64TH GRADUATE STUDY PROGRAMME

Compilation of the Group Works' Papers

United Nations Information Service, Geneva
July 2026

Foreword

The 64th Graduate Study Programme (GSP) was held at the Palais des Nations in Geneva, Switzerland from 29 June to 10 July 2026, bringing together 56 graduate students from 42 countries. Organized by the United Nations Information Service Geneva, GSP is the longest-running educational initiative of the United Nations. This report represents a collection of written works of the four student Working Groups.

Opinions, positions, statements, and conclusions expressed in the four reports included in this compilation are exclusively of their authors – graduate students who participated in the Programme. They do not necessarily represent or reflect the views of the United Nations, the group facilitators, or the facilitators' respective organizations.

Published in Geneva, Switzerland on 20 July 2026.

Table of Contents

WORKING GROUP 1

Medtech Market Entry Considerations in Low- and Middle-Income Countries

Facilitated by Jon Mark Walls

Authors:

Karabo Asala
Alan E. Avila Ramirez
George Berry
Karin Blommegård
Amelia Houana
Chelsea Molua
Artem Novobtritskii
Laura V. Ospina Ceron
Weile Qiu
Elena Szegedi
Papa El Hadji Keba Cire Toure
Katherine R. Yao
Denis Zaricinîi

Medtech Market Entry Considerations in Low- and Middle-Income Countries

Facilitated by: Jon Mark Walls

DISCLAIMER: dEEGtal is included in this report as an illustrative case study, not as the subject of a formal business recommendation. The company has voluntarily worked with us to help conceptualize the paper and provide context on AI-assisted EEG technology, but our analysis should not be read as a recommendation that dEEGtal enter Kenya, establish operations in an LMIC, or pursue a specific market-entry strategy.

1. INTRODUCTION AND CASE-STUDY OVERVIEW

AI has rapidly integrated our lives and is already transforming healthcare by improving access to diagnostics and personalized treatment, reducing clinical workload and enabling more informed decision-making (Fahim et al., 2025). Yet, research shows that only 2% of published articles on AI application to healthcare refer to LMICs (Fahim et al., 2025).

From the perspective of private AI MedTech companies, the reluctance to enter LMICs markets is totally understandable. We believe, however, that profitability and social impact should not be seen as competing objectives. Careful consideration of the local healthcare system, regulations, economic potential and cultural specificities can enable companies to establish profitable operations in LMICs, while contributing to the reduction of health inequities.

This paper, therefore, aims to establish a framework for AI MedTech companies on LMIC market entry, listing the most important aspects to consider in this process. As a case study, we analyze the potential entry of dEEGtal - a start-up developing an AI-assisted electroencephalography analysis tool for epilepsy diagnosis, into the Kenyan market.

2. PRINCIPLE MARKET ENTRY CONSIDERATIONS

2.1 Public Health	**2.2 Financing & Economic Feasibility**	**2.3 Technological Readiness**	**2.4 Political, Geopolitical, & Regulatory Risk**	**2.5 Sociocultural Context and Adoption Readiness**
2.1.1 Needs Assessment and Impact Analysis	2.2.1 Market Attractiveness.	2.3.1 Connectivity	2.4.1 Political and Governance	2.5.1 Gender and Social Norms
2.1.2 LMIC Co-Design and Capacity Building	2.2.2 Reimbursement Pathway	2.3.2 Data	2.4.2 Regulatory	2.5.2 Trust in the Healthcare System
2.1.3 Monitoring, Evaluation, and Adaptive Management	2.2.3 Procurement Pathway	2.3.3 Hardware		
	2.2.4 Macroeconomic risk			

2.1 Public Health

Any AI MedTech company entering an LMIC market must demonstrate how its technology serves the three core WHO health system goals: improving population health, providing financial and social risk protection, and ensuring responsiveness to population needs and expectations, while maintaining coverage and equity(Murray 2000). The following steps provide a structured pathway for responsible market entry:

2.1.1 Needs Assessment and Impact Analysis. A needs assessment establishes whether the technology is required by the populations and health systems it would serve. This is a distinct exercise from impact assessment and must come first. It should be conducted at three levels: patients, providers (including community health workers, who are frequently absent from assessments despite being the primary frontline interface), and systems. The needs of each group may differ significantly and must be understood independently before any alignment is attempted.

Once need is established, companies must model the impacts of introducing the technology. A rigorous impact assessment considers whether the technology will improve diagnosis and treatment outcomes; whether it will increase or reduce out-of-pocket costs; whether it will support or deskill health workers; and whether it will reach marginalised populations or deepen existing disparities. The World Economic Forum and London School of Economics 2026 white paper documents the de-skilling risk specifically: routine exposure to AI-assisted diagnosis can reduce clinicians' independent capabilities(van Kessel, 2026). In LMICs, where specialist capacity is already severely constrained, this risk is magnified and must be planned for explicitly.

2.1.2 LMIC Co-Design and Capacity Building. LMIC clinicians, patients, and community health workers must be embedded in the development process, not consulted after the fact. If development has already occurred, local adaptation is required before deployment. This is not a cosmetic exercise: an EEG analysis tool trained exclusively on European patient data may fail to recognise seizure patterns caused by neurocysticercosis or cerebral malaria, conditions common in sub-Saharan Africa but absent from Western training datasets. Local clinical validation, conducted in the target country on the target population, is the minimum standard for responsible deployment. Anything less risks deploying a tool that performs well in trials but fails patients in practice.

Capacity building must be built into market entry plans from the start, training programs for community health workers and clinicians, ongoing technical support, and development of local technical expertise that reduces long-term dependency. AI-enabled health systems also require a broader ecosystem: data scientists, medical device engineers, digital health specialists. Investing in this workforce creates sustainable employment and reduces structural dependency.

2.1.3 Monitoring, Evaluation, and Adaptive Management. Every deployment must be accompanied by a rigorous monitoring and evaluation framework tracking not only adoption metrics but meaningful health outcomes: diagnostic accuracy in the local population, treatment initiation rates, patient outcomes, equity of access, and workforce impacts. This reflects the core argument of the World Economic Forum and London School of Economics 2026 white paper: the winners of the AI race in health will be those who first identify and pursue outcomes that matter to patients and professionals and build the governance architecture to deliver them(van Kessel, 2026). Adaptive management requires a clear protocol for iteration, redesign, or exit. An honest exit plan is not a sign of failure; it is a sign of responsible deployment.

WHO frameworks, including the Global Strategy on Digital Health, SMART Guidelines, and the WHO/ITU Focus Group on AI for Health, and HealthAI's Global Regulatory Network provide the institutional infrastructure for each step (WHO, 2021; Mehl, 2021; ITU, 2025; HealthAI).

2.2 Financing and Economic Feasibility

Financial considerations extend beyond affordability to encompass the economic viability of market entry and long-term commercialization. This section covers financial feasibility through three dimensions: market attractiveness, reimbursement- , and procurement pathways.

2.2.1 Market Attractiveness. The first step for assessing entering a market is evaluating the epidemiological and diagnostic gap, i.e the market attractiveness. For AI-driven medical innovations, market assessment needs to extend beyond disease burden, as healthcare system capacity determines the realistically addressable market (Marwaha et al., 2022). Commercial potential depends on three dimensions: (1) disease burden and target population, (2) healthcare capacity, including infrastructure, equipment and skilled personnel, and (3) operational readiness, including digital infrastructure and workforce capabilities. Given limited primary data availability, market estimates should combine regional benchmarks, comparable countries, and scenario analysis to account for uncertainty (Chi, 2022).

2.2.2 Reimbursement Pathway. Financial sustainability in LMICs is closely linked to the reimbursement pathway. WHO and the World Bank highlight that over-reliance on out-of-pocket (OOP) payments limits financial protection and equitable access (WHO, 2025), making it an unsustainable model for scaling healthcare innovations. Sustainable business models therefore typically depend on pooled financing mechanisms, such as national health insurance, government procurement, or public–private partnerships (Africa CDC, 2025). Where reimbursement pathways are not yet established, demonstrating cost-effectiveness and broader health-system efficiency is critical. Positioning diagnostic AI as an investment that reduces downstream costs and improves resource allocation can strengthen adoption by Ministries of Health and Finance (Atiku & Olakotan, 2026).

2.2.3 Procurement Pathway. The medical procurement system indicates the route to the market and dictates the distribution and pricing strategy for the company. In centralized systems, securing a national tender can provide broad market access and predictable demand due to larger order volumes (OECD, 2000). However, this creates challenges including high entry barriers for smaller firms, dependence on a single buyer, potential reimbursement delays, and pricing pressure that may reduce margins. Missing a major tender can also delay access to the public market for years. Moreover, missing the tender might lock the company out of the public market for years. In decentralized systems, companies can engage directly with local providers and benefit from faster procurement processes, but must manage multiple purchasing decisions and fragmented market access (OECD, 2000).

2.4 Macroeconomic risk. In addition, macroeconomic factors such as inflation and exchange-rate fluctuations may increase the long-term cost of implementation, particularly where technologies rely on imported software or cloud-based services (World Bank, 2025).

2.3 Technological Readiness

Technological aspects are pivotal to consider for the market entry of a Medtech product. These are connectivity and infrastructure, the digitalisation of the public sector as well as the hardware facilities in health institutions.

2.3.1 Connectivity. While 58% of the world population are utilising mobile internet, differences persist for DCs moving from coverage to a usage gap as additional 38% have broadband connectivity but do not use it (Shanahan & Bahia 2025). Consulting the mobile connectivity index of GSMA for a chosen market entry can greatly facilitate medtech product considerations. Beyond broadband infrastructure, reliable energy access, in terms of fluctuations or outages, is crucial for powering medical equipment particularly those exceeding the urban landscape (Franco et al., 2017). The Multi Tier Framework (MTF) provides insights on a country´s energy status and whether the energy service is usable from a household perspective (Bhatia & Angelou, 2015).

2.3.2 Data. As AI MedTech technologies often require that health records are digitalized, the company's data environment needs to be understood. This is particularly important for training the model which makes the existence of local data essential. The WHO recommends the FAIR framework for data management and stewardship: data should be Findable, Accessible, Interoperable and Reusable (WHO, 2026; GO FAIR, 2022). Additionally, data protection regulations vary by country, both in the classification of health data as well as security requirements for the handling of that data. The data regulation framework therefore applied here is made up of three levels. **L1**: No centralised data protection regulation or sufficiently strict data protection regulations that make training and use of the technology very difficult to impossible. Minimal to no training data exists and cybersecurity of data holdings is minimal. **L2**: Centralised data protection regulation with additional strict data protection regulations, often treating health data as a special type of sensitive data (e.g. GDPR). Training data exists and is securely protected. **L3**: Centralised data protection regulation with health data considered as sensitive data. Training data is FAIR and securely protected, regularly undergoing phishing and cyber security testing.

