."

Risks: Fragmented Public Discourse

Platform algorithms optimised for engagement systematically privilege sensational content over substantive deliberation. Ecosystem studies that look at the media and information environment as a whole, rather than isolating individual platforms or actors, have documented how these dynamics can create polarisation spirals and asymmetric amplification, allowing extreme voices to dominate while moderate perspectives disappear.[66] The "information disorder" literature warns against treating this as a single "fake news" problem, framing it instead as a spectrum of disorders in the information ecosystem and advocating for targeted interventions addressing specific pathologies. Misinformation describes false or misleading information shared without intent to deceive or cause harm. Disinformation refers to false information created or shared with the intent to deceive or manipulate. Malinformation, on the other hand, describes genuine information shared to cause harm, often by taking it out of context or exposing private details.[67] Other pathologies that AI may exacerbate include astroturfing (fake grassroots campaigns), agenda flooding (drowning out genuine discourse), and systematic exclusion through linguistic and digital barriers. These techniques can simulate democratic participation while actually undermining it, creating a facade of engagement that masks genuine public sentiment.

Opportunities: AI Enabling Mass Deliberation and Transparency

In contrast to the disruption AI can cause in public discourse, Taiwan's vTaiwan platform provides one example of how AI can enable genuine large-scale deliberation. Launched in 2014, this civic-technology platform enables both online and offline deliberation and consensus building. For example, vTaiwan hosts Polis, an online consensus-finding tool that allows participants to post short statements and vote agree/disagree and uses AI clustering algorithms to group respondents by patterns of opinion and reveal one or more broadly supported "consensus statements". Participants can also take part in facilitated offline resolution opportunities to help them work through their remaining disagreements. The vTaiwan system thus helps groups move beyond binary yes/no conflicts toward more nuanced solutions. The approach has successfully resolved contentious public policy issues, from Uber regulation to alcohol sales policy.[68]

AI tools can offer similar benefits to legislative transparency. For example, Slovenia's Parlameter is a civic-tech platform, developed by the NGO Danes je nov dan ("Today is a new day"), that can transform parliamentary transcripts and voting records into navigable data, enabling media and citizens to track the positions of their representatives and hold public officials accountable.[69]

Another example of AI used for deliberation and transparency is Decidim, an open-source digital democracy platform originally developed by the city of Barcelona to run public consultations, participatory budgeting, online petitions, surveys and deliberations. Decidim provides a participatory infrastructure that allows democratic processes to be delivered at a much larger scale, to reach more participants and handle higher volumes of engagement, all while preserving institutional control and auditability. AI can also be used to confront linguistic exclusion. Masakhane is a pan-African research collective that works to build open, AI-based technologies that seeks to expand linguistic inclusion for many communities and thus addressing a fundamental prerequisite for equitable engagement in multilingual societies.[70]

Democracy Shaping Engagement AI

Public investment strategies can steer engagement AI toward democratic ends. Funding civic tech as a public good, requiring open interfaces and, where feasible, mandating open-source infrastructure can prevent vendor lock-in while enabling innovation. When an organisation becomes so dependent on a particular supplier's technology, software, or services it can be very difficult or costly to switch to another provider and thus reduce agility and the growth of new ideas. The UK's Introduction to AI Assurance policy (February 2024), which provides standards to verify, validate and evaluate AI systems to ensure they are safe, ethical, compliant and perform as intended, shows organisations how to test the trustworthiness of AI systems, a process known as assurance. Evaluating, documenting and communicating an AI system can provide a framework for the testing, inspection, and independent assessment of public-facing systems, which can in turn help to ensure the system serves democratic rather than commercial imperatives.[71]

Who is responsible for these assessments and what criteria can be used to measure their success in providing the specific service and that they are safe, ethical and compliant? Ownership typically spans parliaments or line ministries, municipalities, platform teams, civil society partners, and independent auditors. Success should be measured through the share of proposals achieving broad rather than narrow support, using transparent thresholds that prevent gaming behaviours.

The example of AI and citizen engagement in Somalia, Sudan and Kenya
Build Up is a peacebuilding organisation active across Somalia, Sudan and Kenya that leverages technology, the arts and research to transform conflict, foster dialogue and strengthen inclusive, participatory approaches to peace. It has piloted peace-building initiatives where AI has been harnessed in innovative ways to expand citizen engagement, particularly where traditional participation approaches face constraints. Build Up has sponsored different initiatives in each of the three partner countries. These cases, cited below, are not presented as examples of consolidated democratic behaviour. They are intended to illustrate how AI-enabled tools can create a channel for citizen voices and for structured consultation in conflict-affected contexts where conventional participation is constrained and where legitimacy depends on transparency about how inputs are collected, clustered and synthesised. Each case illustrates a distinct methodology for making democratic processes more inclusive and responsive.

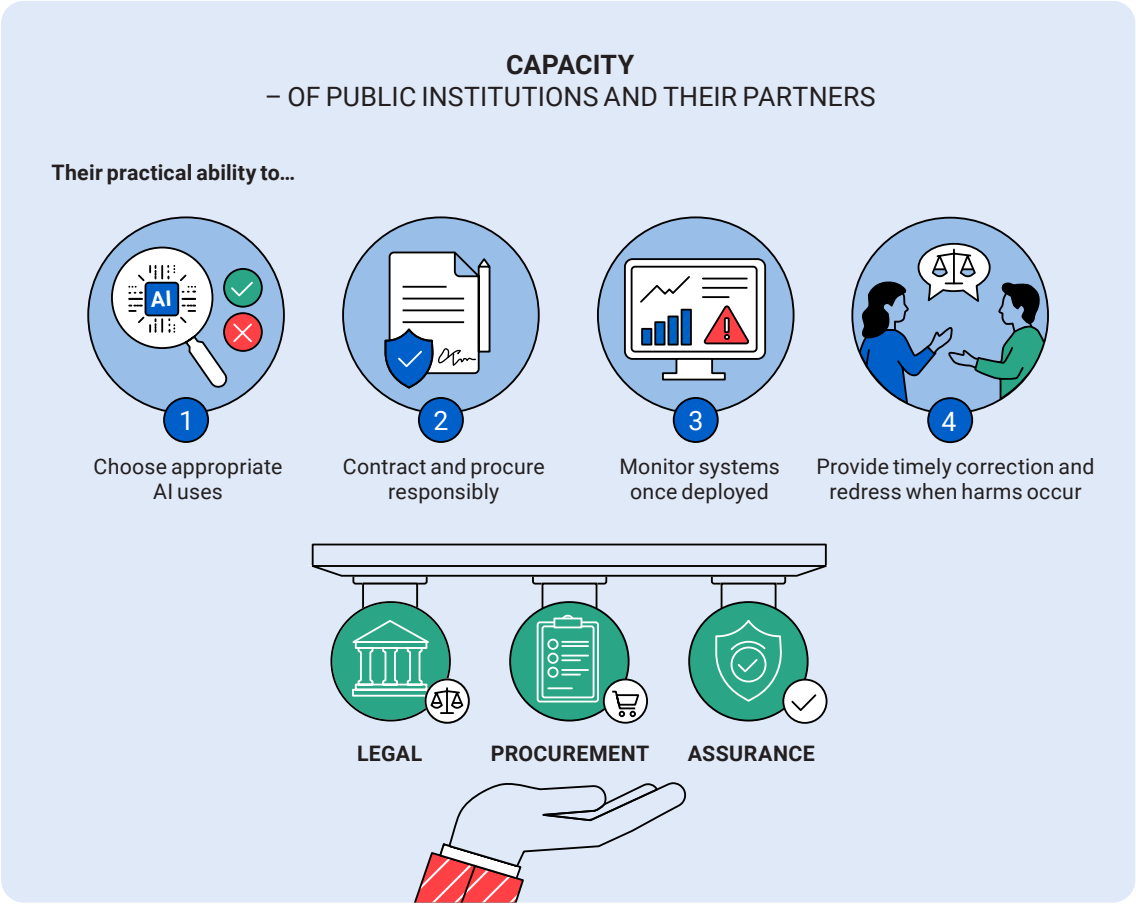
In Somalia, Build Up piloted a WhatsApp chatbot linked to an AI agent that simplified access to the HeshiisBeeg peace barometer: survey data that captures community perceptions of peace, safety and trust in institutions. Available in both Somali and English, the agent allowed citizens and stakeholders to explore complex survey findings through natural, chat-based conversations. Combined with Pol.is, an online tool that maps areas of agreement and disagreement by clustering participants' responses and mapping consensus, the approach transformed technical data into accessible, locally owned insights that reflected genuine community perspectives.

Despite the ongoing conflict in Sudan, AI-assisted platforms have made large-scale public consultations possible. Through AI, citizens have been able to submit voice comments in their local dialect, engage in structured debates on Pol.is, and join in-depth group discussions on Remesh, a platform that uses AI to moderate live conversations and synthesise collective insights. AI's clustering function was able to reveal unexpected areas of agreement and division across thousands of inputs, thus equipping decision makers with grounded evidence about citizen priorities, even amid crises.

In Kenya, Build Up's GroundUp strategy used AI to blend offline barazas (youth convenings) and radio with digital amplification. A WhatsApp bot was developed to support structured policy dialogue while online engagement reached more than 500,000 citizens. The process had a direct impact on the Public Participation Bill and the Youth Bill (2024), showing how grassroots voices, when amplified by AI, can shape national policy. The examples of Somalia and Kenya furthermore illustrate how technology can create meaningful channels for citizen voices in conflict-affected states, potentially laying a groundwork for future democratic development.

2.5. Building capacity to govern AI: legal, procurement, and assurance levers

Throughout this chapter, "capacity" has meant the practical ability of public institutions and their partners to (i) choose appropriate AI uses, (ii) contract and procure responsibly, (iii) monitor systems once deployed, and (iv) provide timely correction and redress when harms occur. Such capacity is made possible through the presence of three sets of levers: legal, procurement and assurance.



These levers are important because, within democracies, election integrity, service delivery, and citizen engagement systems tend to fail in similar ways: unclear accountability, weak documentation, and an inability to respond quickly under pressure. Building institutional capacity within democratic contexts to adequately adopt and manage AI requires coordinated action across multiple domains.

Legal frameworks provide the foundation necessary to control AI in a democratic context. The risk-based duties and post-market monitoring requirements in the EU's AI Act offer a reference model that other jurisdictions can adapt.[72] Risk-based duties mean that the obligations of developers, deployers and users are scaled in line with the level of risk: minimal-risk systems (such as spam filters) face almost no requirements whereas high-risk systems (such as those used in hiring, credit scoring or education) must comply with strict rules on risk management, data quality and human oversight. Unacceptable-risk systems (such as social scoring or manipulative toys) are banned outright. Insistence on post-market monitoring ensures that oversight continues after deployment, obliging providers of high-risk systems to track performance, report incidents, and take corrective action if harms or compliance failures emerge.

National standards, such as the National Institute of Standards and Technology AI Risk Management Framework (2023), a US government toolkit for assessing and mitigating risks across the AI lifecycle, provide practical scaffolds for identifying, measuring and governing risk.[73] The EU Digital Services Act (2022) illustrates how platform duties can be tied to systemic risk mitigation during elections by requiring large platforms to address disinformation, electoral interference, and other harms to fundamental rights while improving transparency and accountability.[74] The clear labelling and disclosure of sponsors required by the EU Political Advertising Regulation (2024) demands an additional level of transparency for political ads as a way to prevent covert influence and ensure voters know who is behind a political message.[75] By contrast, the UK's national standard adopts a "pro-innovation" approach to AI regulation. Rather than constructing a set of prescriptive rules, this policy emphasises trustworthiness as its overarching goal, with the stress then placed on the adoption of a set of core organising principles and conditions, such as accountability structures, risk management processes and assurance practices.[76]

Procurement and vendor management represent a lever that is powerful but underutilised. Contracts should require comprehensive documentation, including Model Cards and Datasheets (see Section 2.3) as well as clear incident-reporting timelines and ATRS-compatible transparency records (see Section 2.2).[77] The UK Guidelines for AI Procurement emphasise the importance of multidisciplinary decision-making, insist that suppliers be transparent about their data and model origins, and demand that buyers be allowed to retain the contractual right to audit systems and exit agreements if needed.[78]

Civic-tech partnerships offer a third type of lever, particularly when structured as a long-term collaboration rather than a one-off project. Funding open platforms and research partnerships widens participation, reduces translation and accessibility costs, and creates learning feedback loops that improve both technology and governance.[79]

"institutional capacity for successful AI adoption depends on effective change management within the public organization adopting the respective AI system."

Perhaps most critically, institutional capacity for successful AI adoption depends on effective change management within the public organization adopting the respective AI system. To ensure that new practices take hold, the implementation process requires clear backing from senior leaders, a compelling case for change, adjustments to roles and incentives, and regular reinforcement. In public organisations, this process must also include staff consultations and equality impact assessments, training linked to new job requirements, phased rollouts that gather feedback, and other practical monitoring tools for tracking progress, such as impact assessments and adoption dashboards.[80]

For practitioners, the immediate priority is not national infrastructure planning but adopting a repeatable delivery lifecycle that makes AI deployments auditable, contestable, and safe to pause or roll back.