2.3.3 Hardware. The hardware readiness framework maps out what a medical facility actually needs, physically and technologically, before a device like dEEGtal can be deployed. We split this into three levels. At **L1, Absence of Infrastructure**, a facility has no formal power distribution, no reliable fuel supply chain, and often not even piped water. Rural facilities sit here far more often than urban ones (WHO, 2023), and this is where hardware medtech implementation meets the most resistance, not because the clinical need is lower, but because there's nothing to build on yet. **L2, Basic Equipment**, is the point where a power core, generators and separated life-safety, critical, and equipment branches (Trystar, 2023), exists alongside a network core: routers, switches, firewalls, and proper hospital cabling [2]. Once these two cores are in place, implementation becomes realistic, provided the local hospital can support it. **L3, Fully Equipped**, is an established clinical ecosystem: defibrillators, patient monitors, surgical instruments, and diagnostic tools including X-ray, ultrasound, MRI, CT, ECG, and EEG (WHO, n.d.) (AliMed, 2024). This is where a new device faces the least resistance on entry, since the surrounding infrastructure already exists. Mapped onto this, the EEG-specific pathway looks like this: at L1, rural areas with no power supply, deployment isn't viable at all. At L2, power and network cores are in place, so the device can run on generator or battery power and transmit over the hospital's existing connectivity (scalp electrodes → computer/monitors → clinical response). At L3, the full ecosystem is present and EEG becomes part of the wider diagnostic stack, feeding into a proper e-health platform, clinical response, and continuous monitoring loop.

2.4 Political, Geopolitical, and Regulatory Risk

2.4.1 Political and Governance. The effectiveness of the government is directly tied to the level of corruption of a country, which can involve embezzlement, nepotism, and extortion (Ministry of Justice, n.d.). Economists, public policy makers, and human rights experts should be consulted, but "street level” bureaucrats such as social workers and public healthcare professionals who will be involved with the product in any capacity should also advise the implementation of the product (Runst, 2021). Importantly, the foreign relationship between the country of the founding company and the target country should be kept in mind when determining market entry fit. Sanctions, geopolitical restrictions, and other foreign ownership constraints result in the home country and host country (Council of the European Union, 2026). Current world events and their predicted impacts in the future, related to the home or host country, should be analyzed and a plan should be developed to mitigate risks and barriers for market entry.

2.4.2. Regulatory. Market entry for an AI MedTech company requires a clear understanding of the target country’s legal and regulatory environment, including privacy and cybersecurity governance, foreign market entry rules, medical software licensing, labor regulations, tax/reporting obligations, and environmental compliance (Pharmacy and Poisons Board, n.d.).

2.5 Cultural Context and Considerations

2.5.1 Gender Norms. For a MedTech company entering a new market, gender dynamics matter because household decision-making power can shape who is able to seek diagnosis, pay for care, and follow through with treatment.

2.5.2 Trust in the Healthcare System. Trust in the healthcare system also matters because patients and clinicians may hesitate to adopt biomedical or AI-based tools if they conflict with local beliefs, prior experiences, or concerns about data and accountability (Azad et al., 2020). If these social factors are ignored, an AI diagnostic tool may fail to reach the people most in need, especially when household decision-making power, fear of stigma, or distrust of biomedical and AI-based care limits access to formal clinical services (Azad et al., 2020).

3. COUNTRY APPLICATION OF FRAMEWORK (KENYA)

According to the World Health Organization (WHO), epilepsy is one of the most common neurological disorders worldwide, affecting an estimated 50 million people. Moreover, 80% of those affected live in low- and middle-income countries (WHO, 2024). Yet, the treatment gap of epilepsy in medium- and low-income settings is above 50% and 75% respectively (The Lancet Regional Health – Western Pacific, 2024).

These statistics underline the importance of timely and accurate diagnosis of epilepsia, since a patient who is never diagnosed is never treated. Classifying epilepsy, however, is an inherently complex process, requiring access to reliable testing tools and professional expertise. In middle- and low-income contexts, therefore, due to the lack of resources, diagnosis of epilepsy might be limited (The Lancet Regional Health – Western Pacific, 2024).

An established seizure etiology and abnormal electroencephalography (EEG) findings are the primary predictors of seizure recurrence (WHO, 2024). The EEG records the brain's electrical activity through scalp electrodes and can pick up abnormal spiking patterns (Mayo Clinic, 2024). dEEGtal’s AI tool works alongside a conventional EEG recording, flagging abnormal patterns, including subtle seizure activity that's easy to miss on manual review, to make interpretation faster and more consistent. To make the underlying technology legible to a non-specialist audience, regulators, hospital administrators, policymakers, we also built a companion website walking through what EEG is and what dEEGtal's tool actually consists of. *(Link: https://alanavilarmz.github.io/UNGSP_2026_WP1_EEGuide/)*

Kenya was chosen as a case study, since despite being recognized as Africa's leading technology hub in 2025, it has a relatively small neurology workforce (Aderinto, 2023). It therefore provides a good example of a potential destination for dEEGtal and for applying the framework described in Section 2.

3.1 Public Health

Epilepsy affects an estimated 500,000 Kenyans, with an 80% diagnostic gap in Nairobi informal settlements and a treatment gap of 62–70% in rural areas. The burden falls hardest on adolescents aged 13–18, unemployed individuals, and those with no formal education. Non-convulsive epilepsy, already at a 100% diagnostic gap, is systematically invisible to current clinical systems (Kinyanjui et al., 2024; Ngugi et al., 2012).

Applying the framework to dEEGtal's Kenyan market entry reveals an immediate structural tension at Step 1. dEEGtal's current product requires connection to an EEG machine to enhance diagnostic accuracy. Kenya has fewer than 50 EEG machines nationally, almost all concentrated in Nairobi and a handful of county referral hospitals (WHO Regional Office for Africa, 2023). The 75–80% treatment gap lives in rural counties where no EEG machine exists within 200 kilometres (Ngugi et al., 2012; Sokhi et al., 2016). A tool that requires EEG connectivity therefore reaches approximately 5–10% of patients with epilepsy, those already closest to specialist care, while leaving the majority entirely unreached. This is the value inversion the WEF/LSE 2026 white paper describes: optimising for what is measurable rather than what matters (van Kessel et al., 2026).

The framework surfaces three options: tertiary deployment in urban facilities as a clinically valuable but equity-limited tool; redesign toward a lightweight community-level version operable without EEG on basic smartphones by community health workers; or a staged approach combining both. Kenya's Digital Health Act 2023 (Republic of Kenya, 2023), its 100,000-strong community health volunteer network (WHO Regional Office for Africa, 2023), and KEMRI's clinical research infrastructure provide the institutional foundation for whichever path is chosen. Kenya's devolved health system, 47 county governments each with independent health budgets, means national regulatory approval is necessary but not sufficient: county-level engagement is required for any deployment to reach scale.

3.2 Financial

3.2.1 Market Attractiveness: Epilepsy affects an estimated 500,000 Kenyans, with an 80% diagnostic gap in Nairobi informal settlements and a treatment gap of 62–70% in rural areas. The burden falls hardest on adolescents aged 13–18, unemployed individuals, and those with no formal education. Non-convulsive epilepsy, already at a 100% diagnostic gap, is systematically invisible to current clinical systems (Kinyanjui et al., 2024; Ngugi et al., 2012). However, the potential user base of the product extends beyond confirmed epilepsy cases to include patients presenting with seizure-like symptoms, unexplained loss of consciousness, episodic confusion, and other paroxysmal events — all of which require EEG to determine whether epilepsy is present or to rule it out. Based on Kenyan EEG data, approximately 30% of recordings show abnormal findings, while 70% are normal, suggesting that EEG workflows include both epilepsy-related diagnosis and broader neurological assessment (Lehn-Schiøler et al., 2026). Using this distribution as an assumption, the broader diagnostic population could represent approximately 1.6 million potential clients. Assuming 50 EEG machines in Kenya and a throughput of 1,250–3,000 studies per machine per year, the national EEG market can serve approximately 62,500–150,000 patient studies annually, or about 100,000 in the base case (Miskin et al., 2015).

3.2.2 Reimbursement pathway: Long-term financial sustainability for dEEGtal will depend on integration into Kenya's evolving pooled health financing system. The Social Health Authority (SHA), established under the Social Health Insurance Act (2023), replaces the National Health Insurance Fund (NHIF) and aims to expand universal health coverage through strategic purchasing and pooled financing (Government of Kenya, 2024; Government of Kenya, 2025). This creates a more favorable environment for digital diagnostics, as payment can shift away from out-of-pocket spending toward institutional reimbursement. However, reimbursement pathways for AI-assisted diagnostics remain under development, making evidence of cost-effectiveness and health-system efficiency critical for adoption.

3.2.3 Procurement pathway: Kenya's healthcare procurement is largely centralized, with county governments playing a significant role in purchasing medical supplies and services (Mirangi Kiambo et al., 2024). To access the public market, AI diagnostic solutions require regulatory approval through the Pharmacy and Poisons Board. International regulatory approvals, such as US FDA authorization, may support a more streamlined review process (Pharmacy and Poisons Board, n.d.). Successful market entry will therefore depend on navigating both regulatory approval and public procurement processes.

3.3 Technological Readiness

3.3.1 Connectivity: In Sub-Saharan Africa, Kenya is among the leading nations on health technology innovations, particularly as the country has a widespread mobile phone utilisation enabling accessibility to the mobile health platform (mHealth). Disparities remain between urban and rural contexts on internet connectivity, high implementation costs and a lack of public awareness on digital health solutions (Agbeyangi & Lukose, 2025). The mobile connectivity index demonstrates Kenya's advanced role in the region (GSMA). Datapoints such as the network coverage of 98 out of 100 emphasise the country's wide 4G population coverage. For the EEG reading, connectivity is negligible. Only when communication occurs between the headquarters and a team on-site, mobile broadband presents importance. Yet, reliable energy access remains a crucial part of the EEG reading and dEEGtal's analysis. Kenya's MTF report reveals a clear urban-rural disparity of household energy access as well as regular energy disruptions via outages or other voltage fluctuations (Dubey et al., 2020). Bearing in mind to reassess and investigate the latest data.

3.3.2 Data: Kenya's data protection regulation defines health as sensitive data under the Data Protection Act (2021), which provides guidelines similar to EU's GDPR (2016) but without the strict technical requirements. Additional regulations are in progress that require that datasets for AI training be representative of the intended patient population. Health records are digitalized as part of the Digital Health Act (2023), and have been previously used to develop other AI MedTech technologies, such as the deployment of AI-powered X-rays in 2025 (Kenya Pharmacy and Poisons Board, 2025). An initial assessment of Kenya's data environment would therefore be L3, dependent on the availability of the EEG data and existence of cybersecurity measures.

3.3.3 Hardware: Access to medical hardware in Kenya varies significantly between urban and rural populations. In cities, hospitals such as Nairobi Hospital and Aga Khan University Hospital operate dedicated EEG clinics, while rural areas largely lack access to this technology (). Smartphone-based EEGs have been piloted to extend access to rural areas but are not yet adopted (Enkhtsetseg et al., 2026). Because of this disparity, the hardware assessment ranges from L1 to L3 depending on region.

3.4. Political and Regulatory Risk: Kenya appears moderately ready for dEEGtal's market entry because it has recently strengthened its digital health infrastructure through the Digital Health Act, which created a national framework for health data governance, interoperability, and digital health services (Digital Health Act, 2023). However, because dEEGtal is AI-powered medical software for EEG analysis, the company would need to comply with Kenya's medical device software regulations, including requirements related to safety, cybersecurity, data protection, AI/ML performance, and post-market monitoring (Pharmacy and Poisons Board, 2026). dEEGtal should consult Kenyan health regulators, neurologists, hospital administrators, data-protection experts, and frontline healthcare workers because implementation will depend not only on formal approval, but also on hospital workflows, clinician trust, affordability, and local EEG capacity (Runst, 2021). Overall, Kenya is a promising market, but dEEGtal should enter through a regulated pilot that addresses clinical validation, patient data privacy, and accountability for AI-supported diagnosis. For dEEGtal, these rules would shape not only whether the company can legally operate in Kenya, but also how it stores patient EEG data, hires or partners locally, secures regulatory approval, reports revenue, and manages any hardware-related environmental responsibilities.

3.5 Cultural Considerations:

3.5.1 Gender norms: Gender dynamics represent a significant but often overlooked barrier to equitable healthcare access and by extension, to the effective deployment of AI diagnostic tools. In Kenya, as across much of sub-Saharan Africa, men typically hold decision-making authority over household health expenditures. Women are the primary caregivers responsible for identifying and monitoring illness but face limited autonomy due to male authority and financial control, with men's roles as household heads and financial providers granting them ultimate decision-making power over healthcare access (Azad et al., 2020). If a tool requires a formal clinical setting, referral from a provider or out-of-pocket or insurance-backed expenditure, this dynamic is directly relevant. A woman or female caregiver who recognizes a seizure disorder may not have the financial autonomy to pursue diagnostic services without male household authorization (Azad et al., 2020).

Inaccurate beliefs about the causes of epilepsy, stigma, treatment costs, distance from care, life stage and gender also play a role in the massive epilepsy treatment gap (Wabulya et al., 2024). Epilepsy in particular carries profound social stigma across East African contexts. Culturally driven health-seeking behaviours contribute to a low rate of epilepsy diagnosis, as well as poor access to treatment and are thought to be key contributors to the treatment gap (Samia et al., 2019). Fear of stigma is likely to discourage those affected from seeking medical attention, with children and adolescents being disproportionately affected (Samia et al., 2019). Stigma manifests as social exclusion, barriers to healthcare and reduced quality of life, disproportionately affecting vulnerable groups such as women, the less educated and those with early-onset epilepsy (Alege et al., 2026).

For a medical AI tool, these dynamics have concrete implications. A diagnostic tool that confirms epilepsy may inadvertently expose patients to social stigma, creating perverse disincentives to seek diagnosis in the first place. Deployment strategies must therefore incorporate community sensitization and stigma-reduction programming, not as an add-on, but as a prerequisite to clinical rollout. Engaging community health volunteers to provide culturally grounded education about epilepsy's neurological causes before a diagnostic tool is introduced could be a great first step in shifting the beliefs that drive avoidance.

3.5.2 Trust in the Healthcare System: Trust in a formal healthcare system is complex and layered. Traditional medicine remains the first point of care for up to 80 per cent of populations in sub-Saharan Africa, largely due to affordability and cultural acceptance (Wanza, 2026). In the context of epilepsy specifically, failure to seek biomedical treatment was associated with holding traditional animistic religious beliefs, reporting negative attitudes about biomedical treatment and living more than 30 kilometres from health facilities (Mbuba et al., 2012).

Trust in AI-mediated diagnostics introduces a further layer of complexity beyond trust in biomedicine generally. Lower scores were reported for regular utilization of AI disease prediction capabilities and staff trust in AI-generated predictions, highlighting the need for greater staff sensitization (Tira & Kihara, 2025). If clinicians themselves harbour scepticism about AI-assisted EEG interpretation, patient-level trust is even less likely to follow automatically (Tira & Kihara, 2025), an implication that is important for dEEGtal.

Kenya's evolving digital health ecosystem offers both cause for cautious optimism and pointed warnings. Kenya's 2025 draft AI Strategy explicitly identifies healthcare as a priority sector and stresses data sovereignty, yet health communication and trust-building remain underdeveloped (Addie & McNealy, 2025). Critically, none of the existing frameworks explicitly engage with epistemic justice and the question of whose knowledge is legitimized and how cultural expertise is incorporated into systems (Addie & McNealy, 2025). A tool developed in Geneva like dEEGtal that is validated on European or North American datasets and deployed by foreign operators risks being perceived as an extension of what scholars describe as digital colonialism: technological systems that extract data and impose frameworks without genuine local co-ownership (Addie & McNealy, 2025).

Implications for dEEGtal: Trust cannot be assumed or purchased through marketing. It must be built through co-design with Kenyan clinicians, community health volunteers and patient communities, through transparent data governance that guarantees data sovereignty; through local clinical validation; and through honest communication about what dEEGtal can and cannot do (Wabulya et al., 2024; Smith-Sreen et al., 2026). A phased pilot approach, beginning with trusted tertiary institutions like Kenyatta National Hospital or Moi Teaching and Referral Hospital, where digital health innovations have been previously piloted, would allow trust to develop organically before broader rollout is attempted.

4. CONCLUSION

In this report, bringing medical technology into LMICs is not solely a question of whether a technology is functional. The questions should also explore how the surrounding systems, clinical facilities, financial status, technological readiness in connectivity, hardware, and data, but also cultural matters, are part of the integration of new ideas, concepts, and technology.

In well-developed ecosystems, an AI tool is more of a new layer added to the workflow, but in an LMIC, it has to work alongside gaps in infrastructure, financing, and regulation, as well as the cultural and social constraints of the regions of interest. Therefore, the question is not whether the technology works, but rather whether the system around it is ready to hold it. Based on this report, several factors have to line up.

As such, dEEGtal, a Swiss-based start-up, served as an example to test these new AI technologies. The case explores the possibility of implementation in a country such as Kenya, which has access to fewer than 50 EEG machines at a national level, almost all of them in Nairobi. A large share of the treatment burden falls on rural counties with no EEG access at all, whereas an urban approach appears to offer a more immediate possibility for integration.

dEEGtal's financial sustainability depends on integrating into Kenya's financing system and clearing both regulatory approval and procurement processes, none of which are guaranteed, given that AI-diagnostic reimbursement pathways are still developing. Even if financed, the tool creates no value if the population most affected can't reach a facility that has it.

From a cultural perspective, the problem of adoption does not come from technical accuracy. The real barrier may be stigma, health spending, and layered distrust (e.g., western medicine, AI, and foreign-developed tools), all of which can suppress uptake regardless of diagnostic performance. Responsible deployment requires treating community trust-building and local co-design as prerequisites.

Therefore, in conclusion, the integration of novel technologies in LMICs requires a more holistic analysis, one that includes not just the innovation perspective but also how financial, cultural, and infrastructural aspects shape the integration of these new tools.

References

Addie, Y.O. & McNealy, J. (2025). Lessons from Nigeria and Kenya on digital colonialism in AI health messaging. TechPolicy.Press. https://www.techpolicy.press/lessons-from-nigeria-and-kenya-on-digital-colonialism-in-ai-health-messaging/

Addressing the epilepsy treatment gap in low resource settings. The Lancet Regional Health - Western Pacific, 45. https://doi.org/10.1016/j.lanwpc.2024.101081

Aderinto, N. (2023). Unlocking the hidden burden of epilepsy in Africa: Understanding the challenges and harnessing opportunities for improved care. Health Science Reports. https://onlinelibrary.wiley.com/doi/full/10.1002/hsr2.1220

Africa CDC. (2025). Africa’s Health Financing in a New Era. Concept Paper. https://africacdc.org/download/africas-health-financing-in-a-new-era-april-2025/

Africa's Health Financing in a New Era April, 2025 — Africa Health Knowledge Hub. (n.d.). Africa CDC. Retrieved July 9, 2026, from https://khub.africacdc.org/records/resource/africas-health-financing-in-a-new-era-april-2025

Aga Kahn University Hospital Neurology Department https://hospitals.aku.edu/nairobi/Departments/Pages/Neurology.aspx

Alege, B., Agbo, C., & Olagunju, A. (2026). Epilepsy-related stigma in Nigeria: A systematic review of manifestations, impacts, and socio-cultural drivers. African Journal of Disability, 15, 8 pages. doi:https://doi.org/10.4102/ajod.v15i0.1875

AliMed. (2024). *A brief guide to hospital equipment*. https://www.alimed.com/blogs/clinical-furnishings/a-brief-guide-to-hospital-equipment

Atiku S, Olakotan O. (2026). Digital Health and Fiscal Credibility in Low- and Middle-Income Countries: A Scoping Review of Practice-Based Evidence. J Eval Clin Pract. 2026 https://pmc.ncbi.nlm.nih.gov/articles/PMC12867509/pdf/JEP-32-0.pdf

Azad, A. D., Charles, A. G., Ding, Q., Trickey, A. W., & Wren, S. M. (2020). The gender gap and healthcare: associations between gender roles and factors affecting healthcare access in

Central Malawi, June-August 2017. Archives of public health = Archives belges de sante publique, 78(1), 119. https://doi.org/10.1186/s13690-020-00497-w https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7672876/

Bhatia, M., & Angelou, N. (2015). Beyond connections: Energy access redefined (ESMAP Technical Report 008/15). World Bank. https://openknowledge.worldbank.org/server/api/core/bitstreams/3eb83026-2c51-49fd-b8c4-3d1aca9ad419/content

Ba-Diop, A., et al. (2024). Addressing the epilepsy treatment gap in low resource settings. The Lancet Regional Health – Western Pacific. https://doi.org/10.1016/j.lanwpc.2024.101081

Chi, G. (2022). Micro-Estimates of Wealth for all Low- and Middle-Income Countries. Proceedings of the National Academy of Science, 119. https://www.pnas.org/doi/10.1073/pnas.2113658119

Council of the European Union. (2026, June 8). Foreign direct investment screening explained. https://www.consilium.europa.eu/en/policies/fdi-screening-explained/

Data Protection Act, 2021 (Kenya). https://www.odpc.go.ke/data-protection-laws-kenya/

Digital Health Act, No. 15 of 2023 (Kenya). https://new.kenyalaw.org/akn/ke/act/2023/15/eng@2023-11-24

Digital Health Act, No. 15 of 2023 (Kenya). https://new.kenyalaw.org/akn/ke/act/2023/15/eng@2023-11-24

Enkhtsetseg, N., Lehn-Schiøler, W., Storgaard, A. M., Pedersen, M. G., Rice, D., Wambugu, G., Fidele, N. J., Hribljan, M. C., Arbune, A. A., Larsen, S. A., Beniczky, S., & Mateen, F. J. (2026). Clinical utility and feasibility of smartphone-based EEG in Kenya: A multicenter observational study. arXiv. https://arxiv.org/abs/2605.08157

Fahim, Y. A., Hasani, I. W., Kabba, S., & Ragab, W. M. (2025). Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. European Journal of Medical Research. https://doi.org/10.1186/s40001-025-03196-w The Lancet Regional Health – Western Pacific. (2024).

Franco, A., Shaker, M., Kalubi, D., & Hostettler, S. (2017). A review of sustainable energy access and technologies for healthcare facilities in the Global South. Sustainable Energy Technologies and Assessments, 22, 92-105.

GDPR Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1. https://eur-lex.europa.eu/eli/reg/2016/679/oj

GO FAIR. (2022). FAIR principles. https://www.go-fair.org/fair-principles/

Government of Kenya. (2024). Social Health Insurance Act, 2023 (implementation information available through the Social Health Authority). https://www.sha.go.ke/

Government of Kenya. (2025). Digital Health (Health Information Systems) Regulations, 2025. https://new.kenyalaw.org/akn/ke/act/ln/2025/77/eng%402025-04-11

HealthAI. (n.d.). *Global Regulatory Network*. Retrieved July 10, 2026, from https://healthai.agency/grn/

International Telecommunication Union (ITU). (2025). *FG-AI4H: Focus Group on Artificial Intelligence for Health*. Retrieved July 10, 2026, from https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx

van Kessel R, et al. Meaningful Outcomes Determine the Winners of the Health AI Race. WEF / LSE Health, June 2026.