2.6. Operating Model for Democratic AI Projects

Public sector institutions need a repeatable operating model that aligns with established service standards while addressing AI-specific challenges. The UK government's Discovery-Alpha-Beta-Live framework, used across its Government Digital Service, provides a clear step-by-step structure for AI development, with defined outputs at each stage.[81]

During the Discovery stage, teams validate whether the problem is well defined, check if the proposed solution fits, and evaluate non-AI alternatives. This phase must include an appropriateness test, mapping potential harms and their reversibility, and identifying data sources and quality gaps. [82] In the Alpha stage, teams build a minimal viable model or procure a pilot, decide when people must check or overrule system outputs, design how the system will be evaluated and switched off if needed, and prepare clear information for the public on how the system should be used.[83]

During the Beta stage, limited-scale deployment of the system with rigorous testing for bias and robustness takes place. Teams must publish transparency records, establish appeal and redress routes, agree incident thresholds and reporting procedures, and finalise contracts with appropriate audit and exit rights.[84] Once in the Live stage, the focus shifts to post-market monitoring, periodic evaluation, publishing change notes, independent assurance routines, refreshed transparency records, and conducting rollback procedures when thresholds are breached.[85]

To prevent the diffusion of responsibility within AI projects, clear roles and accountability structures are necessary. Each project needs a Senior Responsible Owner with authority and a budget, supported by Product or Service Owners, Data Stewards, Assurance Leads, Vendor Managers, Ethics or Legal Advisors, Security Specialists, and Communications Leads.[86] Publishing a simple chart that maps these roles against key tasks or decisions in a project – a Responsible, Accountable, Consulted and Informed (RACI) matrix – can clarify responsibilities across the lifecycle, including consultation points with staff bodies and equality impact assessments.

Effective assurance also needs to define what "good evidence" looks like. Some examples are testing plans, a clean dataset set aside for independent validation, the careful analysis of errors, the application of appropriate fairness metrics, genuine user research, clear records of system changes, and independent external audits. The assurance framework established by the UK's Centre for Data Ethics and Innovation supports proportionate third-party evaluation without adding unnecessary bureaucracy.[87] At the same time, these public guidelines stress that development teams must maintain a clear public narrative that communicates the system's purpose, safeguards, and avenues for redress in plain language, by default publishing transparency records and evaluation summaries.[88]

2.7 Conclusions

Democracies can retain agency in the age of AI. The most effective way to exercise this influence is by closing the capacity gap in the three domains where AI intersects with core democratic functions: elections, public service delivery and citizen engagement. Minimum safeguards, clear ownership, simple ways to measure performance in each area, disciplined change management and enabling infrastructure, all implementable, can make AI deployments governable now. Cross-cutting levers and a sequenced roadmap can convert pilot projects into lasting institutional muscle. Ultimately, the choice facing democracies should not be between acceleration and restraint, but between governed and ungoverned acceleration.[89]

Implementing an AI governance regime requires sequenced action across three horizons. The roadmap below is written for (i) EMBs and their operational partners (courts, telecoms regulators, platforms, internet service providers); (ii) public service agencies deploying AI-assisted decision making; and (iii) organisations running citizen-engagement platforms (including civic tech), with supporting roles for regulators, auditors and funders.

With these thoughts in mind, this report makes the following key recommendations outlined in the next section.

2.8 Recommendations

Horizon 1 (immediate safeguards):
establish minimum controls that can operate under time pressure

Elections:
- ☐ Adopt a simple AI-election incident workflow (reporting channel, evidence capture standard, verification triage, escalation contacts with platforms/telecoms, and public communications templates).
- ☐ Rehearse a "degraded mode" for critical election functions (including analogue fallbacks where relevant).
- ☐ Publish a short post-incident learning template so that each event produces reusable operational lessons.

Public service delivery:
- ☐ Require a basic register of AI-assisted decisions.
- ☐ Ensure decision logging (inputs, model/version, operator, outcome), human override, and a clearly signposted appeal route before scaling.
- ☐ Put rollback and manual fallback procedures in place up front.

Citizen engagement: Implement baseline protections against manipulation, with
- ☐ Clear moderation rules.
- ☐ Rate-limiting and abuse controls appropriate to context.
- ☐ Transparency about when AI can be used for summarisation/clustering.
- ☐ An independent mechanism to review contested moderation or summaries.

Across all three domains:
- ☐ Assign a single accountable senior owner per AI deployment.
- ☐ Ensure procurement contracts include documentation, audit rights, incident reporting duties, and exit/termination rights.

Horizon 2 (routine capability):
move from ad hoc responses to repeatable practice

Elections:
- ☐ Formalise court–platform–EMB standard operating procedures for urgent labelling/takedown decisions during regulated periods.
- ☐ Create intake routines for rapid content verification.
- ☐ Join information-sharing arrangements on election interference (including cross-border cooperation where feasible).

Public service delivery:
- ☐ Embed audit-ready artefacts (logs, system cards, evaluation summaries) into procurement.
- ☐ Establish an oversight function including audit and data-protection expertise.
- ☐ Shift from one-off impact assessments to ongoing monitoring with documented thresholds that trigger pause/rollback.

Citizen engagement:
- ☐ Commission periodic independent audits of clustering/summarisation and moderation.
- ☐ Strengthen inclusion measures (accessibility and language coverage).
- ☐ Publish evaluation summaries that enable external scrutiny without exposing personal data.

Horizon 3 (institutionalisation):
make AI governance durable and contestable

Elections:
- ☐ Codify minimum election-integrity requirements for AI-era risks (impersonation rules, political advertising transparency, incident reporting expectations, and enforceable duties for key intermediaries where legally available).
- ☐ Create post-election independent review routines.

Public service delivery:
- ☐ Formalise due process guarantees for AI-assisted decisions (contestability, reasons, timeliness, non-discrimination).
- ☐ Require periodic external assurance for high-risk systems.

Citizen engagement:
- ☐ Institutionalise standards for transparency and accountability in AI-mediated participation tools.
- ☐ Create sustainable partnership/funding models so high-integrity platforms are not dependent on short-term pilots.

03 AI and Peace Processes

by Martin Wählisch

Overview

by Catalina Perdomo

Peace processes are complex and delicate undertakings that seek to resolve wars – often the outcome of power struggles between entrenched interests – that through their operation cause immense human suffering and deep emotional and psychological trauma.[90] For conceptual clarity, a peace process in this report is understood as a process that involves persistent initiatives to achieve peace between the main antagonists in a protracted conflict, a process that is likely to require more than an isolated peace initiative.[91] This understanding of a peace process includes not only the peace negotiations itself, but also any ceasefires, peace agreements and its implementation mechanisms, and post-agreement peacebuilding.

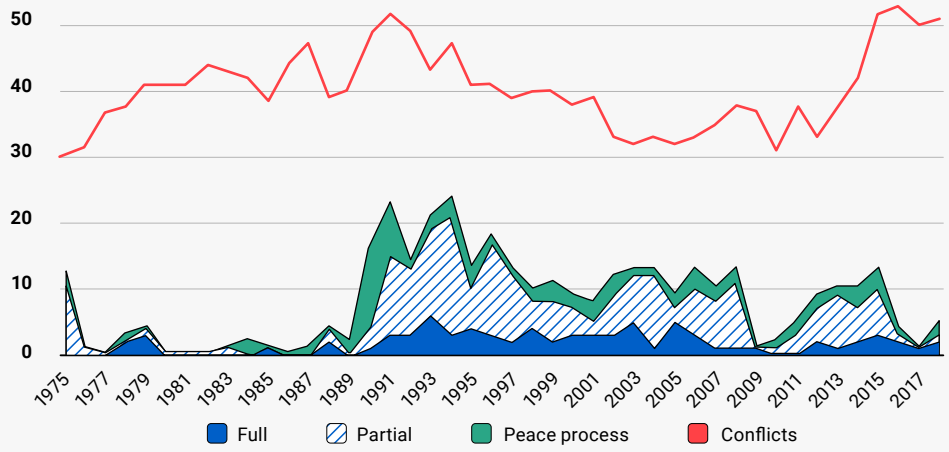
Of the various stages in a peace process there is, typically, the preparation for peace, including the so-called talks about the talks and possibly a ceasefire agreement, the peace negotiations phase, including both framework and comprehensive peace negotiations; and the peacebuilding stage, which encompasses the implementation of the agreements and any renegotiation or renewal of the original accords. All these stages are influenced by a potential spike in conflict-related violence or renewed political upheaval. Such disruptions make a peace process fragile and usually not linear.[92]

Effectively supporting a peace process requires highly skilled diplomatic and political efforts and no shortage of resources. Technical advisers, mediators, guarantors and observers need to have a profound understanding of the context and actors involved in the conflict and the means necessary to respond to the process's specific needs. Designing a peace process, requires both understanding and expertise in a range of areas, including community deliberations; territorial options; power sharing provisions; the disarmament, demobilisation and reintegration of former combatants; transitional security arrangements; transitional justice and reconciliation; among other areas of technical knowledge.[93]

These process components require also detailed analyses that encompass vast amounts of information, such as national legislation, local consultations, comparative international cases, and international law. With the growth of technology, especially social media, the pace of peace has increased, with actors involved in a peace process expected to respond rapidly to any official statements. Additionally, to improve the probability of sustainable and long-lasting peace, negotiation processes should be more inclusive and require more consultations and participation from various actors. Given that a peace process is liable to rapid and unexpected change, negotiators, advisors, guarantors and observer stakeholders, often need to rethink their analysis and even revisit their interests, priorities and concerns at short notice. The intricacy and sophistication a peace process requires can often overwhelm its supporters and participants, making failure to reach an agreement a likely outcome.[94]

The future for peace processes today seems cloudy. More and more peace processes seem to be failing: long lasting agreements seem to be in decline.[95] According to the Peace Agreement Dataset from the Uppsala Conflict Data Programme (UCDP), between 2015–2018 the world reached a peak in the number of active conflicts, similar to the levels reached during the period 1991–1994. The difference between these two peaks is that during the period 2015-2018 fewer peace agreements took place, compared to the period 1991-1994 (see Graph 1). In fact, only one peace accord ended an armed conflict during the period 2015-2018, the agreement between the government of Colombia and the Revolutionary Armed Forces of Colombia (FARC-EP) (see Graph 1).

Graph 1. Conflicts and peace agreements by type, 1975–2018



Source: Högbladh, S. (2021). Peace agreements in armed conflicts: focusing on finding a solution to the conflict incompatibility. Pathways to Peace and Security, (2), 11–23. https://doi.org/10.20542/2307-1494-2021-2-11-23, p. 20.

This murky future for peace processes has been accompanied by a global political landscape that is less supportive than it once was of multilateralism and its role in peace dialogues.[96] Due to unprecedented cuts in UN funding in 2024 it is likely that support for peacebuilding will fall dramatically in the years to come.[97] As an example, between 2015 and 2024 the number of personnel deployed to multilateral peace operations fell by more than 40 per cent.[98] Decreasing support for multilateral peace-building is taking place alongside a changing landscape of decreasing number of peace agreements, rise in global conflicts and threats to the status and capabilities of the UN's diplomatic and peacekeeping architecture. A variety of other actors, such as regional organizations, major and middle powers and civil society organisations, are all claiming more space in the peace processes support and mediation field.[99]

"In light of this evolving political landscape, one key question emerges: how can support to peace processes remain effective despite shrinking resources, a weakened multilateral system, increasingly fragmented diplomatic agendas among major and middle powers, and a declining political appetite for peace dialogues in a world marked by a growing number of armed conflicts?"

In light of this evolving political landscape, one key question emerges: how can support to peace processes remain effective despite shrinking resources, a weakened multilateral system, increasingly fragmented diplomatic agendas among major and middle powers, and a declining political appetite for peace dialogues in a world marked by a growing number of armed conflicts?

Within this climate of reduced funding and lost legitimacy, AI has the potential to enhance the capacity and efficiency of peace process support. At the same time, integrating new technologies into traditional peacemaking also carries risks that must be carefully considered. As a dual-use technology, AI can serve both beneficial and harmful purposes.

Potentials in the use of AI in peace processes

A key aspect of AI's potential to support peace processes lies in its ability to process big data, large datasets that are known for their volume, variety and velocity.[100] AI can facilitate data collection, analysis, trend identification and the detection of correlations that might otherwise go unnoticed when doing conflict analysis or strategizing for a peace negotiation.

Through various algorithms, called supervised learning, AI can be used to train machines to identify relationships between datasets and make predictions based on historical data.[101] For example, Uppsala University employs supervised learning on historical data from the Uppsala Conflict Data Programme to forecast conflict trends.[102] In this way, AI can support knowledge management and background research, both crucial for mediators and conflict parties seeking a comprehensive understanding of conflict dynamics.[103]

AI's support potential is further amplified when combined with other technologies. Ceasefire monitoring, for instance, increasingly leverages the tech capacities of satellites, crowdsourcing, data scraping, aerial platforms and AI-driven analytics. Machine-assisted data acquisition enables real-time ceasefire monitoring and verification, particularly in areas where direct access to conflict zones is hazardous.[104] The Centre for Humanitarian Dialogue, for example, has used this combined approach in its work on ceasefire agreements in eastern Ghouta, Syria, to monitor and assess the implementation of humanitarian corridors.[105] The rapid, well-informed responses to ceasefire violations that AI can generate can help to prevent the escalation of violence and sustain peace negotiations.