Kinyanjui DW, et al. Prevalence of all epilepsies in urban informal settlements in Nairobi, Kenya. Lancet Global Health, 2024.

Kipruto, I., & Mwathi, E. (2025, October 14). Kenya rolls out AI-powered X-ray machines to boost fight against tuberculosis and lung diseases. Health Business. https://healthbusiness.co.ke/9560/kenya-rolls-out-ai-powered-x-ray-machines-to-boost-fight-against-tuberculosis-and-lung-diseases/

Lehn-Schiøler, W., Enkhtsetseg, N., Storgaard, A. M., Pedersen, M. G., Rice, G., Wambugu, G., Fidele, N. J., Hribljan, M. C., Arbune, A. A., Larsen, S. A., & Beniczky, S. (2026). Clinical Feasibility of Smartphone-based EEG in Kenya. preprint. https://arxiv.org/pdf/2605.08157

Marwaha, J. S., Landman, A. B., Brat, G. A., Dunn, T., & Gordon, W. J. (2022). Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. npj Digital Medicine, 5(1). https://www.nature.com/articles/s41746-022-00557-1

Mayo Clinic. (2024, May 29). EEG (electroencephalogram). Mayo Clinic. Retrieved July 9, 2026, from https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875

Mbuba, C. K., Ngugi, A. K., Fegan, G., Ibinda, F., Muchohi, S. N., Nyundo, C., Odhiambo, R., Edwards, T., Odermatt, P., Carter, J. A., & Newton, C. R. (2012). Risk factors associated with the epilepsy treatment gap in Kilifi, Kenya: a cross-sectional study. The Lancet. Neurology, 11(8), 688–696. https://doi.org/10.1016/S1474-4422(12)70155-2

Mehl, G., Tunçalp, Ö., Ratanaprayul, N., Tamrat, T., Barreix, M., Lowrance, D., Bartolomeos, K., Say, L., Kostanjsek, N., Jakob, R., Grove, J., Mariano, B., Jr., & Swaminathan, S. (2021). WHO SMART guidelines: Optimising country-level use of guideline recommendations in the digital age. *The Lancet Digital Health*, *3*(4), e213–e216. https://doi.org/10.1016/S2589-7500(21)00038-8

Ministry of Justice. (n.d.). Examples of corruption. Anti-Corruption.fi. Retrieved July 7, 2026, from https://korruptiontorjunta.fi/en/examples-of-corruption

Miskin, C., Carvalho, K. S., Valencia, I., Legido, A., & Khurana, D. S. (2015). EEG Duration. The Long and the Short of It. J Child Neurol, 30(13). https://pubmed.ncbi.nlm.nih.gov/25814476/

Mirangi Kiambo, E., Murira, F., & Pattnaik, A. (2024). “Financing Essential Medicines and Supplies in Kenya.” [Financing Supplies Factsheet 1]. Health Systems Insight. https://thinkwell.global/wp-content/uploads/2024/11/Kenya-factsheet-11.19.24.pdf

Most countries make progress towards universal health coverage, but major challenges remain, WHO–World Bank report finds. (2025, December 6). World Health Organization (WHO). Retrieved July 9, 2026, from https://www.who.int/news/item/06-12-2025-most-countries-make-progress-towards-universal-health-coverage-but-major-challenges-remain-who-world-bank-report-finds

Murray, C. J. L., & Frenk, J. (2000). A WHO framework for health system performance assessment. *Bulletin of the World Health Organization*, *78*(6), 717–731.

Nairobi Hospital Neurophysiology services https://www.neurophysiologyservices.co.ke/clinical-physiology-2/

Ngugi AK, et al. Risk factors associated with the epilepsy treatment gap in Kilifi, Kenya. Lancet Neurology, 2012.

Odhiambo, J., & Barsoget, J. (2024). Robotic Medical Services and the Future of Healthcare in Kenya.

OECD (2000),Centralised and Decentralised Public Procurement, SIGMA Papers, No. 29,OECD Publishing, Paris, https://doi.org/10.1787/5kml60w5dxr1-en.

Pharmacy and Poisons Board. (2026, April 16). Guideline on regulation of medical device software in Kenya (MDSW). https://web.pharmacyboardkenya.org/download/guideline-on-regulation-of-medical-device-software-in-kenya-mdsw/

Republic of Kenya. Digital Health Act No. 15 of 2023. Kenya Law.

Runst, P. (2021). A case study of bureaucratic discretion: Heterogeneous application of market entry regulation in Germany. Journal of Institutional Economics, 17(4), 681–700. https://doi.org/10.1017/S1744137421000096

Samia, P., Hassell, J., Hudson, J. A., Murithi, M. K., Kariuki, S. M., Newton, C. R., & Wilmshurst, J. M. (2019). Epilepsy diagnosis and management of children in Kenya: review of current literature. Research and reports in tropical medicine, 10, 91–102. https://doi.org/10.2147/RRTM.S201159

Shanahan, M., & Bahia, K. (2025). The state of mobile internet connectivity 2025: Overview report. GSMA. https://www.gsma.com/somic/wp-content/uploads/2025/09/The-State-of-Mobile-Internet-Connectivity-2025-Overview-Report.pdf

Sibiya, S. E., Hurchund, R., Omondi, B., & Owira, P. (2025). Artificial intelligence for digital healthcare in the low and medium income countries. Health and Technology, 15, 323–332. https://doi.org/10.1007/s12553-025-00950-2

Smith-Sreen, J., Timothy, B., Ngila, B., Maina, J. W., Pirirei, S., Kinuthia, J., Bukusi, D., Waweru, H., Bosire, R., Farquhar, C., Mello, M. J., & Aluisio, A. R. (2026). Acceptability and use determinants of digital health technologies for HIV services: a qualitative study of emergency care patients in Nairobi, Kenya. Frontiers in digital health, 7, 1697814. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12876169/

Sokhi D, Diaz M, Ngugi A, et al. Epilepsy Prevalence, Treatment Gap, and Stigma in Western Kenya. Neurology, 2016.

Tira, Corretta & Kihara, Allan. (2025). Effect of Artificial Intelligence Adoption on Performance of Healthcare Sector in Kenya: A Case of Nairobi Public Hospitals. Journal of Business and Strategic Management. 10. 31-52. 10.47941/jbsm.3150.

Trystar. (2023). *Hospital emergency power requirements*.

van Kessel R, et al. Meaningful Outcomes Determine the Winners of the Health AI Race. WEF / LSE Health, June 2026.

Wabulya, A., Odero, N. A., Kimathi, A. M., Kwobah, E. K., Malawsky, D. S., Agans, R., ... & Shiloh-Malawsky, Y. (2024). Epilepsy Care in Kenya: Gaps and Opportunities from Targeted Survey of Health Workers. Medical Research Archives, 12(10). https://esmed.org/MRA/mra/article/view/5923

Wanza, P. (2026). Kenya to integrate traditional medicine into national health system. KBC News. https://www.kbc.co.ke/kenya-to-integrate-traditional-medicine-into-national-health-system-oluga-says/

World Health Organization. Regional Office for Africa. Kenya: Strengthening the Health Workforce. Based on Kenya 2023 Health Labour Market Analysis

World Health Organization. (2024, February 7). Epilepsy. World Health Organization (WHO). Retrieved July 9, 2026, from https://www.who.int/news-room/fact-sheets/detail/epilepsy

World Health Organization. (2025). Most countries make progress towards universal health coverage, but major challenges remain, WHO–World Bank report finds https://www.who.int/news/item/06-12-2025-most-countries-make-progress-towards-universal-health-coverage-but-major-challenges-remain-who-world-bank-report-finds

World Health Organization. (2021). Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200

World Health Organization. (2021). *Global strategy on digital health 2020–2025*. https://www.who.int/publications/i/item/9789240020924

World Health Organization. (2024). Epilepsy [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/epilepsy

World Health Organization. (n.d.). *Priority medical devices / MeDevIS*. Retrieved July 10, 2026, from https://www.who.int/activities/prioritizing-medical-devices

World Health Organization. (2026, April 25). Artificial intelligence and evidence-informed policy: Emerging challenges and opportunities: Discussion paper. https://www.who.int/publications/i/item/B09667

Author Contributions

All authors contributed to the writing of this document.

ANNEX

Individual opinión by Laura Ospina Ceron

2.5.1 Health Beliefs and Traditional Practices. Although the framework acknowledges the importance of stakeholder engagement and local validation, it approaches implementation primarily from a biomedical and institutional perspective. This reflects the implicit assumption that healthcare systems operate mainly through formal health institutions and that scientific evidence alone insufficient to promote the adoption of new technologies. However, in many LMICs, health-related decisions are shaped by a much broader social and cultural ecosystem in which traditional medicine, local knowledge systems, religious beliefs, family networks, and community norms play a central role. Consequently, one of the greatest barriers to implementing AI-based medical technologies is cultural legitimacy. A diagnostic tool may demonstrate excellent clinical performance and still fail if it cannot be integrated into local understandings of illness, healing, and trust.

Although the framework proposes a robust public health approach, it continues to rely primarily on a biomedical and institutional vision of development. This perspective assumes that technological innovation, regulation, and scientific evidence are sufficient to transform healthcare systems. However, research in medical anthropology, critical development studies, and global health has consistently shown that technologies are never implemented in a social, political, or cultural vacuum. Health interventions are embedded within power relations, colonial histories, diverse knowledge systems, and different conceptions of health and well-being.

2.5.2 Decolonial Health Context. The framework is grounded in evidence-based medicine and WHO standards, both of which are essential for ensuring quality of care and patient safety. However, these standards are primarily rooted in the Western biomedical paradigm, where disease is understood as a biological disorder that should be diagnosed through scientific procedures and treated with medical interventions. This perspective is not universal. In many societies, health is understood as the result of the interaction between physical, spiritual, environmental, social, and cultural dimensions. When AI technologies are implemented without recognizing these belief systems, they may be perceived as disconnected from local realities and, consequently, face resistance or limited acceptance.

1. **Traditional Medicine as Part of the Healthcare System**

The framework identifies governments, regulators, healthcare professionals, and community health workers as the key stakeholders involved in the co-design process. However, in many contexts these actors do not represent the healthcare system in its entirety.

Traditional healers, traditional birth attendants, religious leaders, herbalists, community authorities, and elders are often the first point of contact for healthcare. Many patients move between traditional and Western medicine, combining both approaches according to their needs, experiences, and beliefs. Ignoring these therapeutic pathways results in incomplete needs assessments and may lead to an overestimation of the actual reach and adoption of AI-based medical technologies.

2. **Historical Critiques of Western Medicine**

Resistance to Western medicine should not automatically be interpreted as misinformation or irrationality. In many countries, there is a historical legacy of colonialism, unethical medical experimentation, unequal access to healthcare, and asymmetric power relations.

Today, these concerns also extend to what has been described as data colonialism: the fear that the clinical data of local populations will be used to improve algorithms developed by foreign companies without the resulting benefits being returned to the communities that generated those data.

Therefore, trust constitutes not only a technological challenge but also a political and social one.

Local Knowledge

Another limitation of the framework is that it conceptualizes local adaptation primarily as clinical validation using locally generated data. However, community knowledge also constitutes a valuable form of evidence.

Communities possess knowledge about disease patterns, barriers to healthcare access, seasonality, stigma, caregiving practices, and cultural factors that are rarely captured in the datasets used to train AI algorithms.

Incorporating this knowledge can improve both the design and the implementation of AI-based solutions.

AI Does Not Operate in a Cultural Vacuum

Algorithms may classify diseases with high accuracy, but people interpret diagnoses through their own cultural frameworks. An AI system capable of detecting epilepsy, for example, may still fail if the condition is socially understood as spiritual possession, divine punishment, or a source of stigma.

Therefore, the success of AI depends not only on technical accuracy but also on its cultural compatibility.

Community Trust as a Public Health Outcome

In addition to measuring diagnostic accuracy, financial sustainability, and equity, the framework should incorporate indicators related to social trust, such as:

· The level of community trust in AI.

· Cultural acceptance of AI-based diagnostic technologies.

· Perceived legitimacy of AI-generated recommendations.

· Participation of community leaders and traditional practitioners.

· Integration of local knowledge throughout the co-design and implementation processes.

Beyond the Conventional Development Paradigm

Critical development scholars challenge the assumption that there is a single universal pathway to progress. Alternative paradigms, such as *Buen Vivir* (Good Living), propose understanding well-being through the relationships between people, communities, and nature, rather than measuring success solely in terms of economic growth or technological innovation.

From this perspective, classifications such as "high-income countries" and "low- and middle-income countries (LMICs)" oversimplify highly diverse realities and reproduce historical hierarchies of power. These categories can encourage the assumption that solutions developed in the Global North should be transferred to the Global South, while overlooking local capacities, knowledge systems, and forms of social organization.

The Political Construction of Development and AI

Artificial intelligence is often presented as a technical solution to health challenges. However, many health problems are primarily rooted in political and structural factors, including poverty, inequality, armed conflict, territorial exclusion, underinvestment in public services, and weak institutions. There is a risk of transforming political problems into technical ones, shifting attention away from their structural causes toward the adoption of new technologies.

This logic echoes longstanding critiques of conventional development and the so-called *tyranny of experts*: the tendency to privilege solutions designed by international experts while minimizing democratic participation and the knowledge generated by local communities themselves.

Critique of Top-Down International Aid

Historical experience shows that many international development programmes have failed because they were designed by external institutions without sufficiently understanding local priorities. This top-down approach reproduces relationships of dependency and reinforces power asymmetries between those who produce knowledge and those who are expected to receive interventions.

Responsible AI implementation requires horizontal processes of knowledge co-production in which communities participate not merely as beneficiaries or data providers but as actors with genuine decision-making power.

Medical Anthropology and Medical Pluralism

Medical anthropology demonstrates that medicine is not culturally neutral. Western biomedicine represents only one among many possible systems for understanding illness and well-being.

In many LMICs, traditional medicine, Indigenous knowledge systems, religious practices, and Western medicine coexist. This medical pluralism means that patients frequently move between different forms of healthcare according to their beliefs, available resources, and lived experiences. Focusing exclusively on the biomedical system results in incomplete assessments of the population's actual healthcare needs.

Local Knowledge and Health Ontologies

Local knowledge should not be regarded as a barrier to innovation but as a legitimate form of evidence. Communities possess valuable knowledge about disease patterns, environmental conditions, cultural barriers, stigma, and caregiving practices that is rarely captured in the clinical datasets used to train AI algorithms.

Likewise, different societies hold distinct ontologies of health and well-being. While biomedicine typically defines disease as an individual biological condition, many communities understand illness as the result of spiritual, social, familial, or environmental relationships. Ignoring these differences limits both the legitimacy and the long-term sustainability of AI-based health technologies.

Towards a Pluriverse of Health Innovation

Rather than assuming that there is a single correct model of innovation, the concept of the *pluriverse* recognizes the coexistence of multiple legitimate ways of producing knowledge and organizing healthcare systems. From this perspective, responsible AI implementation should engage with diverse epistemologies, integrate local knowledge systems, respect traditional practices whenever they are compatible with patient safety, and promote genuinely intercultural co-design processes.

Rather than adapting communities to technology, the challenge is to adapt technology to the diversity of worlds, knowledge systems, and ways of understanding health.

The framework provides a robust analysis of regulatory, epidemiological, and governance dimensions but could be strengthened by incorporating a more critical perspective on development, the coloniality of knowledge, and epistemic diversity. A truly responsible approach to AI in healthcare should not be limited to validating algorithms in local settings; it should also question who defines health problems, whose knowledge is considered legitimate, and how power and the benefits of technological innovation are distributed.

Critical Reflection

Overall, the framework effectively addresses structural barriers such as governance, regulation, infrastructure, financing, and capacity building. However, it pays relatively limited attention to the cultural determinants of health.

A truly context-sensitive framework should recognize that the successful implementation of AI depends not only on technical performance and regulatory compliance but also on whether technologies are perceived as culturally legitimate, socially acceptable, and respectful of local knowledge systems. Rather than viewing traditional practices as obstacles, they should be understood as valuable sources of knowledge and as potential partners in the co-design and implementation of AI-based healthcare innovations.

WORKING GROUP 2

AI Supply Chains: Uncovering and Addressing the Hidden Costs of AI Development

Facilitated by Morgan Williams and Caetano Patta Barros (ILO)

Authors:

Adriana Caballero
Divya Dey
Mario Fernandez Hernandez
Killian Foloppe
Bryson Frank
Vasiliki Elisavet Georgiadou
Maria Stephania Guzman Serna
Pamela Eyre Victoria Lira
Vera Minder
Aria Mousavifar
Julia Ratcliffe
Ethan Spencer
Madhurya Tarafder
Mohamed Tawhid
Juliana Zanubi

1 AI Supply Chains: Uncovering and addressing the hidden costs of AI development

1.1 Introduction

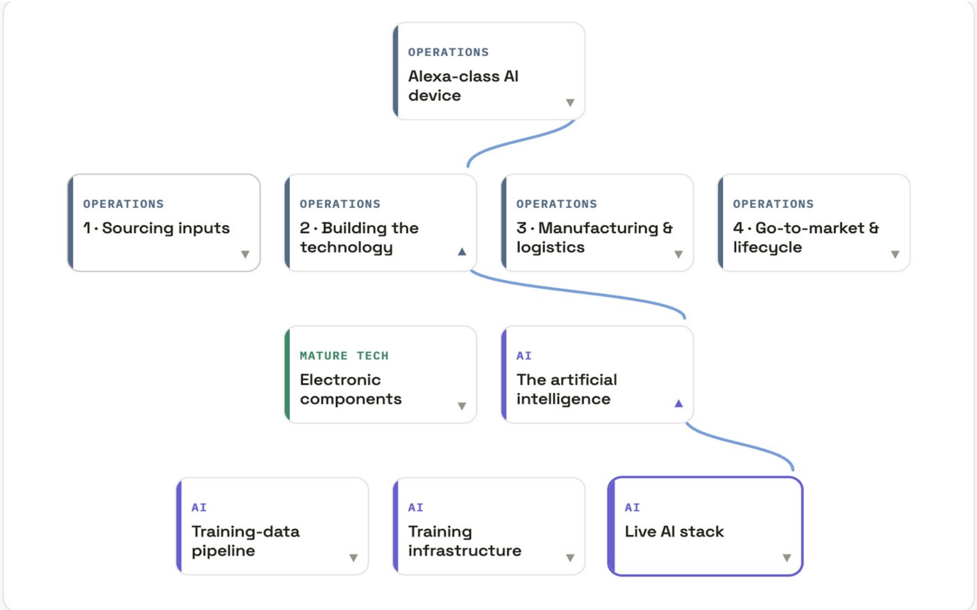
Across the globe, Artificial Intelligence (AI) has emerged as a technology transforming societies, economies and daily lives, with many opportunities and challenges still being explored in these domains. In this context, AI solutions are often perceived as immaterial technologies, depending on little to no physical resources, an account which resonates with wider narratives of immateriality associated with digitalization (Coyle 1997; Ensmenger 2018; Negroponte 1996). However, despite the prevalence of such accounts, the existence and development of AI technologies rest on extensive but often invisibilized global supply chains that are deeply material, mobilizing a wide range of physical and human resources, and raising numerous social, labour and environmental risks. Given the exponential increase in AI use and the growing economic, social and geopolitical power of AI companies, it is critical to map and analyze the risks in these supply chains, so as to ensure that this technology develops within planetary boundaries while advancing SDGs. As such the following document shall analyze these social, labour and environmental dimensions across the supply chains of three popular AI products.

1.2 Case study 1: Uncovering the supply chains of voice assistants

1.2.1 Mapping the supply chain

The development of AI voice assistants relies on an extensive global supply chain that mobilizes significant material resources and human labour, with social and environmental costs at each stage. This section traces this supply chain across its four main phases (Figure 1).

Figure 1. Mapping the supply chain of voice assistants[1]



Source: Authors' own elaboration

Sourcing inputs: At the initial upstream phase of this supply chain, inputs such as raw materials are sourced to create the future components of the hardware and digital infrastructure supporting AI voice assistants. This involves mining and refining key minerals and metals such as rare earth metals, silicon, cobalt, copper tungsten and gold (Integrated Microfabrication Lab n.d.; Manthey 2025; SEC 2017).

Building the technology: The second phase is focused on building the technologies that sustain voice assistants, which can be divided into two further subcategories: the development of electronic components (hardware), and the development of the AI voice assistant in itself (software). In the first case, hardware components such as microphones, speakers, circuit boards, and AI-accelerator chips are produced, to be placed in a shell made from moulded plastic and aluminium (Manthey, 2025). The second case involves developing, training and deploying the actual AI voice assistant, which begins with the collection of voice data from users, hired speakers and other audio sources (Day et al. 2019; Jee 2019). These recordings are then transcribed, cleaned and labelled by human annotation teams so that spoken language can be converted into structured training data. The resulting datasets are then used to train and fine-tune models for speech recognition, language understanding, dialogue management and speech generation (Day et al. 2019; Jee 2019; Panay 2025). Once evaluated and integrated into products, the assistant is deployed to users, whose interactions may generate further data for monitoring, improvement and retraining.

Manufacturing and logistics: In the third phase, the device is manufactured. This includes the assembly of parts, quality validation, as well as the boxing and shipping of the products. AI voice assistant devices are contract-manufactured through intensive labour, primarily in East Asian countries like China (Day 2018). Quality control and packaging follow standard procedures involving a combination of automatisation and manual checks. The coordination of products shipping to their destination is handled by the internal logistics division of the company, with AI forecasting demands and deciding the quantity in production (Emilio 2026).

1 Access the interactive map here: https://claude.ai/public/artifacts/b88b96a4-cc30-4dc3-ae1f-9cd32d1f8fa5

Go-to-market and lifecycle: In the fourth and final stage, the device reaches its destination, however, its lifecycle does not end there. Once the product is sold through the company's store, third-party retail, online or in person, the customer connects the product to their company account and home network. This starts a cycle of inference and finetuning: the software learns the customers specific preferences, and adjusts its behaviour and response pattern to optimise itself as a home assistant over time. At the same time, the customer trains the underlying software through their interactions, which are transcribed, annotated and fed as training data back into phase two (Tubaro and Casilli 2022). Both human agents and AI chatbots serve as customer support. At the end of the device's life, its parts are returned, recycled and/or disposed of.