Research is starting to experiment with AI's potential to support peace. One study has examined the potential of supervised machine learning models to support decision-making in the demobilisation of ex-combatants in Colombia.[106] In Ukraine, AI has been used to aid disarmament, in particular to consolidate and analyse data from multiple sources to support landmine and unexploded ordnance clearance.[107] AI and blockchain technologies have also been used within the framework of transitional justice as a way to document war crimes. In 2024 Ukraine submitted to the International Criminal Court an AI-assisted evidence dossier detailing Russia's possible breaches to international humanitarian law.[108] A blockchain is a cryptographically interconnected chain of data blocks whose design ensures that modifying the contents of any single block would demand an alteration of all linked blocks.[109]

AI also has the potential to perform unsupervised learning, an ability that is particularly useful when the nature of a problem is unclear or the relationships between available datasets are unknown. This form of learning can help to uncover correlations and identify new ways of categorising data. Such a capability is invaluable when processing large volumes of unstructured information, such as national consultation dialogues within peace processes.[110] The United Nations Department of Political and Peacebuilding Affairs has applied AI-driven unsupervised learning to assess public receptivity to peace agreements in Yemen by analysing data from focus groups and large constituencies representing a variety of Arabic dialect groups.[111] The application of AI in this case both enhances inclusivity and community ownership in peace processes while enabling peace process supporters and other actors to interpret large-scale consultation outcomes more systematically and efficiently.[112]

Risks in the use of AI in peace processes

Risks related to algorithmic bias, poor data quality, insufficient explainability of results, lack of transparency, frail data security, and weak trust in technology must be carefully managed. An awareness of these risks and a willingness to implement appropriate safeguards are essential for the responsible use of AI. As noted by the Nordic Council of Ministers for Digitalisation, "countries that are successful in utilising and realising the benefits of AI, while managing risks responsibly, will have advantages in international competition and in developing more efficient and relevant public sector activities".[113] The challenge in maintaining this balance is not unique to peace processes; it reflects the broader difficulties associated with AI adoption in general and demonstrates the risks encountered when applying AI to peacebuilding contexts. Ensuring confidentiality and fostering AI literacy among peacemakers is crucial for the safe and informed use of technology.

The effective use of AI in peace support depends largely on how it is used. Ethically grounded technology use can potentialize human capabilities if risks such as security breaches, data biases, corporate influence or even the oversimplification of problems can be mitigated.[114]

Other possible applications of AI in peace processes are outlined later on in this chapter. Rather than attempting to present a comprehensive literature review, this chapter brings together a curated set of case studies to illustrate the opportunities associated with the use of AI to support peace processes. The focus on opportunities is intentional, aiming to take a constructive approach to the analysis. There are multiple risks in using AI to support peace processes, and some are mentioned in the following text, but that is not the focus of this chapter. The below presented cases were part of the workshop session "Ai-Pax: Exploring Ai Innovations in Peace Processes", held at the 2025 Stockholm Forum for Peace and Development, organized by FBA together with the University of Birmingham.

The integration of AI into peace processes, particularly in the realm of peace mediation, marks a pivotal advancement in the modernization of conflict resolution approaches. Current armed conflicts involve a multitude of actors, numerous cross-cutting transnational influences, and deeply entrenched grievances so the tools required to mediate and resolve these conflicts must evolve. AI, with its capacity to process vast volumes of information, detect patterns, and provide real-time analysis, offers a unique opportunity to support and enhance every stage of the peace process, from preparatory work to active dialogue facilitation and long-term peacebuilding.

In the preparatory phase of a peace process, AI can significantly improve the quality and depth of conflict analysis. Through techniques such as natural language processing and machine learning, AI systems can sift through massive amounts of unstructured data quickly and with purpose. It can survey news reports, social media content, historical documents, and diplomatic communications with the goal of mapping conflict dynamics, identifying key actors and influencers, detecting grievance narratives, and highlighting potential spoilers or flashpoints. This analytical capacity can equip mediators and peace support teams with sharper situational awareness, enabling more informed strategic planning and stakeholder engagement prior to formal dialogue.

During the mediation and dialogue phase, AI can play a supportive role in real-time translation, sentiment analysis, and discourse mapping, making cross-cultural and multilingual communication more seamless and inclusive. AI tools can also assist mediators by tracking the positions and preferences of negotiating parties over time, simulating potential negotiation outcomes, and flagging contradictions or inconsistencies in evolving proposals. These tools, while not a substitute for human intuition or empathy, can complement human facilitation by reducing information overload and improving process management, particularly in complex, multi-track dialogues.[115]

In the post-agreement phase, AI can contribute to peacebuilding and implementation monitoring by tracking commitments, analysing implementation data, and detecting early signs of relapse into conflict. For example, AI-driven sentiment analysis and social listening tools can assess public perception and community-level tensions, allowing peacebuilders to respond proactively to emerging grievances. Predictive analytics can support early warning systems, helping to prevent violence before it escalates while AI-enabled platforms can foster civic participation and feedback loops between the affected communities and implementing bodies.

Despite this potential, there remains a significant degree of hesitation and scepticism among peacebuilding professionals regarding the adoption of AI. Concerns include the risk of dehumanizing inherently human processes such as dialogue and reconciliation, the potential for algorithmic bias to exacerbate existing inequalities, and the fear that reliance on opaque technological systems could further undermine trust among conflict parties. A further concern is feasibility: in sensitive mediation and negotiation contexts, parties may be unwilling to share information with AI systems if doing so could expose vulnerabilities or place them at a relative disadvantage. This limits the practicality of many advanced analytical functions and constrains the conditions under which AI can be meaningfully applied. Furthermore, the ethical, political, and cultural sensitivities involved in peace mediation necessitate a cautious and context-sensitive approach to technology deployment.[116] Nonetheless, the challenges posed by today's conflicts make it increasingly difficult to justify a complete avoidance of technical innovation. Rather than shying away from AI, the peacebuilding community must engage with it critically and constructively. It should be acting to ensure that the design and implementation of AI systems are grounded in ethical principles, local ownership, and complementarity with human expertise.

"Nonetheless, the challenges posed by today's conflicts make it increasingly difficult to justify a complete avoidance of technical innovation. Rather than shying away from AI, the peacebuilding community must engage with it critically and constructively. It should be acting to ensure that the design and implementation of AI systems are grounded in ethical principles, local ownership, and complementarity with human expertise."

The rest of this chapter thus aims to explore the potential benefit AI can bring to peace processes, not as a replacement for human mediators, facilitators or analysts, but as a set of tools that can support, enhance, and scale the impact of peace processes when used thoughtfully. The following sections investigate current AI applications, assess evolving trends, and consider the implications of broader AI integration while also addressing the operational and normative questions that must be resolved to enable responsible and effective adoption in the field. The analysis refers only to selected aspects and stages of peace processes, first because the number of available empirical examples remains limited, and second because the intention is to be illustrative rather than to provide a comprehensive treatment of the issue in its entirety.

3.1. Parameters: Unlocking Opportunities and Navigating the Risks of AI in Peace Processes

AI technologies can offer valuable support across different phases of peace processes. Early warning, dialogue facilitation, monitoring, and post-agreement peacebuilding can all benefit from AI's technical capabilities. While adoption is not yet widespread, growing experimentation by international organizations, research institutions, and NGOs signals an expanding frontier. At the same time, the known structural, ethical, and political risks associated with AI suggest that its deployment in sensitive conflict-affected settings must be approached with caution.

Where AI Can Make a Difference: Emerging Opportunities

Early conflict warning systems all present an opportunity for AI enhancement. By analysing diverse and large-scale data sources such as satellite imagery, social media trends, news coverage, and humanitarian reports early signs of conflict escalation can be easily detected. When properly trained and localized, an AI model could be able to identify patterns that a human analyst might overlook. The result would be more timely alerts and an early start to preventive diplomacy.[117]

In ceasefire monitoring, AI tools linked to satellite data and remote sensing technologies can identify troop movements, weapons deployments, or the destruction of infrastructure, evidence that could support the verification of violations. AI tools may increase the objectivity, frequency, and geographic scope of monitoring missions and can be particularly valuable in hard-to-reach or insecure areas. AI can also support digital dialogue and remote mediation efforts. Natural language processing (NLP), speech recognition, and real-time translation tools may help bridge linguistic and cultural divides, enabling more inclusive and accessible communication between negotiating parties. AI can assist in managing large volumes of information, summarizing discussions, and tracking shifts in positions over time.

In addition, AI offers utility in monitoring online narratives and public sentiment. Machine learning algorithms can help detect hate speech, incitements to violence, and other disinformation campaigns on social media. These insights can inform peacebuilders' strategic communication efforts and allow for more responsive engagement with communities vulnerable to digital manipulation.

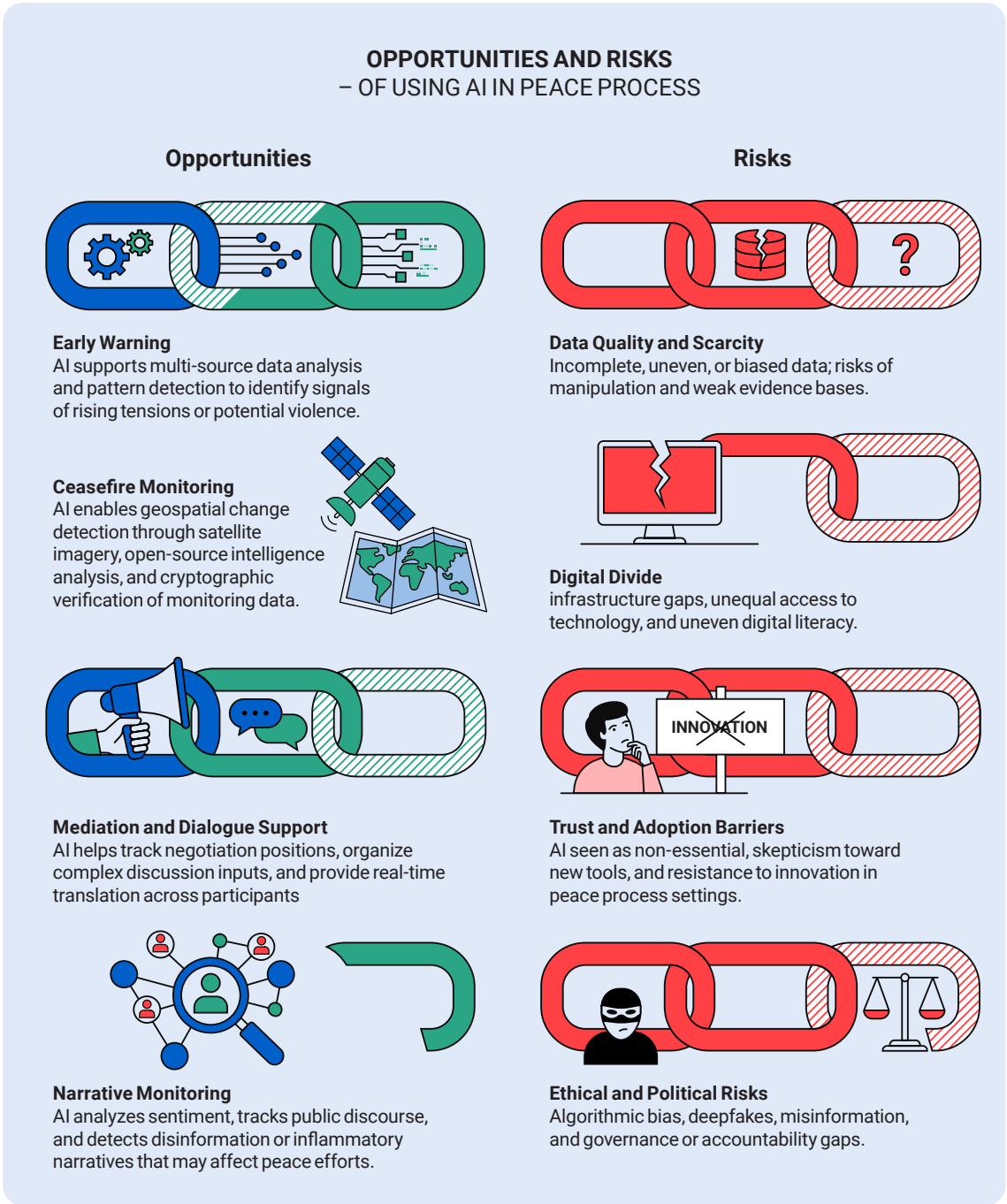
The Roadblocks Ahead: Critical Risks and Barriers

Despite its potential, the use of AI in peace processes is accompanied by significant constraints.

"Without accurate, representative, and trusted datasets, AI tools risk producing flawed outputs: misidentifying threats, reinforcing dominant narratives, or excluding marginalized voices from analysis and decision making"

Data scarcity and data quality are foundational concerns. In conflict-affected settings, available data are often incomplete, outdated, or politically distorted. It is also likely that conflict parties will withhold information, share selectively, or provide misleading inputs. Where the use of AI systems is known, actors may actively attempt to game models by manipulating signals or narratives. Without accurate, representative, and trusted datasets, AI tools risk producing flawed outputs: misidentifying threats, reinforcing dominant narratives, or excluding marginalized voices from analysis and decision making. The digital divide presents another major barrier. In many conflict-affected areas, internet access is

unreliable or entirely absent, and electricity, hardware, and secure servers, a basic digital infrastructure, is often limited. These conditions reduce the feasibility of using cloud-based AI systems and restrict the reach of digital peace initiatives. As a result, AI tools often remain inaccessible to the very communities most in need of inclusive peacebuilding efforts.



Additionally, there is an enduring perception within many peacebuilding circles that AI and other digital tools are supplementary rather than essential.[118] In high-stakes mediation contexts, technology can be seen as peripheral to core political work. Unless AI demonstrates a clear and context-specific value, peace professionals may continue to prioritize traditional methods over digital innovation.

Ethical risks and the misuse of AI are also serious concerns. Algorithmic bias, whether stemming from poor data or opaque model design, can undermine trust in peace processes. Furthermore, AI technologies can be co-opted to produce deepfakes, spread disinformation, or enable targeted surveillance – tools that may be used by actors seeking to sabotage negotiations or manipulate public perception.