1.2.2 Identifying hidden costs across the supply chain

Sourcing inputs: While the supply chains behind AI voice assistants are rarely fully mapped, their social and environmental costs are even harder to trace. One of the more well-documented stages lies upstream, in the mining of minerals used in digital devices and infrastructures. This includes tin, tantalum, tungsten and gold, which are subject to conflict-mineral due diligence because their extraction has been linked to armed groups and conflict financing, especially in and around the Democratic Republic of Congo (SEC 2017). Mining also carries labour risks, including accidents, toxic exposure, long shifts, low pay, weak protections and, in some contexts, child or forced labor (Integrated Microfabrication Lab n.d.; Manthey 2025; Savinova et al. 2023; SEC 2017). Its environmental impacts include water depletion and contamination, land degradation and habitat loss, toxic tailings and processing waste, and emissions from extraction, processing and transport. Other minerals used in batteries, speakers and electronic components, including cobalt, lithium and rare earth elements, raise further concerns around concentrated supply and environmental pressure. Together, these risks show that AI systems depend on extractive processes long before data collection or model training begins (Ang Li Ern et al. 2024; Azadi et al. 2023).

Building the technology: This second phase shifts the risks from extraction to manufacturing, data work and computation. On the hardware side, labour risks include demanding employment conditions in electronics manufacturing, such as repetitive factory work, long hours, low pay and subcontracting, as well as occupational health and safety risks linked to exposure to chemicals, fine particles and industrial processes (Day 2018). Environmental risks can be grouped into hardware resource demand, including electricity and water use in semiconductor and electronics production; pollution and waste, including chemical residues and production scrap; and climate impacts linked to energy-intensive fabrication, assembly and transport (Chamberlain 2019; Day 2018; Smith 2025).

On the software side, voice-data collection raises privacy, consent and surveillance concerns, while transcription and annotation can depend on low-paid, precarious and largely invisible data labor (Casilli et al. 2022; Day 2019; Eadicicco 2016; Hern 2019; Jee 2019; Tonella 2025). Training, deployment and continuous use add further environmental pressure through data-centre electricity demand, cooling requirements, server turnover and related emissions (Morley et al. 2018).

Manufacturing and logistics: The manual tasks in the assembly phase are mostly outsourced. Although contract manufacturers tend to require a Supplier Code of Conduct, allegations of precarious work conditions such as illegal overtime or underage work have been raised, reporting the involvement of teenagers in breach of local labour laws (Chamberlain 2019).

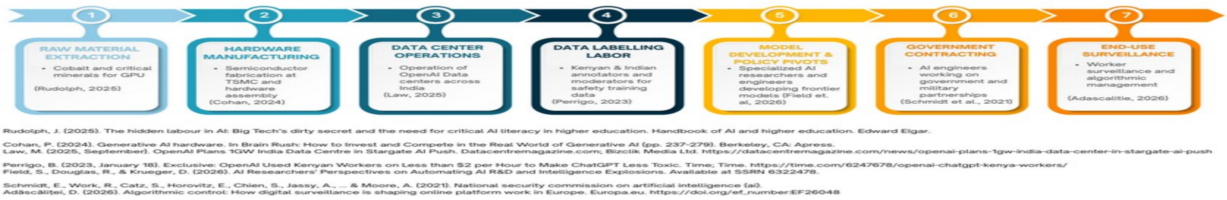
Go-to-market and lifecycle: Risks associated with the last phase of the supply chain reach from the customers’ living rooms to the annotators' workplaces. The mostly female AI assistant voices have been found to promote gender stereotypes, painting women as obedient and eager to help (UNESCO and EQUALS Skills Coalition 2019). Since the devices collect audio data to feed back into the training phase, as well as user data that can be used to adjust recommendations, advertisement and personal services, this raises serious privacy concerns (Day et al. 2019; Hern 2019; Jee, 2019). The technology shows the same increasing environmental footprint as all AI use, causing detrimental costs in regards to water, land, and climate ("AI's environmental costs threaten water, land and climate," 2026). The workers involved in annotating the audio are mostly engaged in unstable repetitive microtask-work, validating the automatic transcription (Tubaro and Casilli 2022). This labour includes the confrontation with possibly emotionally distressing audio, such as witnessing sexual assault (Day et al. 2019).

1.3 Case study 2: Uncovering the supply chain of military applications of LLMs

1.3.1 Mapping the supply chain

OpenAI was established in 2015 as a nonprofit for profit hybrid dedicated to ensuring that artificial general intelligence benefits humanity. Since January 2024, however, it has expanded into defence and national security markets by removing explicit prohibitions on military applications from its usage policies (CNBC 2024), culminating in a US$200 million Department of Defense contract in June 2025 and a February 2026 agreement with the United States Department of War (CNBC 2026). Its flagship model, ChatGPT, launched in November 2022, provides a useful case study for examining the hidden costs of AI supply chains because its development depends on a global network spanning critical mineral extraction, semiconductor manufacturing, cloud infrastructure, outsourced data annotation, and government deployment. While this supply chain enables advanced AI capabilities, it also externalizes significant labour, environmental, and governance costs that remain largely invisible to end users (Figure 2).

Figure 2. ChatGPT Supply Chain Process[2]



Source: Authors' own elaboration

Raw material extraction: ChatGPT's hardware depends on cobalt, lithium, copper, and rare earth elements. Cobalt extraction is concentrated in the Democratic Republic of Congo, which supplies over 70% of global output, through a mix of artisanal, small-scale, and industrial mining operations (Sovacool 2021). Rare earth elements are sourced primarily from China, while lithium is extracted mainly in Chile and Australia. These raw materials are processed and sold to semiconductor manufacturers before entering the chip fabrication stage.

Hardware manufacturing: Refined minerals are transformed into semiconductors, a stage dominated by TSMC, which produces around 90 percent of the world's most advanced chips. OpenAI's computing

2 Access interactive map here: https://claude.ai/public/artifacts/054df656-77b3-41ec-b812-23f735289563

infrastructure relies primarily on NVIDIA GPUs, alongside AMD and Broadcom hardware. The company is expanding its Stargate infrastructure while developing a custom inference chip, codenamed Titan, fabricated on TSMC's 3 nm process and expected to enter production in late 2026.

Data center operations: Assembled hardware is deployed in data centers, primarily on Microsoft Azure infrastructure, alongside OpenAI's own Stargate campuses under construction in Texas, New Mexico, and Wisconsin. This stage completes the physical infrastructure chain linking mineral inputs and chip fabrication to the operational compute capacity that trains and serves ChatGPT.

Data labeling labor: Once infrastructure is operational, models require human labelled data to improve performance and safety. OpenAI contracted Sama to employ Kenyan workers between 2021 and 2022 to classify harmful text and images for ChatGPT's moderation systems (Time 2023). This business process outsourcing model now coexists with platform based systems such as Scale AI's Outlier, which recruits independent contractors worldwide for data annotation, reinforcement learning from human feedback, and model evaluation (Forbes 2025). Unlike traditional outsourcing, platform workers bear unpaid training time, irregular task availability, and limited labour protections, leading to multiple lawsuits alleging wage theft and poor working conditions (AlgorithmWatch 2025).

Model development & policy pivots: In January 2024, OpenAI removed explicit prohibitions on military and warfare use from its usage policies. In June 2024, retired General Paul Nakasone, former NSA Director and U.S. Cyber Command chief, joined OpenAI's board and its Safety and Security Committee. In December 2024, OpenAI announced a partnership with defense contractor Anduril to deploy AI systems for national security applications.

Government contracting: OpenAI signed a US$200 million contract with the Department of Defense in June 2025 covering both warfighting and enterprise applications under its OpenAI for Government initiative. In February 2026, following the collapse of Anthropic's negotiations with the Pentagon, OpenAI concluded a second agreement to deploy its models on the Department of War's classified network. After criticism, CEO Sam Altman acknowledged that the initial contract language had "looked opportunistic and sloppy," prompting subsequent revisions.

End-use surveillance: The February 2026 agreement prohibits mass domestic surveillance, autonomous weapons targeting, and certain high stakes automated decisions while requiring compliance with FISA, Executive Order 12333, and the Fourth Amendment. Deployment occurs through OpenAI Public Sector LLC using a cloud API architecture rather than direct battlefield integration. Nevertheless, the agreement authorizes use on classified government networks to support data analysis, situational awareness, and military decision making (Forbes, 2026). OpenAI argues that cloud based deployment, rather than edge deployment, prevents its models from operating as fully autonomous weapons (OpenAI 2026).

1.3.2 Identifying hidden costs across the supply chain

Labour Risks: Labour risks emerge at two distinct points in the AI supply chain: upstream mineral extraction and midstream data work. At the upstream level, production of advanced semiconductors depends on critical minerals, particularly cobalt from the Democratic Republic of Congo, where artisanal mining is associated with hazardous conditions, child labour, and weak occupational safety protections (Sovacool 2021). These workers provide essential inputs for AI infrastructure while capturing only a minimal share of downstream value. At the midstream level, human labour remains indispensable for model development. A landmark 2023 Time investigation documented how Kenyan workers contracted through Sama, an OpenAI subcontractor, were paid between US$1.32 and US$2 per hour to label graphic content depicting violence, sexual abuse, and hate speech (Perrigo 2023). Workers consistently reported psychological distress, insufficient mental health support, and abrupt contract terminations without pay. Structural conditions deepen the vulnerability: youth unemployment in Kenya exceeds 67%, and extensive subcontracting and confidentiality agreements further insulate AI developers from direct accountability while reducing transparency throughout the chain (Rani and Williams 2026). In response, affected workers established the Data Labelers Association in early 2025, now representing over 800 members advocating for fair pay, mental health protections, and a platform code of conduct.

A further hidden cost concerns military end-use opacity. A February 2026 investigation by Rest of World, in partnership with the Bureau of Investigative Journalism and the Pulitzer Center's AI Accountability Network, found that gig workers employed by data-annotation firm Appen were largely unaware that their work fed into U.S. military applications. Joan Kinyua, president of the Data Labelers Association, identified this as a structural problem: companies rarely disclose to workers who they are ultimately working for or what the purpose of a given task is. This opacity constitutes a hidden cost distinct from low wages or psychological harm, it denies workers informed consent, any possibility of differentiated compensation for high-stakes end-uses, and any mechanism to object to, or even be aware of, the military application of their labour.

Environmental Costs: ChatGPT's environmental footprint extends well beyond running the models themselves. Mineral extraction drives deforestation, biodiversity loss, soil degradation, and water contamination in producing regions, impacts structurally disconnected from the companies that benefit from these resources. Cobalt for AI chips is mined predominantly in the DRC, where an estimated 40,000 children work in artisanal mines (UNICEF 2012), with no binding traceability requirement linking OpenAI's chip supply chain back to those sites. Greenpeace East Asia documented a 4.5 fold increase in AI chip manufacturing emissions in a single year (Greenpeace 2025), costs outsourced to suppliers in Taiwan and South Korea still heavily dependent on fossil power. Downstream, e-waste from rapid hardware cycling follows a well documented geography, processed predominantly in Ghana, Nigeria, and Pakistan (Njoku et al. 2023), where workers dismantle toxic components, including heavy metals and PFAS compounds, with no regulatory protection and no connection to the corporations whose products generated the waste.