Finally, regulatory and political challenges can complicate implementation. Many conflict-affected states lack clear policies on AI governance, data protection, and digital cooperation. Peace mediators often operate in politically sensitive environments where neutrality is critical; the use of AI tools – especially those developed or managed externally – can raise questions about control, transparency, and legitimacy.

The following section introduces a set of case studies designed to illustrate how AI has been applied in practice as a tool to support peace processes. The cases provide concrete examples of how AI-enabled tools and methods have been used to inform analysis, facilitate dialogue, support decision-making, and enhance inclusivity in peace-related contexts. These narratives place their primary emphasis on the opportunities that the use of AI presents to peace processes, a position that is intended to counterbalance the fact that, as noted earlier, the existing literature in this field has tended to focus on AI's constraints and risks.

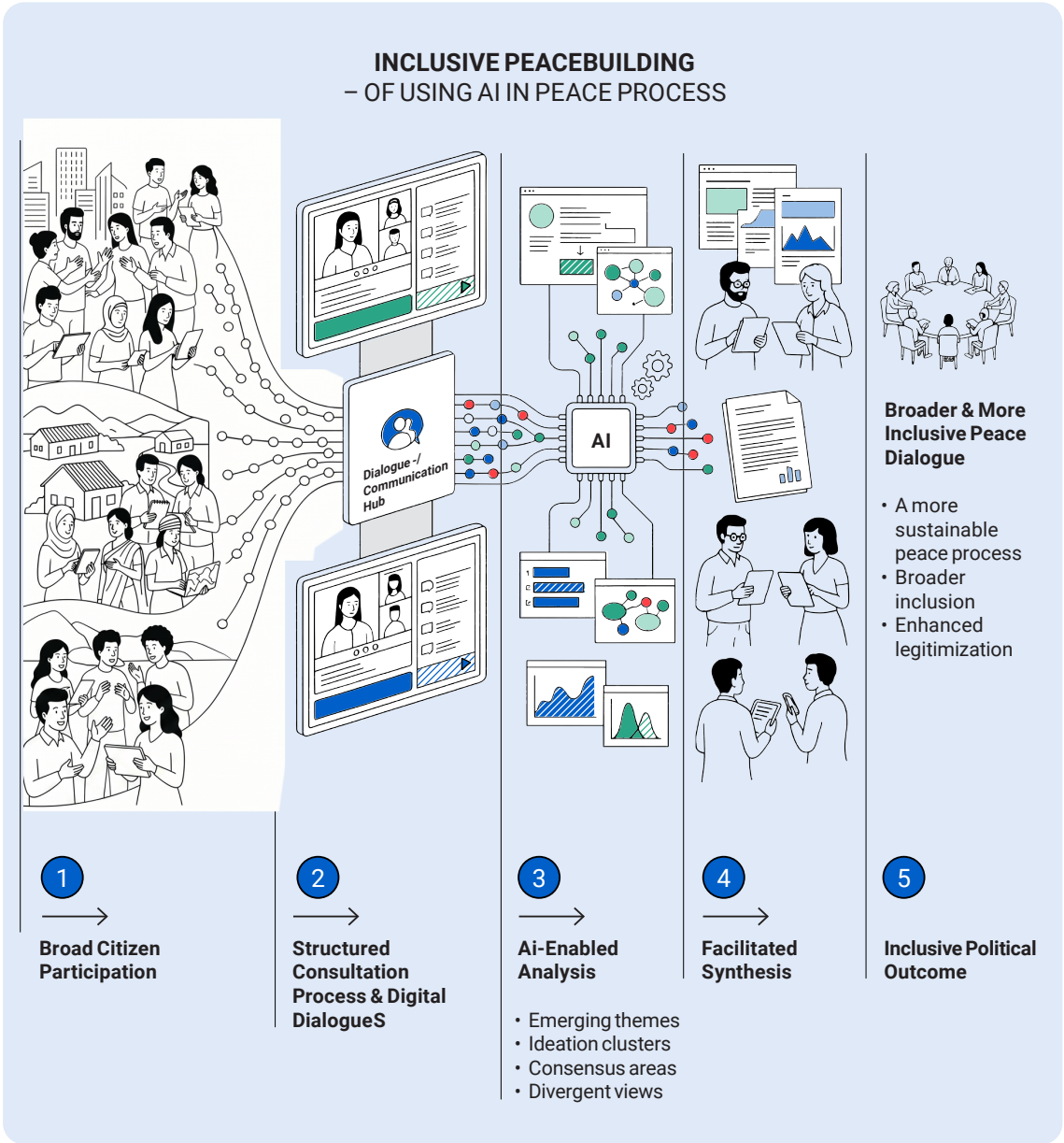
3.2. Case Studies: AI Applications in Peace Processes

Digital Dialogues

Digital dialogues represent a growing area of innovation in peace processes and offer scalable and inclusive formats for engaging communities across political, geographic, and cultural divides. AI-enabled platforms are increasingly being explored as tools to support structured deliberation, demonstrate common ground and reduce power asymmetries in conflict-affected settings where face-to-face engagement is limited or unsafe.[119]

Researchers from institutions such as the University of Notre Dame and the University of Birmingham have examined how factors like platform architecture, moderation design, and linguistic accessibility can shape the quality and legitimacy of digital engagement. Their findings emphasize the importance of culturally responsive, ethically grounded, and context-sensitive approaches to designing AI-enabled peace technologies.

These principles are being practically implemented by several organizations and multilateral actors. The United Nations Support Mission in Libya (UNSMIL), for example, conducted a series of large-scale digital dialogues in 2020, with over 1,000 participants from across the country taking part in discussion sessions. These consultations aimed to include a broad spectrum of Libyan voices – particularly those of youth, women, and civil society actors – in shaping the future of the peace process. AI-enabled tools helped to process these inputs in real time, identify emerging themes, and support facilitators in recognizing areas of consensus and divergence. The end result was ultimately improved transparency and a more inclusive political dialogue.



One civil society example is the AI Pulse initiative directed by the Alliance for Middle East Peace (ALLMEP) and intended for use in the Israel-Palestine conflict. The initiative integrates bridging algorithms and large language models (LLMs) to support structured digital dialogues between individuals from historically opposing communities. These tools have helped to detect patterns of agreement, highlight shared values, and overcome linguistic and ideological barriers. The result has been a more equitable degree of participation in the dialogues and the production of outputs that reflect collective visions for coexistence – even in these deeply polarized environments.

In many other national contexts, chatbot-enabled tools and platforms such as Pol.is and Remesh have demonstrated the potential to enable large-scale, multilingual, and anonymous participation. In Afghanistan, for example, peacebuilders adapted dialogue technologies to local realities, including limited internet connectivity, linguistic diversity, and uneven levels of digital literacy. These platforms provided safe and accessible spaces for citizens to engage in dialogue on governance and societal priorities while minimizing facilitator bias and preserving participant anonymity.

"The effectiveness of such approaches depends on digital access and literacy, trust in platforms and facilitators, and safeguards that protect participants from surveillance, manipulation, or data misuse. AI-supported dialogues cannot substitute for offline engagement, contextual knowledge, or political will, and may reinforce exclusion or power imbalances if deployed without careful design."

Taken together, these initiatives suggest that AI-supported digital dialogues can expand participation, enhance analytical capacity, and support more inclusive deliberation in peacebuilding processes. At the same time, they reveal important limitations and enabling conditions. The effectiveness of such approaches depends on digital access and literacy, trust in platforms and facilitators, and safeguards that protect participants from surveillance, manipulation, or data misuse. AI-supported dialogues cannot substitute for offline engagement, contextual knowledge, or political will, and may reinforce exclusion or power imbalances if deployed without careful design. Their value therefore lies not in technological novelty, but in how they are embedded within trusted institutions, ethical data practices, and broader peacebuilding strategies.

Social Media Monitoring: From Hashtags to Hotspots

AI-powered social media monitoring has become a critical tool in modern peacebuilding, particularly for identifying and responding to the online dynamics that can fuel conflict. In increasingly digital nature of conflict environments, platforms like Instagram, Twitter (X), WhatsApp, TikTok and Telegram are not just communication channels; they are also battlegrounds for disinformation, hate speech, and polarization.[120] AI technologies, including NLP, sentiment analysis, and automated content moderation, are being harnessed to detect these threats in real time and help peacebuilders respond to them more effectively.

"AI tools have helped to identify coordinated propaganda efforts, allowing platforms and civil society actors to intervene before online narratives translated into real-world harm. This underscores the central role that digital conflict prevention plays as a pillar of contemporary peace work."

One stark illustration of this potential is the case of Myanmar, where AI-assisted content moderation has been used to reduce the spread of online hate speech and incitement to violence, particularly that targeting minority communities. In this context, AI tools have helped to identify coordinated propaganda efforts, allowing platforms and civil society actors to intervene before online narratives translated into real-world harm. This underscores the central role that digital conflict prevention plays as a pillar of contemporary peace work.

A leading example of social media monitoring in civil society is the work of Build Up, a peacebuilding organization that develops tools to manage conflict in the digital age. Their open-source platform, Phoenix, is designed to ethically and participatorily track public discourse on social media.[121] Phoenix uses AI-assisted classification and visualization tools to analyse content and network patterns, allowing peacebuilders to map digital narratives, identify risks, and better understand the online dynamics that influence conflict escalation. By making this kind of analysis accessible to local actors, Build Up empowers communities and mediators to act on early digital warning signs.

In addition to Phoenix, Build Up has led projects such as The Digital Us, which supports marginalized groups – like youth experiencing racism – to reclaim digital spaces and build inclusive narratives. These initiatives show how AI-enabled monitoring can both de-escalate conflict and create new opportunities for empowerment and inclusion in digital civic life.

However, the success of AI-based social media monitoring depends heavily on language capabilities.[122] These can pose a major challenge in conflict settings where many spoken languages are low resource: they lack sufficient digital data and NLP tools. Arabic, for example, is one of the most widely spoken languages across the Middle East and North Africa, where many conflict zones are located, yet it remains underrepresented in the development of AI tools due to its complex morphology, regional dialect diversity, and limited annotated datasets.

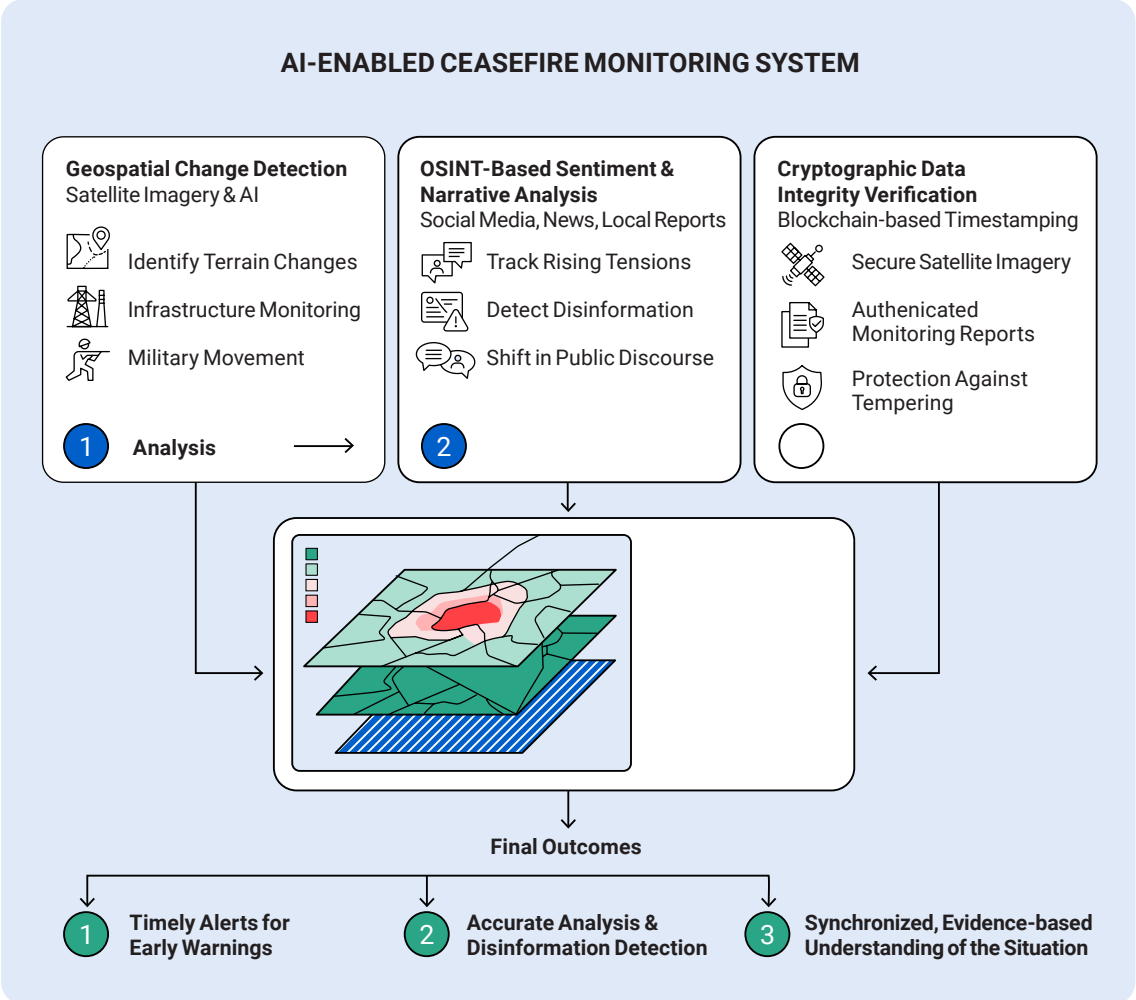
To address this, institutions like the Doha Institute for Graduate Studies have been advancing research in Arabic NLP and automated reasoning. Through projects such as the Doha Historical Dictionary of the Arabic Language, the Institute is building the linguistic infrastructure needed for more sophisticated language understanding and information extraction in Arabic. Tools like these will not only enhance the ability to process large volumes of Arabic social media content but will also contribute to building fairer and more localized AI systems for peacebuilding applications in Arab language contexts. Improving AI's ability to interpret and classify Arabic text and other low-resource languages is essential for monitoring digital spaces in many of the world's most fragile contexts, including Iraq, Libya, Sudan, Syria and Yemen. Without this capacity, harmful content can go undetected, leaving peacebuilders without the situational awareness necessary to respond effectively.

These innovations reflect a growing ecosystem of actors – from NGOs to academic institutions – working to make social media monitoring a more inclusive and robust pillar of peacebuilding strategy. By combining ethical AI, participatory design, and attention to linguistic justice, AI-powered monitoring tools can help detect early warning signals and emerging risks. However, their effectiveness depends on several conditions, including data quality, sustained

local engagement, and transparency about system limitations. Such tools cannot replace contextual expertise or human judgment; indeed, they risk reinforcing existing biases or surveillance dynamics if they are deployed without clear safeguards. When embedded in accountable governance structures and complemented by offline knowledge and response mechanisms, social media monitoring can contribute not only to risk detection but also to more resilient, inclusive, and peace-oriented digital spaces.

Ceasefire Monitoring: Enhancing Oversight and Accountability through AI

The application of AI in ceasefire monitoring is gradually reshaping how stakeholders observe, verify, and respond to violations in volatile environments. One of the field-tested examples comes from the work of Hala Systems, an organization that has developed cutting-edge AI-driven tools to support monitoring efforts in conflict zones such as Yemen and Ukraine. Their model demonstrates how AI can be deployed in complex operational settings to generate timely, credible, and actionable insights that are difficult to obtain through conventional monitoring mechanisms alone.



Hala's ceasefire monitoring system integrates several core technologies:

- Geospatial change detection: By using satellite imagery and AI-powered image recognition algorithms, Hala's platform can automatically identify changes in terrain, infrastructure, and military positioning. For example, the sudden appearance of fortifications, vehicle tracks, or damaged buildings in a demilitarized area may signal a ceasefire breach. AI tools significantly reduce the time and human labour required to process satellite data and enable near real-time alerts, allowing for rapid response and investigation by ceasefire guarantors or third-party monitors.
- OSINT-based sentiment and narrative analysis: Through its leveraging of open-source intelligence (OSINT), the Hala system is able to monitor digital platforms for changes in public sentiment, emerging disinformation campaigns, and other politically charged narratives. Through natural language processing and machine learning, their system tracks trends across social media, news outlets, and local reporting to detect shifts in public discourse that may indicate rising tensions or ceasefire arrangements being undermined. This aspect is particularly important in contexts like Ukraine, where information warfare and hybrid conflict are key elements of the security environment.
- Cryptographic data integrity verification: To address growing concerns about misinformation and data tampering – especially in highly polarized conflicts – Hala employs cryptographic methods to timestamp and verify all incoming data. This ensures that satellite imagery, incident reports, and other monitoring inputs remain unaltered from the moment they are captured, enabling stakeholders to trust the authenticity of the evidence base. In disputes where parties may question the credibility of ceasefire reports, this level of data security is critical for maintaining transparency and accountability.

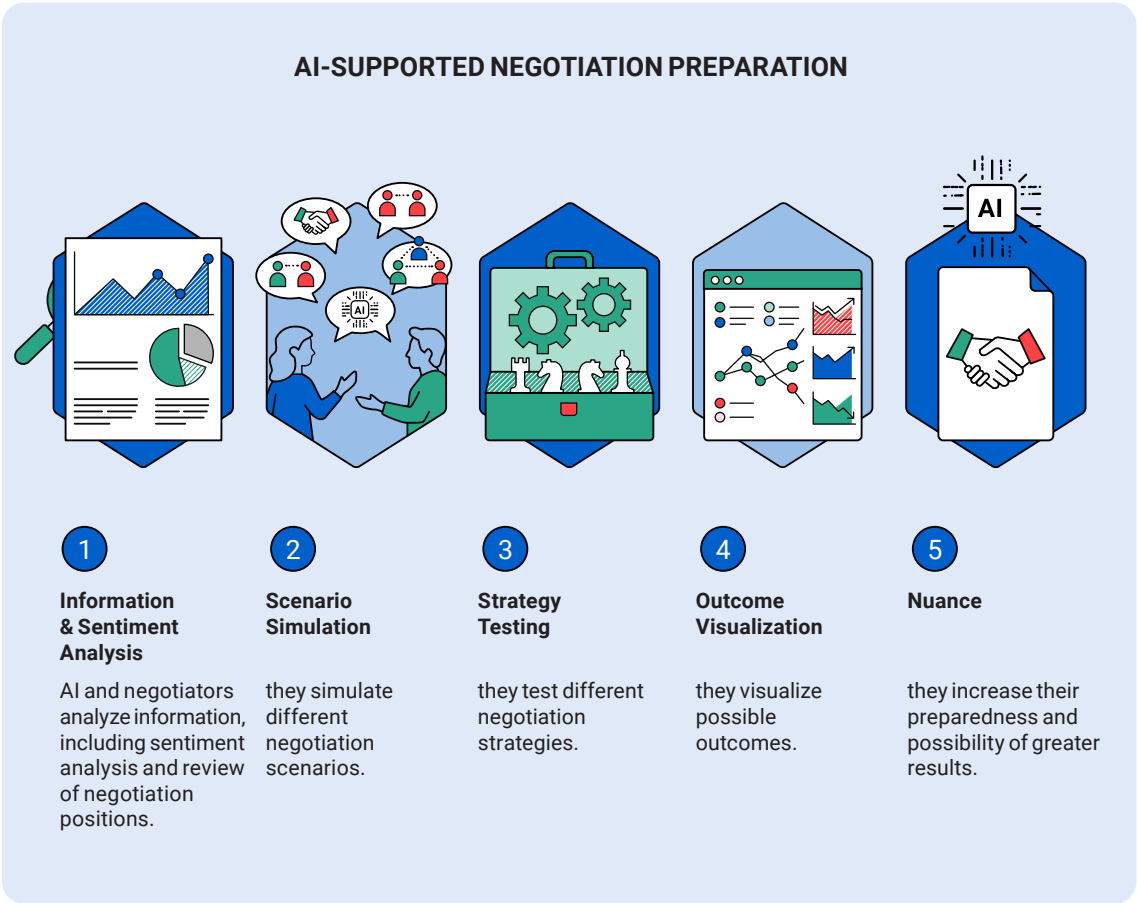
"The combined use of these tools supports the creation of a "shared operational picture", a synchronized and evidence-based understanding of the situation on the ground that can be accessed by mediators, monitors, and, in some cases, the public."

The combined use of these tools supports the creation of a "shared operational picture", a synchronized and evidence-based understanding of the situation on the ground that can be accessed by mediators, monitors, and, in some cases, the public. In practice, AI in ceasefire monitoring enhances mutual confidence among conflict parties, facilitates more objective dialogue, and reduces the risk of miscommunication or manipulation of data.

The lessons from Hala Systems' deployments emphasize the strategic value of AI not just in technical verification, but also in building institutional trust and procedural legitimacy in fragile ceasefire arrangements. However, they also point to the need for capacity-building and infrastructure support to scale such models in environments where digital access, technical expertise, or political will may be limited.

3.3. Future Frontiers of AI in Peace Processes

Several emerging trends are set to redefine the role of AI in peace processes over the coming decade. These developments are likely to shift the emphasis from current applications into more dynamic, immersive, and emotionally attuned technologies. These advances are likely to transform not only how peace is negotiated and monitored, but also how it is experienced, taught, and sustained.



Generative AI is at the forefront of this evolution. With increasing capability to produce realistic human dialogue and complex narratives, it will soon be possible for generative models to enable high-fidelity peace dialogue simulations. These simulations can be used by mediators, negotiators, and community leaders to rehearse sensitive discussions, test different negotiation strategies, and visualize potential outcomes. This is particularly valuable in high-stakes environments where real-world trial and error is not an option. AI-powered role-playing systems can also be used to support training programs for peacebuilders, helping them to engage with diverse conflict scenarios in emotionally nuanced and culturally specific ways.

Another significant trend is the convergence of AI and blockchain technology. The secure and decentralized nature of blockchain makes it an ideal companion for AI in developing transparent, tamper-proof platforms for documenting peace agreements, tracking implementation progress, and storing sensitive negotiation data. When integrated with smart contracts and automated monitoring systems, these tools could increase confidence among parties and help address issues of mistrust, data integrity, and accountability – key barriers in many peace processes.

Meanwhile, immersive technologies like Virtual Reality and Augmented Reality are beginning to reshape how empathy and understanding are cultivated in divided societies. When enhanced by AI, these platforms can create tailored experiences that immerse individuals in the perspectives of others, such as walking through someone else's daily life under conflict conditions or visualizing the shared history of a contested space. Such experiences have been shown to reduce prejudice and increase openness to dialogue, making them powerful tools for reconciliation, peace education, and trauma-informed community engagement.

Looking ahead, the next generation of peace tech may also involve agentic AI: systems capable of taking goal-directed action in collaborative processes. These AI agents could play a supporting role in live peace negotiations by helping track conversation dynamics, suggest consensus language, flag emotionally charged moments, or generate inclusive talking points based on prior engagement. If developed responsibly, it might be possible for agentic systems to act as real-time facilitators or "co-mediators" and to augment human judgment rather than replacing it.

Another future trend in AI use in peacebuilding efforts might be a shift toward more intuitive systems designed to respond to human emotional cues. Such tools could incorporate emotion recognition capabilities and draw on voice patterns, facial expressions, or physiological signals to better assess participants' states of mind and adapt communication in real time. In digitally mediated or hybrid dialogues, these systems could help identify stress or escalation risks and suggest de-escalation strategies or facilitator interventions.

Wearable technologies, including biosensors, emotion-sensing headsets, or augmented-reality devices could also possibly enter peacebuilders' toolkits. Used cautiously, these tools could provide real-time feedback on emotional dynamics during facilitation or negotiation, supporting greater self-awareness and emotional regulation in high-stakes settings.

At the same time, these applications raise significant concerns. Emotion detection systems rely on contested assumptions about the universality and interpretability of emotional signals and risk misclassification, bias, or cultural misreading. The collection of intimate biometric data introduces serious privacy and consent challenges, particularly in conflict-affected contexts marked by power asymmetries and low trust. AI-mediated interventions may also shift authority in negotiations, obscure accountability, or undermine human judgment if human actors come to rely on them too heavily. Without clear safeguards, such technologies could unintentionally exacerbate harm rather than reduce it.

As AI technologies evolve, so too must the frameworks for their governance. Ensuring that AI remains a servant to human peacebuilding values rather than a driver of technological determinism will require robust ethical safeguards, local ownership, and cross-disciplinary collaboration. Importantly, the future of AI in peace processes must remain inclusive and context-sensitive, especially in low-connectivity environments and linguistically diverse regions.

3.4 Conclusions

The integration of Artificial Intelligence into peace processes is no longer a distant proposition – it is an unfolding reality. This report has demonstrated that AI can play a meaningful role across the entire peace process continuum, from early warning and ceasefire monitoring to digital dialogue and public sentiment analysis. It can support more inclusive engagement, improve situational awareness, and offer scalable tools to navigate the growing complexity of contemporary conflict environments.

At the time of writing, AI adoption in peacebuilding is uneven and largely experimental. While there is cautious optimism, the ethical, political, and operational challenges of algorithmic bias, the digital divide, data quality, and trust persist in many quarters. These challenges underscore the need for deliberate design choices, interdisciplinary coordination, and evidence-based policy development. As AI technologies evolve toward more immersive, agentic, and emotionally responsive forms, the peacebuilding community faces both a challenge and an opportunity: to harness the potential of these tools responsibly while safeguarding the deeply human values at the heart of peacebuilding.

In sum, the integration of AI into peace processes presents a complex but promising frontier. With thoughtful investment, ethical foresight, and collaborative engagement across sectors and disciplines, AI can serve as a powerful enabler of more inclusive, adaptive, and effective peacebuilding. The challenge now is to move beyond conceptual enthusiasm and toward sustained, evidence-driven, and values-based action.

With these optimistic yet cautious thoughts in mind, this report makes five recommendations, outlined in the following section.