By 2030, computing is projected to account for up to 8 percent of global energy consumption, driven by expanding AI adoption in civilian and military applications. A professor in social and ethical AI at Sweden's Umeå University noted that some data farms training and operating AI systems in Europe and Canada consume energy equivalent to a small city (Teeratanabodee 2022). Large data centers compound these costs globally: training GPT-3 alone consumed roughly 700,000 liters of freshwater for on-site cooling, rising to 5.4 million liters including electricity-related water use (Li et al. 2025), and inference now drives 80 to 90 percent of total AI energy and water consumption through billions of daily interactions. By 2030, AI-powering data centers are projected to consume 945 terawatt-hours of electricity (UNU-INWEH 2026), nearly triple the combined annual electricity use of Pakistan, Bangladesh, and Nigeria, while their water footprint will equal the basic annual domestic water needs of all 1.3 billion people in Sub-Saharan Africa.

Military & Surveillance risks: OpenAI's 2023 usage policy originally banned military use outright. That prohibition was quietly removed in January 2024, and by February 2026 the company had signed a classified agreement with the U.S. Department of Defense, rushed through within hours of Anthropic's refusal to accept the same terms, to deploy AI systems across classified military networks.

The agreement's central clause, permitting use for "any lawful purpose," has been widely criticized as structurally inadequate: it ties surveillance and weapons restrictions to existing U.S. legal frameworks while effectively shifting interpretation to the Pentagon, leaving enforceability dependent on authorities the government controls unilaterally. Critically, the contract's stated protections apply only to U.S. persons, offering no legal protection to the non-U.S. populations most likely to face AI-assisted surveillance, targeting, and lethal decision-making, including communities in conflict zones, border regions, and refugee contexts across the Global South. A Bureau of Investigative Journalism investigation published in early 2026 found that gig workers in Africa had unknowingly contributed labor to U.S. military AI supply chains, illustrating how the human costs of military AI are dispersed globally while governance and accountability stay concentrated among a few states and corporations.

Beyond this accountability gap, the technical risks of deploying general-purpose AI in military contexts remain substantial and poorly understood. AI-supported targeting systems identify targets through pattern recognition and probabilistic prediction rather than contextual human judgment, making them vulnerable to hallucinations, adversarial attacks, and cybermanipulation, any of which can misidentify civilian infrastructure or fabricate intelligence, raising the risk of unlawful attacks. These risks are compounded by growing reliance on synthetic training data, where one model trains on another's outputs, a process documented to cause model collapse and to amplify errors and biases recursively.

A large-scale King's College London study found that leading AI models, including ChatGPT, opted for nuclear escalation in 95% of simulated conflict scenarios, raising concerns about deploying these systems in high-stakes military decisions without binding international oversight. While legal frameworks governing biometric data and privacy exist in some jurisdictions, these protections do not extend to armed conflict, leaving a governance vacuum precisely where AI military deployment is expanding fastest. Accountability is fragmented across developers, cloud providers, subcontractors, and government agencies, making responsibility for adverse outcomes structurally impossible to establish, while the communities bearing the consequences are, by design, those least able to seek redress.

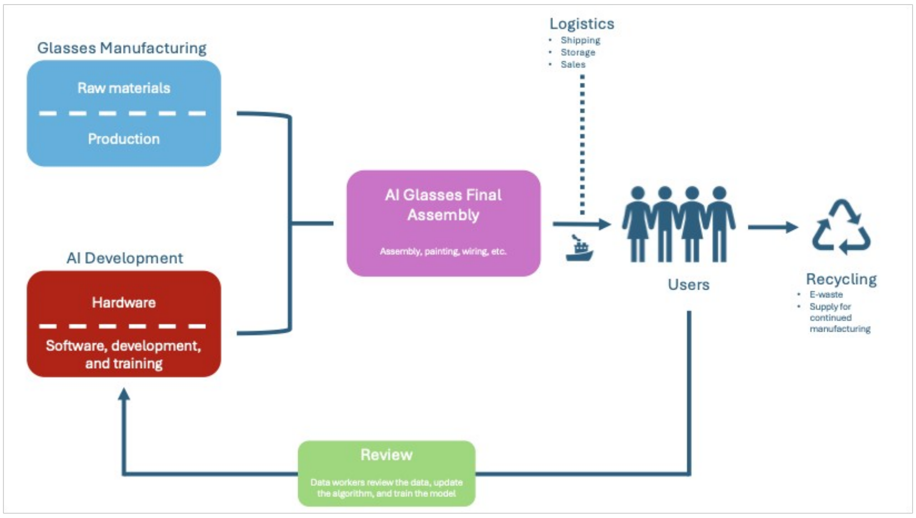
1.4 Case study 3: Uncovering the supply chains of smart glasses

1.4.1 Mapping the supply chain

The AI glasses market is experiencing explosive growth. In 2025, global AI glasses shipments reached approximately 8.7 million units, representing a dramatThe AI glasses market is experiencing explosive growth. In 2025, global AI glasses shipments reached approximately 8.7 million units, representing a dramatic 322% increase (China Daily 2026; VarIndia 2026). By 2026, the market is expected to surpass 10 million units (Omdia 2025), with some projections reaching 22.67 million units under the most optimistic scenarios (Omdia 2025). Camera-equipped smart glasses combine traditional eyewear aesthetics with AI-powered capabilities including voice commands, image capture, visual recognition, live translation, and facial recognition. They are powered by advanced processors, image sensors, lithium-ion batteries, and microphones, and rely on cloud-based AI infrastructure for many of their core functions.

The AI glasses supply chain is global and multi-layered, spanning raw material extraction, component fabrication, assembly, distribution, and cloud AI infrastructure (Figure 3). Critical minerals such as cobalt, lithium, rare earth elements, and copper are essential for AI infrastructure. China controls over 60% of global refining capacity for critical raw materials and dominates supply chains for cobalt and lithium (Visual Capitalist 2025). Semiconductor fabrication is heavily concentrated in a few key Asian economies. Final assembly occurs primarily in China and Vietnam, while AI training and data processing rely on data centres in the United States, Europe, and Asia (World Trade Organization 2025).

Figure 3: Supply chain for AI glasses[3]



Source: Author’s own elaboration.

1.4.2 Identifying hidden costs across the supply chain

This section details the labour and environmental risks embedded across AI glasses supply chains, from raw material extraction through end-of-life disposal.

Labour Impacts

Glasses supplies/hardware: Roughly 90% of the world's optical frames are made in China, concentrating governance risk (OECD Roundtable on Corporate Responsibility 2002). Investigations have found breaches of Chinese labour law on wages, hours, and social security (OECD Roundtable on Corporate Responsibility 2002). Lens grinding, polishing, and coating expose workers to toxic solvents, linked to elevated cancer risk (Tsai et al. 2010; Chang et al. 2013; Lee et al. 2020).

AI Models (Infrastructure, Hardware, Usage): The workforce behind AI is paid and protected very unevenly. At the bottom, Kenyan labellers filtered violent and sexual content for ChatGPT at roughly 1.32 to 2 dollars an hour (TIME 2023), chip workers behind Samsung's 2018 apology suffered 360-plus illnesses and 110-plus deaths (Inquirer Business 2018), about 28 percent of Malaysian electronics workers were found in forced-labour conditions (Verité 2014), and Bangladeshi migrants building Malaysia's data centres face debt bondage and passport confiscation (Bloomberg 2026). Near the top, US construction trades gained 25 to 30 percent pay rises (CNBC 2026) and expert annotators earn up to about 200 dollars an hour (TIME 2026).

---

3 Access the interactive map here: https://claude.ai/public/artifacts/f8ded57b-f114-4a8e-8d91-ed61b30bed27

For a more detailed description of the AI development component of this supply chain, see Chapter 2 of the OECD’s *Competition in artificial intelligence infrastructure.*

Manufacturing: Women comprise 60–90% of the electronics workforce, yet occupy fewer higher-paying roles, facing wage inequality and harassment (OECD 2025; ILO 2025). Migrant and temporary workers face forced labour, recruitment debt, and passport retention, often unable to organise due to employer-tied visas (OECD 2025; ILO 2024). Chemical exposure is linked to miscarriage and reproductive disorders (Kim et al. 2014; Chan, Ngai, and Selden 2013).

Logistics: Algorithmic warehouse management drives work intensification, injury, and psychological stress (Hennum Nilsson et al. 2025). Port automation weakens dock workers' bargaining power absent transition policies (International Transport Workers' Federation 2025). Last-mile delivery relies on subcontracted, migrant, and platform workers facing insecure, low-wage employment (Chan 2023).

Data review: Facebook agreed in 2020 to a 52 million dollar settlement with more than 10,000 US content moderators who developed PTSD from reviewing graphic material (NPR 2020), and Kenyan courts have since ruled that Meta can be sued locally over moderator conditions despite its outsourcing structure, ordering it to provide proper psychiatric care in June 2023 (Computer Weekly; LSE Africa 2025).

E-waste: Manual dismantling in informal recycling exposes workers to lead, cadmium, mercury, and flame retardants from batteries and components, risking respiratory and neurological harm (WHO 2021).

Environmental Impacts

Glasses supplies/hardware: Eyewear manufacturing is wasteful: only 20% of an acetate plate is used, with 80% discarded (Association of Optometrists 2019; EOS 2025). Most frames rely on petroleum-based, non-biodegradable plastics (EyeO Eyewear). Adding electronics compounds this: smart glasses generate up to nine times more CO_2 than traditional glasses (SILMO Next Expert Committee), and acetate production relies on petrochemicals and fossil-fuel plasticisers which contributes to a worsening climate (VisionPlus Magazine).

AI Hardware, Development, and Coding : The computation behind AI carries a large and rising resource footprint. Global data centres used about 448 terawatt-hours of electricity in 2024, more than all but 10 countries, emitting roughly 208 million tons of CO2 and consuming about 1.2 trillion gallons of water, with demand projected to nearly double by 2030 (PBS / UN University / IEA 2026). A single large data centre can use up to 5 million gallons of water a day, and each chip fabrication plant about 10 million (Lincoln Institute 2026). Concentration strains local grids, pushing electricity prices up 267 percent in high-density areas (Consumer Reports 2026), while roughly 56 percent of US data-centre power comes from fossil fuels and the sector stays among the least transparent industrial water users (EESI 2026; AGU Advances 2026).

Manufacturing: AI glasses draw on semiconductor supply chains the OECD identifies as high-risk for emissions, water use, and pollution (OECD 2025). Semiconductor manufacturing generates roughly 50 megatons of CO_2 annually, and nearly 75% of a battery-powered device's carbon footprint originates in manufacturing, almost half from chip fabrication (imec, n.d.), outweighing device-operation emissions (imec n.d.). If the EU Chips Act reaches its 20%-of-global-semiconductors target by 2030, manufacturing emissions could rise up to eightfold, with chip production potentially reaching 3% of global emissions by 2040 (imec n.d.).

Logistics: AI hardware moves primarily via container ship, sometimes air freight for high-value products; both carbon-intensive. Maritime shipping alone accounts for roughly 2–3% of global CO_2 emissions (International Maritime Organization 2021; International Council on Clean Transportation 2024).