3.5 Recommendations

- ☐ **Build AI Literacy Among Peacebuilders and Diplomats**
  AI will increasingly shape the spaces where dialogue, negotiation, and public discourse occur. It is therefore essential that mediators, diplomats, facilitators, analysts and other peace professionals possess a working understanding of AI systems. This includes how they function, where their strengths and limitations lie, and how to engage critically with AI-generated outputs. Developing tailored capacity-building programs and cross-sector training initiatives can help bridge the current knowledge gap and ensure that peace actors are empowered, not sidelined, by technological innovation.
- ☐ **Expand Pilot Projects for AI-Enhanced Digital Dialogue and Deliberation**
  While promising initiatives such as ALLMEP's AI Pulse and UNSMIL's digital consultations in Libya provide valuable proof-of-concept, far more empirical evidence is needed. Governments, multilaterals, and civil society organizations should invest in pilot projects that test, refine, and scale AI-supported digital dialogue platforms, especially in fragile, multilingual, and digitally constrained settings. Pilots should prioritize participatory design, user feedback, and rigorous impact evaluation to ensure that the AI tools are fit for purpose and locally relevant.
- ☐ **Develop Clear Norms and Guidance on the Responsible Use of Technology**
  A shared normative framework for the responsible use of AI in peacebuilding is urgently needed. This includes practical guidance on ethical design, risk mitigation, data governance, privacy, and inclusivity. Peace actors cannot wait for global regulatory frameworks to fully emerge but should closely engage with and draw on ongoing international policy debates and normative processes. In parallel, they should begin developing their own principles, safeguards, and accountability mechanisms tailored to mediation and dialogue contexts. These efforts should be grounded in the principles of "do no harm", technological humility, and equity.
- ☐ **Strengthen Cross-Disciplinary Collaboration in the Field of Crisis Computing**
  There is a growing need to institutionalize collaboration between peace practitioners, computer scientists, data engineers, ethicists, and social scientists. The field of crisis computing – which encompasses the use of computational tools in disaster response, humanitarian work, and peacebuilding – provides a fertile ground for this exchange. Joint research programs, innovation labs, and collaborative fellowships can help integrate technical innovation with deep contextual knowledge, ensuring that AI tools address real-world needs and operate safely in volatile environments.
- ☐ **Invest in Empirical Research and Behavioural Insights on Tech Use in Peace Processes**
  While theoretical and technical work on AI in peacebuilding has advanced, the field still lacks a strong empirical foundation. More systematic studies are needed to assess the actual uptake, effectiveness, and limitations of AI tools in real-world peace processes. This includes investigating behavioural barriers to adoption, trust in technology, cultural attitudes toward AI, and unintended consequences. Research should also focus on user experience and the ways technology shapes negotiation dynamics, participation, and power asymmetries.

04 Conclusions and Recommendations

by Catalina Perdomo and Catalina Uribe Burcher

AI is transforming the way societies function, with profound implications for democratic governance and conflict resolution. Its rapid diffusion across political, social and peace domains demands responses that are ethical, forward-looking and grounded in human agency. As this report has shown, while AI can undermine democracy and peace, it can also reinforce them, depending on how governments, institutions and societies choose to design, govern and use these technologies. AI is not a deterministic force but a probabilistic one. It is therefore the choices made by policymakers, practitioners and citizens that will determine whether AI becomes a driver of conflict and inequality, or a force for trust, dialogue and sustainable peace.

As Chapter 2 highlights, AI is already altering how elections are protected, how public services are delivered and how citizens participate in public life. AI is already having an effect on the trust relationships, decision-making authority and accountability mechanisms that exist at the core of democratic systems. At the same time, Chapter 3 shows that AI is beginning to influence every stage of peace processes, from early warning and conflict analysis, to mediation, ceasefire monitoring and post-agreement implementation.

Much of the current debate focuses on whether democracies are capable of governing AI. Far less attention has been paid to how AI itself is shaping democratic practice. Across both democratic practice and peacebuilding efforts, this report identifies a dual reality. On the one hand, AI presents a real risk of amplifying polarization, deepening data and power asymmetries, enabling new forms of manipulation, reinforcing bias in analysis and participation and, in some cases, weakening human judgement in high-stakes decision making. On the other hand, AI also offers significant benefits. For democratic governance, these include gains in administrative efficiency, broader participation and enhanced transparency in elections. For peace processes, AI offers the potential to scale conflict analysis, support mediation and negotiations, and strengthen the monitoring and verification of the implementation of agreements. It also offers the opportunity to better measure public sentiment in support of peace efforts.

"A central finding of this report is that the decisive factor in determining whether AI strengthens or weakens democracy and peace processes is institutional capacity."

A central finding of this report is that the decisive factor in determining whether AI strengthens or weakens democracy and peace processes is institutional capacity. As described in Chapter 2, democracies retain agency when minimum safeguards, clear ownership, defined performance measures, disciplined change management and enabling infrastructure are put in place. When these elements are missing, even technically sophisticated systems become ungovernable in practice.

Institutional capacity cannot be created overnight; it must be built progressively. The horizon-based model set out in Chapter 2 shows how institutions can move from immediate protective measures (Horizon 1), through systematic capability building (Horizon 2), to the full institutionalization of democratic AI governance (Horizon 3). This applies across elections, public service delivery and citizen engagement, and offers a practical roadmap for moving from pilot projects to durable institutional muscle.

In examining the context of peace processes, Chapter 3 makes the clear point that AI should be understood, not as a replacement for human mediation and diplomacy, but as a set of enabling tools that can augment analysis, inclusion and responsiveness across the peace process continuum. At the same time, persistent constraints, such as data quality, the digital divide, algorithmic bias and trust deficits, mean that responsible adoption will require ethical design, contextual sensitivity and sustained investment in human capabilities.

"AI governance is no longer only about regulating technology. It is about renewing democratic and peacebuilding institutions themselves under the conditions of rapidly accelerating technological change."

Taken together, the findings of this report point to a clear conclusion: AI governance is no longer only about regulating technology. It is about renewing democratic and peacebuilding institutions themselves under the conditions of rapidly accelerating technological change.

To ensure that AI strengthens rather than undermines democratic governance and peace processes, this report proposes the following priorities for governments, public institutions, peacebuilding actors, civil society and international partners:

- **Assess why and when AI technologies are needed.**
  The adoption of any technology should not occur merely for its own sake. Whether in democracy support or peacebuilding, the underlying processes must first be carefully examined to determine the specific gaps or limitations that cannot be addressed through existing methods. Only then should AI be assessed for its ability to offer added value in meeting those identified needs. In other words, needs should be defined first, followed by an evaluation of the relevance and appropriateness of employing AI technologies to address them.
- **Align AI with human rights and democratic values.**
  Governments and public authorities should ensure that all AI systems in general – especially those used in democratic governance and peace processes – are aligned with international human rights standards, including principles of non-discrimination, transparency and due process. The aim is to prevent the normalization of bias, automated exclusion and arbitrary decision-making.
- **Embed ethical and conflict-sensitive safeguards from the outset.**
  Policymakers and public institutions should require ethical, participatory and conflict-sensitive design from the earliest stages of AI development and procurement, particularly for systems used to support democratic governance and peace processes. Experience from emerging "DemTech" and "PeaceTech" initiatives shows that early safeguards are decisive for sustaining public trust.
- **Strengthen institutional capacity across the full AI lifecycle.**
  Democratic institutions, civil society organizations and peacebuilding actors must invest in skills, lawful and responsible data, appropriate tools, clear accountability structures and structured approaches to organizational change. AI will only strengthen democracy and peace processes where governance capacity keeps pace with technical capability.

- **Design AI adoption for accountability and public trust.**
  Public institutions should ensure that transparency, contestability, security and clear lines of responsibility are built into every stage of AI procurement, deployment and use. This includes enabling individuals and institutions to question, appeal and, where necessary, challenge AI-supported decisions and request accountability in peace process support tools.
- **Ensure whole-of-government coherence.**
  Governments should align legal, regulatory, technical and infrastructural measures so that responsibility for AI is not fragmented across institutions or departments. Coherent governance frameworks are essential to avoid regulatory gaps, clarify lines of authority and ensure responsible use across sectors and agencies.
- **Promote inclusive and accessible AI innovation.**
  Governments and international partners should uphold inclusive innovation as a core objective of AI adoption. Linguistic, digital and accessibility divides must be addressed so that participation in democratic and peace processes is broadened, not restricted, by AI. This is particularly critical in fragile, low-connectivity and conflict-affected settings.

05 References

Amiri, F. (2025, October 9). United Nations to cut 25% of its global peacekeeping force in response to US funding strains. Associated Press. https://apnews.com/article/trump-united-nations-cuts-review-154e7b2756a08140c6fcecd6c38ce0f5

Amnesty International. (2021). Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch child benefits scandal. https://www.amnesty.org/en/documents/eur35/4686/2021/en/

Amnesty International. (2022, September 29). Myanmar: Facebook's systems promoted violence against Rohingya; Meta owes reparations – new report. https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/#:~:text=Facebook%20owner%20Meta's%20dangerous%20algorithms,a%20new%20report%20published%20today

AP News. (2025). New Hampshire jury acquits consultant behind AI robocalls mimicking Biden on all charges. https://apnews.com/article/e87ef8848a2a6515202535a6e800da07

Amiri, F. (2025, October 9). United Nations to cut 25% of its global peacekeeping force in response to US funding strains. Associated Press. https://apnews.com/article/trump-united-nations-cuts-review-154e7b2756a08140c6fcecd6c38ce0f5

Amnesty International. (2021). Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch child benefits scandal. https://www.amnesty.org/en/documents/eur35/4686/2021/en/

Amnesty International. (2022, September 29). Myanmar: Facebook's systems promoted violence against Rohingya; Meta owes reparations – new report. https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/#:~:text=Facebook%20owner%20Meta's%20dangerous%20algorithms,a%20new%20report%20published%20today

AP News. (2025). New Hampshire jury acquits consultant behind AI robocalls mimicking Biden on all charges. https://apnews.com/article/e87ef8848a2a6515202535a6e800da07

Arana-Catania, M., van Lier, F. A., & Procter, R. (2022, September 2). Supporting peace negotiations in the Yemen war through machine learning. Data & Policy 4: e28. DOI: 10.1017/dap.2022.19.

Arandjelovic, O. (2021, September 10). AI, democracy, and the importance of asking the right questions. AI & Ethics Journal, 2(1): 2. DOI: 10.47289/aiej20210910.

Ashby, H. (2023, December 6). A role for AI in peacebuilding. United States Institute for Peace.

Barr, L. (2024, September 25). Russia and Iran using AI to influence US election: DNI.

ABC News. https://abcnews.go.com/Politics/russia-iran-ai-influence-us-election-dni/story?id=113941680

Barltrop, R. (2025). Review of How is peace made? Conflict mediation in the Arab world, edited by Ibrahim Fraihat and Isak Svensson. Journal of Intervention and Statebuilding, 19(4): 558–565. DOI: 10.1080/17502977.2025.2513718.

Beckman, L., Hultin Rosenberg, J., & Jebari, K. (2022, July 2). Artificial intelligence and democratic legitimacy: The problem of publicity in public authority. AI & Society, 39: 975-984. DOI: 10.1007/s00146-022-01493-0.

Bell, C. (2024). PeaceTech: Digital transformation to end wars. Springer International. DOI: 10.1007/978-3-031-38894-1.

Benkler, Y., Faris, R. & Roberts, H. (2018). Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press. DOI: 10.1093/oso/9780190923624.001.0001.

Brkan, M. (2019). Artificial intelligence and democracy: The impact of disinformation, social bots and political targeting. Delphi, 2: 66–71. https://delphi.lexxion.eu/data/article/14587/pdf/delphi_2019_02-005.pdf

Bürokratt (RIA). (2025). The vision of Bürokratt. https://www.ria.ee/en/state-information-system/personal-services/burokratt

Busiek, J. (2024, March 21). Three fixes for AI's bias problem. University of California. https://www.universityofcalifornia.edu/news/three-fixes-ais-bias-problem

Centre for Data Ethics and Innovation. (2020). Review into bias in algorithmic decision-making. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf

Christodoulou, E. & Iordanou, K. (2021, July 21). Democracy under attack: Challenges of addressing ethical issues of AI and Big Data for more democratic digital media and societies. Frontiers in Political Science, 3. DOI: 10.3389/fpos.2021.682945.

Coalition for Content Provenance and Authenticity (C2PA). (2024). C2PA Specifications. https://spec.c2pa.org/specifications/specifications/2.2/index.html

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Bernhard, M., Cornell, A., Fish, S. M., Gastaldi, L., Gjerløw, H., Glynn, A., Grahn, S., Hicken, A., Kinzelbach, K., Marquardt, K. L., McMann, K., Mechkova, V., Neundorf, A. et al. (2023, March). V-Dem Codebook v13 – March 2023. V-Dem Institute, University of Gothenburg. https://www.v-dem.net/documents/24/codebook_v13.pdf

Darby, J., & Mac Ginty, R., eds. (2003). Contemporary peacemaking: conflict, peace processes and post-war reconstruction. Palgrave Macmillan. DOI: 10.1057/9781403918475.

Barandiaran, X.E., Calleja-López, A., Monterde, A. & Romero C. (2024). Decidim: A Technopolitical Network for Participatory Democracy: Philosophy, Practice and Autonomy of a Collective Platform in the Age of Digital Intelligence Springer Cham. DOI: 10.1007/978-3-031-50784-7.

Delahoz-Domínguez, E., Carrillo-Naranjo, J., Camelo-Guarín, A., and Zuluaga-Ortiz, R. (2023, March). Machine learning application in the ex-combatant demobilization process of the Colombian armed conflict. Intelligent Data Analysis: An International Journal, 27(2): 501–517. DOI: 10.3233/IDA-216397.

Demirjian, K. (2024, September 26). 'Deepfake' caller poses as Ukrainian official in exchange with key senator. The New York Times. https://www.nytimes.com/2024/09/26/us/senator-cardin-deepfake.html

Department of Finance, Sweden. (2025, October 6). Anpassningar till AI-förordningen: säker användning, effektiv kontroll och stöd för innovation [Adapting to the AI regulation: secure use, effective control and support for innovation]. SOU 2025:101. https://www.regeringen.se/contentassets/759c2d42b55843c587bfbc93540f4daa/anpassningar-till-ai-forordningen-sou-2025101.pdf

Department for Science, Innovation and Technology (DSIT). (2024). Introduction to AI assurance. https://www.gov.uk/government/publications/introduction-to-ai-assurance/introduction-to-ai-assurance

DSIT, Office for AI, & the Centre for Data Ethics and Innovation. (2019a, June 10). A guide to using artificial intelligence in the public sector. https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector

DSIT, Office for AI, & the Centre for Data Ethics and Innovation. (2019b, June 10). Assessing if artificial intelligence is the right solution. https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution

DSIT, Office for AI, & the Centre for Data Ethics and Innovation. (2019c, June 10). Planning and preparing for artificial intelligence implementation. https://www.gov.uk/guidance/planning-and-preparing-for-artificial-intelligence-implementation

DSIT, Office for AI, & the Centre for Data Ethics and Innovation. (2019d, June 10). Managing your artificial intelligence project. https://www.gov.uk/guidance/managing-your-artificial-intelligence-project

DSIT et al. (2020, June 8, refreshed 2025). Guidelines for AI procurement. https://www.gov.uk/government/publications/guidelines-for-ai-procurement

DSIT, Centre for Data Ethics and Innovation, Cabinet Office & Office for Artificial Intelligence. (2021, May 13). Ethics, Transparency and Accountability Framework for Automated Decision-Making. https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making

Duberry, J. (2022). Artificial intelligence and democracy: risks and promises of AI-mediated citizen–government relations. Edward Elgar. ISBN 978 1 78897 731 9.

Ellis-Petersen, H. (2024, May 20). Revealed: Meta approved political ads in India that incited violence. The Guardian. https://www.theguardian.com/world/article/2024/may/20/revealed-meta-approved-political-ads-in-india-that-incited-violence

Epstein, R., & Robertson, R. (2015, August 4). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. PNAS, 112(33): E4512–E4521. DOI: 10.1073/pnas.1419828112.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St Martin's Press.

European Commission. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Union. (2024a, September 2). Artificial Intelligence Act. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

European Union. (2024b, March 13) Regulation on the transparency and targeting of political advertising. https://eur-lex.europa.eu/eli/reg/2024/900/oj

Farquhar, A., Henry, N., Beaujouan, J., Wise, L., Hodge, J., Vancisin, T., Xia, W., Wilson, R., & Epple, T. (2024). Peace agreements in 2023: Insights from the PA-X Database [PA-X Data Series].

PeaceRep: The Peace and Conflict Resolution Evidence Platform, University of Edinburgh. https://peacerep.org/wp-content/uploads/2024/07/Peace-Agreements-in-2023-Insights-from-the-PA-X-Database-DIGITAL.pdf

Federal Communications Commission (FCC). (2024, May 23). FCC proposes $6 million fine for illegal robocalls that used Biden deepfake generative AI voice message. https://docs.fcc.gov/public/attachments/DOC-402762A1.pdf

Feldstein, S. (2019, April 22). How artificial intelligence systems could threaten democracy. The Conversation. https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1101&context=pubadmin_facpubs

Financial Times. (2024). TikTok probed by EU over alleged interference in Romania election. https://www.ft.com/content/268682a6-45a0-4cd8-b37b-c67a211455ac

Firchow, P., Martin-Shields, C., Omer, A., & Mac Ginty, R. (2017, February 1). PeaceTech: The liminal spaces of digital technology in peacebuilding. International Studies Perspectives, 18(1): 4–42. DOI: 10.1093/isp/ekw007.

Firera Colmenares, S.C., & Vakil, M. (2022). Bias in the context of artificial intelligence systems: Analysing the risks and contributors from a data perspective. Lund University. https://lup.lub.lu.se/student-papers/record/9091994/file/9091997.pdf

Freedom House. (2025). Freedom in the world 2025: The uphill battle to safeguard rights. https://freedomhouse.org/report/freedom-world/2025/uphill-battle-to-safeguard-rights

Freeman, M. (2025). "Fast Track Negotiation": A White Paper. A 21st century paradigm for negotiating peace and stability. Proceedings of the Paris Institute for Advanced Study. DOI: 10.5281/zenodo.14646340.

Garcia, S. (2024, September 19). Americans in both parties are concerned over the impact of AI on the 2024 presidential campaign. Pew Research Center. https://www.pewresearch.org/short-reads/2024/09/19/concern-over-the-impact-of-ai-on-2024-presidential-campaign/

Gambacorta, L., Jappelli, T., & Oliviero, T. (2025, October 10). Exploring household adoption and usage of generative AI: New evidence from Italy. BIS Working Papers, 1298. https://www.bis.org/publ/work1298.pdf

Garimella, K., & Chauchard, S. (2024, June 5). How prevalent is AI misinformation? What our studies in India show so far. Nature. https://www.nature.com/articles/d41586-024-01588-2

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé, III H., and Crawford, K. (2021, December 1). Datasheets for datasets: Documentation to facilitate communication between dataset creators and consumers. Communications of the ACM, 64(12): 86–92. DOI: 10.1145/3458723.

Giovardani, M. (2024). AI for peace: Mitigating the risks and enhancing opportunities. Data & Policy, 6: e41. DOI: 10.1017/dap.2024.37.

Goldman, S. (2025, July 10). How a deepfake of Marco Rubio exposed the alarming ease of AI voice scams. Fortune. https://fortune.com/2025/07/10/deepfake-marco-rubio-ai-voice-scams/

Goncharuk, V. (2024). Artificial intelligence in defence of Ukraine–Russia's war in Ukraine. Series 2, No. 6. International Centre for Defence and Security, Tallin. https://icds.ee/wp-content/uploads/dlm_uploads/2024/09/Layout-AI-in-Defence-of-Ukraine.pdf

Government Digital Service (GDS). (2023, January 5, updated 2025, May 8). Algorithmic Transparency Recording Standard (ATRS) Hub. https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub

GDS. (2024, December 17). Algorithmic Transparency Recording Standard (ATRS) mandatory scope and exemptions policy. https://www.gov.uk/government/publications/algorithmic-transparency-recording-standard-mandatory-scope-and-exemptions-policy/algorithmic-transparency-recording-standard-atrs-mandatory-scope-and-exemptions-policy

GDS. (2025a, February 10). Artificial Intelligence Playbook for the UK Government (HTML). https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html

GDS. (2025b, May 8). Algorithmic Transparency Recording Standard – guidance for public sector bodies. https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies

Government of Canada. (2025, June 24). Directive on automated decision-making. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

Government of Canada. (2026, January 7). Algorithmic Impact Assessment tool. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

Harroch, D., & Harroch, R. (2025, April 25).15 quotes on the future of AI. Time. https://time.com/partner-article/7279245/15-quotes-on-the-future-of-ai/

Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., and Zwitter, A. (2017, February 25). Will democracy survive big data and artificial intelligence? Scientific American. https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

Hiatt, J.M. (2006). ADKAR: A model for change in business, government and our community. Prosci. ISBN 978-1-930885-50-9.

Hirblinger, A.T. (2022, January 25). When mediators need machines (and vice versa): Towards a research agenda on hybrid peacemaking intelligence. International Negotiation. DOI: 10.1163/15718069-bja10050.

Hirblinger, A.T., Wählisch, M., Keator, K., McNaboe, C., Duursma, A., Karlsrud, J., Sticher, V., Verjee, A., Kyselova, T., Kwaja, C. M. A., and Perera, S. (2024). Forum: Making peace with un-certainty: Reflections on the role of digital technology in peace processes beyond the data hype. International Studies Perspectives 25(2): 185–225. DOI 10.1093/isp/ekad004.

Holroyd, M. & Olorunselu, F. (2022, March 16). 'Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war. EuroNews. https://www.euronews.com/my-europe/2022/03/16/deepfake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war

Horta Ribeiro, M., Ottoni, R., and Almeida, V. (2020, January 27). Auditing radicalisation pathways on YouTube Association for Computer Machinery. https://dl.acm.org/doi/abs/10.1145/3351095.3372879

Hwang, E. (2025, January 30). Lethal autonomous weapons: The next frontier in international security and arms control. Stanford International Policy Review. https://fsi.stanford.edu/sipr/content/lethal-autonomous-weapons-next-frontier-international-security-and-arms-control

Högbladh, S. (2021). Peace agreements in armed conflicts: Focusing on finding a solution to the conflict incompatibility. Pathways to Peace and Security, Report nr 2: 11–23. DOI 10.20542/2307-1494-2021-2-11-23.

Höne, K. (2019, November). Mediation and artificial intelligence: Notes on the future of international conflict resolution. DiploFoundation, Malta. DOI: 10.31752/idea.2021.91.

International IDEA. (2021). The global state of democracy 2021: Building resilience in a pandemic era. Stockholm. DOI: 10.31752/idea.2021.91.

International IDEA. (2022). The global state of democracy 2022: Forging social contracts in a time of discontent. Stockholm. https://idea.int/democracytracker/gsod-report-2022#note41during

International IDEA (2024). The global state of democracy 2024: Strengthening the legitimacy of elections in a time of radical uncertainty. Stockholm. https://www.idea.int/gsod/2024/

International IDEA (2025). Violence against women in the digital space: A growing threat to democracy. Stockholm. https://www.idea.int/blog/violence-against-women-digital-space-growing-threat-democracy

Jacobson, B.R., Höne K. E. and Kurbalija J. (2018, February). Data diplomacy: Updating diplomacy to the big data era. DiploFoundation, Malta. https://www.diplomacy.edu/wp-content/uploads/2019/07/Data_Diplomacy_Report_2018.pdf

Jeevanandam, N. (2025, January 24). Significance of AI on National Voters' Day: Strengthening democracy through innovation. IndiaAI. https://indiaai.gov.in/article/significance-of-ai-on-national-voters-day-strengthening-democracy-through-innovation

Kauer, T. & Vancisin, T. Messy peace processes. PA-X Analytics: Peace and Transition Process Data, University of Edinburgh. https://www.peaceagreements.org/visualizations/messy-peace-processes

Koh, E. (2021). The impact of trauma on peace processes. New England Journal of Public Policy, 33(1): article 4. https://scholarworks.umb.edu/nejpp/vol33/iss1/4

König, P., & Wenzelburger, G. (2020). Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. Government Information Quarterly, 37(3):101489,

ISSN 0740-624X. DOI 10.1016/j.giq.2020.101489

Kosmarski, A. (2020). Blockchain adoption in academia: Promises and challenges. Journal of Open Innovation: Technology, Market, and Complexity, 6(4): 117. DOI: 10.3390/joitmc6040117

Kuziemskia, M. & Misuracab, G. (2020, July). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications Policy, 44(6):101976, ISSN 0308-5961.

Leslie, David. (2019). Understanding artificial intelligence ethics and safety:
A guide for the responsible design and implementation of AI systems.
The Alan Turing Institute. DOI: 10.5281/zenodo.3240529

Liiv, I., Verhulst, S., Tauchnitz, E., Giovanardi, M., Nicolaïdis, K., & Wählisch, M. (2025, January). Data for peace: How novel data sources and technology can enhance peace? Data & Policy 7: e45. DOI: 10.1017/dap.2025.10006.

Liyih, A., Anagaw, a., Yibeyin, M. and Tehone, Y. (2024). Sentiment analysis of the Hamas-Israel war on YouTube comments using deep learning. Scientific Reports, 14: 13647. DOI: 10.1038/s41598-024-63367-3.

Lundgren, M., & Svensson, I. (2020, April). The surprising decline of international mediation in armed conflicts. Research & Politics, 7(2). DOI 10.1177/2053168020917243

Mac Ginty, R. & Silverman, C. (2021, January 14). Facebook has been showing military gear ads next to insurrection posts. BuzzFeed. https://www.buzzfeednews.com/article/ryanmac/facebook-profits-military-gear-ads-capitol-riot

Manheim, K.M. & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. Yale Journal of Law and Technology, VOLUME(ISSUE): 106–188. https://ssrn.com/abstract=3273016

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A.,

Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., et al. (2025, April). The 2025 AI Index Report. Institute for Human-Centered AI, Stanford University. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf

McAleer, K. (2023, October 6). Insights from Dr. Andrew Ng. Sutardja Center for Entrepreneurship & Technology, UC Berkley. https://scet.berkeley.edu/ai-is-the-new-electricity-insights-from-dr-andrew-ng/

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. (2019, January 29). Model cards for model reporting. FAT* '19: Proceedings of the conference on fairness, accountability, and transparency, pp 220–229. DOI: 10.1145/3287560.3287596.

Moshtagi, R., Cortez, J. & Sohn, A. (2025, May 27). AI and the future of conflict resolution: How can artificial intelligence improve peace negotiations? Belfer Center for Science and International Affairs, Harvard Kennedy School. https://www.belfercenter.org/research-analysis/ai-and-future-conflict-resolution-how-can-artificial-intelligence-improve-peace

Muller, C. (2020, March 20). The impact of artificial intelligence on human rights, democracy and the rule of law. Stichting ALLAI Nederland. https://allai.nl/wp-content/uploads/2020/06/The-Impact-of-AI-on-Human-Rights-Democracy-and-the-Rule-of-Law-draft.pdf

Murphy, M., Sharpe, E., & Huang., K. (2024, August 30). The promise of machine learning in violent conflict forecasting. Data & Policy 6: e35. DOI: 10.1017/dap.2024.27.

Murphy, G., Loftus, E.F., & Greene, C.M. (2019). False memories for fake news during Ireland's abortion referendum. Psychological Science, 30(10): 1449–1459. DOI: 10.1177/0956797619864887.

National Democratic Institute (NDI). (2022, October). Interventions for ending online violence against women in politics. NDI. https://www.ndi.org/sites/default/files/NDI%20Interventions%20to%20End%20OVAW-P.pdf

National Institute of Standards Technology (US) (NIST). (2023, January) Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100–1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Nordic Council of Ministers. (2018, May 14). AI in the Nordic-Baltic region https://www.stjornarradid.is/library/04-Raduneytin/ForsAetisraduneytid/Framtidarnefnd/AI%20in%20the%20Nordic-Baltic%20region.pdf

Osimen, G.W., Morola Fulani, O., Xhidozie, F., & Omolara Dada, D. (2024, July 23). The weaponisation of artificial intelligence in modern warfare: Implications for global peace and security. Research Journal in Advanced Humanities, 5(3). DOI: 10.58256/g2p9tf63.

Panic, B. (2024, October 15). Keynote lecture: AI for peace. PRIO AI Days. https://www.prio.org/events/9170

Panic, B., & Arthur, P. (2024). AI for peace. CRC Press. ISBN 9781032418377.

Paris Peace Forum. (2025). Phoenix: Social media tool for peacebuilders. https://parispeaceforum.org/projects/phoenix-social-media-tool-for-peacebuilders/

Ramsbotham, O., Woodhouse, T., & Miall, H. (2016). Contemporary conflict resolution (4th ed.). Polity Press.

Kozelnik, A. (17 October 2023). 'Parlameter: opening parliaments to people'. Innovation in Politics Institute: https://www.democracy-technologies.org/solutions/parlameter/

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press. ISBN 978–0674368279.

Persily, N. & Tucker, J.A., eds. (2020). Social media and democracy. Cambridge University Press. DOI: 10.1017/9781108890960.

Rathje, S., Van Bavel, J.J., & van der Linden, S. (2021, June 23). Out-group animosity drives engagement on social media. PNAS, 118(26): 1-9. DOI: 10.1073/pnas.2024292118.

Ravinutala, S., Botterill, M. & Khan, Z. (2025, March 13). ElectionGPT: How AI-powered tools supported India's elections—and what may be next for AI in the public sector. IDInsight. https://www.idinsight.org/article/electiongpt-how-ai-powered-tools-supported-indias-elections-and-what-may-be-next-for-ai-in-the-public-sector/

Researching Peace. (2023, October 13). #28 AI, peace, and conflict #1 – Conversations with Jim Dale and Mert Can Yilmaz [Audio podcast episode]. Uppsala University. https://researchingpeace.podbean.com/e/28-ai-peace-and-conflict-1-%E2%80%93-conversations-with-jim-dale-and-mert-can-yilmaz/

Responsible Technology Adoption Unit (RTA), DSIT. (2024, November 14). The model for responsible innovation. https://www.gov.uk/government/publications/the-model-for-responsible-innovation/the-model-for-responsible-innovation

Richmond, O.P., Visoka, G. & Tellidis, I. (2023, March 21). Peace in digital international relations: Prospects and limitations. Elements in International Relations. DOI: 10.1017/9781009396752.

Schippers, B. (2020, March). Artificial intelligence and democratic politics. Political Insight, 11(1): 32–35. DOI: 10.1177/2041905820911746.

Schreckinger, B. (2023, January 17). In Ukraine, war crimes go on-chain. Politico. https://www.politico.com/news/2023/01/17/in-ukraine-war-crimes-go-on-chain-00077993.

Silverman, C., Mac, R. & Dixit, P. (2020, September 14). 'I have blood on my hands': A whistleblower says Facebook ignored global political manipulation. BuzzFeed. https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo

SIPRI. (2025). Deployments fall more than 40% in a decade, as geopolitical tensions and funding cuts increase pressure on multilateral peace operations. https://www.sipri.org/media/press-release/2025/peace-operation-deployments-fall-40-per-cent?utm_source=chatgpt.com

Shahrzadi, L., Mansouri, A., Alavi, M., Shabani, A. (2024). Causes, consequences, and strategies to deal with information overload: A scoping review. International Journal of Information Management Data Insights, 4(2):100261. DOI: 10.1016/j.jjimei.2024.100261.

Stedman, S. J. (1997). Spoiler problems in peace processes. International Security, 22(2): 5–53. DOI: 10.2307/2539366.

Stewart, A.J., Mosleh, M., Diakonova, M., Arechar, A. A., Rand, D. G., and Plotkin, J. B. (2019, September 4). Information gerrymandering and undemocratic decisions. Nature, 573: 117–121. DOI: 10.1038/s41586-019-1507-6.

Sticher, V. (2024). War and peace in the age of AI. British Journal of Politics and International Relations, 27(2): 542-550. DOI: 10.1177/13691481241293066.

Sudmann, A. (2019). The democratisation of artificial intelligence: Net politics in the era of learning algorithms. In The democratisation of artificial intelligence: Net politics in the era of learning algorithms, edited by A. Sudmann. Transcript Verlag. pp. 9–32. DOI: 10.25969/mediarep/13537.

Tombe, S. P., Berlin, M., & Sisk, T. D. (2025). Peacemaking in crisis: Amid more wars, fewer peace agreements. Discusssion Paper 25-001. US Institute of Peace.

UNESCO. (2021, September 22). Social Media 4 Peace [video]. https://www.youtube.com/watch?v=T-TtQ8fWLb0

United Nations Department of Political and Peacebuilding Affairs and Centre for Humanitarian Dialogue (2019). Digital technologies and mediation in armed conflict. https://peacemaker.un.org/digitaltoolkit

University of Cambridge (2016, October 19). The best or worst thing to happen to humanity: Stephen Hawking launches Centre for the Future of Intelligence. https://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of

Varsha, P. (2023, April). How can we manage biases in artificial intelligence systems: A systematic literature review. International Journal of Information Management Data Insights 3(1). DOI: 10.1016/j.jjimei.2023.100165.

VIGINUM. (2023). Challenges and opportunities of artificial intelligence in the fight against information manipulation. https://www.sgdsn.gouv.fr/files/files/Publications/20250207_NP_SGDSN_VIGINUM_Rapport%20menace%20informationnelle%20IA_EN_0.pdf

vTaiwan. (2025). About vTaiwan. https://info.vtaiwan.tw/

Wired. (2022, March 17). A Zelensky deepfake was quickly defeated. The next one might not be. https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/

World Economic Forum. (2019, September 20). AI Government Procurement Guidelines. https://www.weforum.org/publications/ai-government-procurement-guidelines/

Yadav, B.L. (2025, May 4). ‘How AI is rewriting the rules of election campaign in India’. The Times of India. https://timesofindia.indiatimes.com/india/how-ai-is-rewriting-the-rules-of-election-campaign-in-india/articleshow/120848499.cms

Zuroski, N., Corrado, M., and Hume, L. (2023). Designing AI for conflict prevention and peacebuilding. Alliance for Peacebuilding. https://static1.squarespace.com/static/5db70e83fc0a966cf4cc42ea/t/6537c999d04a4d0e74bb5cd0/1729877182847/Designing+AI+for+Conflict+Prevention+%26+Peacebuilding.pdf

Endnotes

1 Höne, 2019
2 Transcript of Stephen Hawking's speech at the launch of the Leverhulme Centre for the Future of Intelligence. University of Cambridge, 2016
3 Shahrzadi, Mansouri, Alavi, and Shabani, 2024
4 University of Cambridge, 2016
5 Hwang, 2025; NATO, 2024
6 Maslej et al., 2025
7 Gambacorta et al., 2025
8 CEO of Microsoft Satya Nadella quoted in Harroch & Harroch, 2025
9 McAleer, 2023
10 Osimen et al., 2024
11 Panic and Arthur, 2024; Panic, 2024
12 Bell, 2024
13 Giovardani, 2024
14 Firera et al., 2022
15 Colmenares & Vakil, 2022; Varsha, 2023; Busiek, 2024
16 Sticher, 2024
17 Liiv et al., 2025
18 Department of Finance, Sweden, 2025; European Union, 2024
19 International IDEA, 2021; Coppedge et al., 2023
20 Sorkin, 2024
21 International IDEA, 2024
22 International IDEA, 2024
23 International IDEA, 2024
24 International IDEA, 2024
25 Demirjian, 2024
26 Goldman, 2025
27 Mac Ginty and Silverman, 2021; Rathje, Van Bavel and van der Linden, 2021; Horta Ribeiro et al., 2020
28 International IDEA, 2025
29 National Democratic Institute (NDI), 2022; Amnesty International, 2022; Murphy, Loftus, and Greene, 2019; Epstein and Robertson, 2015; Stewart et al., 2019; Beckman, Rosenberg, and Jebari, 2022; Muller, 2020: 12–1
30 UNESCO 2021; Silverman, Mac Ginty and Dixit, 2020
31 Barr, 2024
32 Garcia, 2024
33 Kuziemskia and Misuracab, 2020; König and Wenzelburger, 2020
34 Kuziemskia and Misuracab, 2020; König and Wenzelburger, 2020
35 Duberry, 2022: 14–39; Arandjelovic, 2021
36 International IDEA, 2022
37 Sudmann, 2019: 9–10; Christodoulou and Iordanou, 2021; International IDEA, 2022; Helbing et al., 2017; Duberry, 2022: 72–92; Brkan, 2019
38 Freedom House, 2025
39 Amodei and Hernandez, 2018
40 National Institute of Standards Technology (US) (NIST), 2023; European Union, 2024a; DSIT, 2024
41 DSIT, 2019b; Kotter, 1996; Hiatt, 2006
42 DSIT, 2019c; DSIT, 2021
43 Holroyd & Olorunselu, 2022; Wired, 2022

44 Benkler, Faris and Roberts, 2018; Persily and Tucker, 2020
45 AP News, 2025; FCC, 2024
46 Financial Times, 2024; European Commission, 2022
47 VIGINUM, 2025
48 National Institute of Standards Technology (US) (NIST), 2023; DSIT, 2021
49 Coalition for Content Provenance and Authenticity (C2PA), 2024
50 GDS, 2023; GDS, 2024
51 The White House, 2025; DSIT, 2019a
52 Yadav, 2025
53 Garimella & Chauchard, 2024
54 Ellis-Petersen, 2024
55 Jeevanandam, 2025
56 Ravinutala, Botterill and Khan, 2025
57 Amnesty International, 2021; IEEE Spectrum, 2022.
58 Eubanks, 2018; Pasquale, 2015
59 DSIT, 2019b; DSIT, 2020
60 Bürokratt RIA, 2025; Interoperable Europe, 2023
61 National Institute of Standards Technology (US) (NIST), 2023; DSIT, 2019a
62 Government of Canada, 2025; Government of Canada, 2026
63 GDS, 2025b
64 Mitchell et al., 2019; Gebru et al., 2021
65 Leslie, 2019
66 Benkler, Faris and Roberts, 2018
67 Wardle and Derakhshan, 2017
68 vTaiwan, 2025
69 Kozelnik, 2023
70 Barandiaran et al., 2024; Masakhane, 2020
71 CDEI, 2021
72 European Union, 2024a
73 National Institute of Standards Technology (US) (NIST), 2023
74 European Commission, 2022
75 European Union, 2024b
76 DSIT, 2024
77 Government of Canada, 2025; Mitchell et al., 2019; Gebru et al., 2021; GDS, 2024
78 Office for AI and World Economic Forum, 2019
79 Barandiaran et al., 2024; Masakhane, 2020
80 Kotter, 1996; Fernandez and Rainey, 2006; Weick and Quinn, 1999; Hiatt, 2006; DSIT, 2019c
81 DSIT, 2021
82 DSIT, 2019b; DSIT, 2019c; Leslie, 2019
83 Mitchell et al., 2019; Gebru et al., 2021; GDS, 2025a
84 CDEI, 2021; Office for AI and World Economic Forum, 2019
85 European Union, 2024a; National Institute of Standards Technology (US) (NIST), 2023; GDS, 2025b
86 DSIT, 2021; CDEI, 2021
87 Centre for Data Ethics and Innovation (CDEI), 2021
88 GDS, 2025b; RTA, 2024

89 National Institute of Standards Technology (US) (NIST), 2023; European Union, 2024a; DSIT,
90 Koh, 2021:11–14; Ramsbotham et al., 2016, Stedman,1997
91 Darby and Mac Ginty, 2003:2
92 Kauer and Vancisin, 2025
93 Darby and Mac Ginty, 2003
94 Freeman, 2025
95 Barltrop, 2025; Lundgren and Svensson, 2020
96 Tombe et al., 2025
97 Amiri, 2025
98 SIPRI, 2025
99 Tombe et al., 2025: 23
100 Jacobson et al., 2018
101 Höne, 2019
102 Researching Peace, 2023
103 Ashby, 2023
104 Grand-Clément, 2022
105 United Nations Department of Political and Peacebuilding Affairs and Centre for Humanitarian Dialogue, 2019
106 Delahoz-Domínguez et al., 2023
107 Goncharuk, 2024
108 Schreckinger, 2023
109 Kosmarski, 2020:1
110 Höne, 2019
111 Zuroski et al., 2023
112 Giovanardi, 2024
113 Nordic Council of Ministers, 2018
114 Moshtagi et al., 2025
115 Hirblinger, 2022
116 Hirblinger et al., 2024
117 Murphy et al., 2024
118 Richmond et al., 2023
119 Arana-Catania et al., 2022
120 Firchow et al., 2017
121 Paris Peace Forum, 2025
122 Liyih et al., 2024

Folke Bernadotte Academy (FBA) is the Swedish government agency for peace, security, and development. As part of Sweden's international development cooperation, we promote peace in conflict-affected countries. We offer training and advice and conduct research to strengthen peacebuilding and governance, in peace and security contexts. Moreover, we deploy civilian personnel to peace operations and election observation missions primarily led by the UN, EU and OSCE. The agency is named after Count Folke Bernadotte, the UN's first peace mediator.

For updates and latest publications, follow FBA on:

linkedIn.com/FolkeBernadotteAcademy

instagram.com/FolkeBernadotteAcademy

facebook com/FolkeBernadotteAcademy

fba.se/en

Folke Bernadotte Academy
Swedish agency for peace, security and development

FBA