E-waste: Integrating cameras, processors, batteries, and sensors into a frame makes AI glasses far harder to repair or recycle than standard eyewear. Global e-waste reached 62 million tonnes in 2022, projected to hit 82 million tonnes by 2030, with recycling lagging (ITU and UNITAR 2024). Glued-in batteries and non-replaceable components make repair impractical (Roskladka et al. 2025), pushing consumers toward full replacement (NextReality 2025; Gutterman 2025). Each discarded pair also drives fresh lithium extraction, adding habitat loss and mining emissions (UNEP 2024).

1.5 Addressing hidden costs across AI supply chains

Environmental and labour abuses in AI supply chains are driven by three core factors:

1) **supply chain opacity**, where global production networks (mining, semiconductors, assembly, logistics) obscure responsibility for environmental damage and labour conditions;

2) **cost and speed pressures**, which incentivise outsourcing to low-cost manufacturers and logistics providers, encouraging overtime, weak safety standards, and environmental externalisation; and

3) **fragmented global governance**, where inconsistent labour and environmental regulation across jurisdictions can mean weak enforcement.

The following policy proposals tackle the largest risks in the supply chain of AI development:

1) **Governments and public procurement bodies** should condition all AI technology procurement on mandatory, independently audited human rights and environmental impact assessments, covering labor conditions across the full subcontracting chain, full-lifecycle environmental disclosure (mirroring EU CSRD standards), and biometric/data protection safeguards extended explicitly to armed conflict and national security contexts, closing current legal carve-outs.

2) **Governments and public procurement bodies**, with support from the ITU and ISO, should require repairable product design, extended producer responsibility, and end-of-life collection and recycling for AI-enabled consumer devices to reduce e-waste and demand for resource extraction.

3) **The International Labour Organization (ILO)**, in coordination with UNCTAD, should establish a binding international framework for AI supply chain labor standards, applying the ILO Living Wage Methodology to data annotation work, formally recognizing worker-led bodies such as the Data Labelers Association as standard-setting stakeholders, and requiring joint employer liability so that lead AI developers share legal responsibility for the mental health and safety of subcontracted labor, regardless of geographic distance from the end task.

4) **AI developers and their governmental clients** should be legally required to maintain a single, designated accountable entity per deployment, enforce mandatory human-in-the-loop validation for any targeting, escalation, or high-stakes intelligence decision, and publicly disclose, at minimum in redacted form, subcontracting arrangements, model versions deployed, and synthetic-vs-human training data ratios.

5) **International and regional regulatory bodies (e.g., EU, UN bodies, International Humanitarian Law forums)** should update existing legal frameworks to explicitly govern algorithmic profiling, biometric data collection, and mass data aggregation. Specifically during armed conflict, these frameworks should also mandate pre-deployment adversarial and escalation testing for any model used in conflict simulation or decision-support roles, given documented risks such as the 95% nuclear escalation rate found by King's College London.

6) **Affected communities and civil society** should be guaranteed an independent, accessible redress mechanism, separate from the contracting government agency, for individuals or communities harmed by AI-enabled surveillance, misidentification, or wrongful targeting, ensuring that accountability gaps do not fall disproportionately on those with the least power to challenge the technology.

7) **The ILO should engage three multi-stakeholder policy coalitions and create a new deliberative body to better include the above labour issues in the conversation**, where such issues currently see little sustained attention. The **Global Digital Compac**t and the **Partnership for AI (PAI)** should engage with the **Global Deal**'s report to include workers in the social dialogue to ensure adequate representation of stakeholders to establish AI-specific standards for labour and good governance.

    a. **PAI**, a coalition of major AI firms, brings the AI industry, **UN specialized agencies** (such as the ILO and UNDP) and **civil society organizations** together for a public mandate and credential. **The Global Deal** brings nation states, private sector actors, and labour unions together through social dialogue to establish joint solutions that deliver positive results for workers, businesses, and society at large.

    b. **Policy direction should be decided by those stakeholders within the coalition framework through agreement**, rather than through executive decree from the ILO. The ILO should act as a broker, rather than an executive branch. This is to preserve organisational autonomy and sectorial and national policy expertise in those venues.

    c. Voluntary entry into the coalition binds the joining organisation to the labour-rights philosophies established within ILO precedent. As such, **any deliberation regarding the divergence of labour policy amongst stakeholders should be on policy implementation only**, rather than on any normative commitments, as agreement on those is a precondition of entry.

References

Acikgoz, Fulya, Rodrigo Perez-Vega, Fevzi Okumus, and Nikolaos Stylos. 2023. "Consumer Engagement with AI-Powered Voice Assistants: A Behavioral Reasoning Perspective." Psychology & Marketing 40 (11). https://doi.org/10.1002/mar.21873.

Aczel, M., Chamanara, S., Matin, M., Farsi, A., Marwala, T., Madani, K. (2026). Environmental Cost of AI's Energy Use: Carbon, Water and Land Footprints. United Nations University Institute for Water, Environment and Health (UNU-INWEH), Richmond Hill, Ontario, Canada. doi: 10.53328/INR26RMA002

Adăscăliței, Dragoș. "Algorithmic Control: How Digital Surveillance Is Shaping Online Platform Work in Europe." Eurofound, 2026. https://doi.org/ef_number:EF26048

AI Chip Map. "OpenAI." 2026. https://www.aichipmap.com/en/company/openai/

AlgorithmWatch. "The AI Revolution Comes With the Exploitation of Gig Workers." August 9, 2025. https://algorithmwatch.org/en/ai-revolution-exploitation-gig-workers/

Building Africa's AI Advantage: Data Centres, Talent, and the Next Digital Economy - ALN. (2026, June 30). ALN | Africa Legal Network. https://aln.africa/news/building-africas-ai-advantage-data-centres-talent-and-the-next-digital-economy/

American Immigration Council. "ICE to Use New AI System to Track Immigrants." 2025. https://www.americanimmigrationcouncil.org

Bradley-Cobb, Hannah. 2024. "Benefits of AI-Powered Voice Assistant for Aging Adults - Randolph County Center | N.C. Cooperative Extension." Randolph County Center. March 21, 2024. https://randolph.ces.ncsu.edu/news/ai-voice-assistant-aging-adults/.

Bureau of Investigative Journalism / Rest of World. "Gig Workers in Africa Had No Idea They Were Helping the U.S. Military." Rest of World, February 23, 2026. https://restofworld.org/2026/gig-workers-us-military-appen/

CNBC. "OpenAI Quietly Removes Ban on Military Use of Its AI Tools." January 16, 2024. https://www.cnbc.com/2024/01/16/openai-quietly-removes-ban-on-military-use-of-its-ai-tools.html

CNBC. "OpenAI Adds Former NSA Chief to Its Board: Paul Nakasone, Sam Altman." June 13, 2024. https://www.cnbc.com/2024/06/13/openai-adds-former-nsa-chief-to-its-board-paul-nakasone-sam-altman.html

CNBC. "OpenAI Wins $200 Million US Defense Contract." June 16, 2025. https://www.cnbc.com/2025/06/16/openai-wins-200-million-us-defense-contract.html

CNBC. "OpenAI Strikes Deal with Pentagon Hours After Rival Anthropic Was Blacklisted by Trump." February 27, 2026. https://www.cnbc.com/2026/02/27/openai-strikes-deal-with-pentagon-hours-after-rival-anthropic-was-blacklisted-by-trump.html

Cohan, Peter. "Generative AI Hardware." In Brain Rush: How to Invest and Compete in the Real World of Generative AI, 237–279. Berkeley, CA: Apress, 2024.

CORDIS, cordis.europa.eu. 2023. "Inclusive Voice AI for People with Speech Impairments." CORDIS | European Commission. September 15, 2023. https://cordis.europa.eu/article/id/446340-inclusive-voice-ai-for-people-with-speech-impairments.

Coyle, Diane. 1997. The Weightless World: Strategies for Managing the Digital Economy. MIT Press.

Cymutta, Sebastian. Biometric Data Processing by the German Armed Forces during Deployment. Tallinn: CCDCOE Publications, 2022.

Data Labelers. "About." Accessed 2026. https://datalabelers.org

Day, Matt (The Seattle Times). 2018. "Watchdog Criticizes Conditions at Chinese Factory That Builds Amazon Echo Speakers." June 9. https://www.seattletimes.com/business/amazon/watchdog-criticizes-conditions-at-chinese-factory-that-builds-amazon-echo-speakers/.

Day, Matt, Giles Turner, and Natalia Drozdiak / Bloomberg. 2019. "Thousands of Amazon Workers Listen to Alexa Users' Conversations." Time, April 11. https://time.com/5568815/amazon-workers-listen-to-alexa/.

The Economist. 2024. "AI Can Bring Back a Person's Own Voice." The Economist. December 11, 2024. https://www.economist.com/science-and-technology/2024/12/11/ai-can-bring-back-a-persons-own-voice?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignid=18151738051&ppcadid=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclsrc=aw.ds&gad_source=1&gad_campaignid=18151761343&gbraid=0aaaaadbuq3j1kkwvcnh8ywzxo5sftwegq&gclid=cjwkcajw6rfsbhaqeiwa_yocpinhvt0syvxba-qvgljlhkpuarux5loz6dssdpine0344jg_tsfl4hocwtgqavd_bwe.

Emilio, Núria. 2026. "Predictive Logistics: How Amazon Uses AI to Anticipate Your Order." Bismart. https://blog.bismart.com/en/predictive-logistics-amazon-ai.

Ensmenger, N. 2018. "The Environmental History of Computing." Technology and Culture 59 (4): S7–33.

European Parliament and Council of the European Union. Regulation (EU) 2016/679 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). Official Journal of the European Union L 119 (2016): 1–88.

Field, Sarah, Ryan Douglas, and David Krueger. "AI Researchers' Perspectives on Automating AI R&D and Intelligence Explosions." SSRN, 2026. https://ssrn.com/abstract=6322478

Francisco, Bryan Paul S. "Artificial Intelligence Is a 'Black Box.' Maybe Not For Long." TIME, May 21, 2024. https://time.com/6980210/anthropic-interpretability-ai-safety-research/

Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM). Responsible by Design: Strategic Guidance Report on the Risks, Opportunities, and Governance of Artificial Intelligence in the Military Domain. The Hague: HCSS, September 2025. https://hcss.nl/gc-reaim-publications/

Greenpeace East Asia, "Chipping Point: Tracking Electricity Consumption and Emissions from AI Chip Manufacturing," April 10, 2025

Human Rights Commission of Australia. "Big Tech Ignoring the Human Cost Behind the Rise of ChatGPT." Accessed 2026. https://humanrights.gov.au/about-us/media-centre/opinion-pieces/opinion-pieces/big-tech-ignoring-human-cost-behind-rise-chatgpt

IEA. (2025, April 10). Energy and AI – Analysis - IEA. IEA. https://www.iea.org/reports/energy-and-ai

Integrated Microfabrication Lab. n.d. "Silicon Water Fabrication Process." Brigham Young University. Accessed May 7, 2026. https://cleanroom.byu.edu/ew_formation.

Intuition Labs. "OpenAI Stargate Datacenter Details." 2025. https://intuitionlabs.ai/articles/openai-stargate-datacenter-details

Jee, Charlotte. 2019. Amazon Workers Are Listening to Some of Your Conversations with Alexa. January 11. https://www.technologyreview.com/2019/04/11/136118/amazon-workers-are-listening-to-some-of-your-conversations-with-alexa/.

Kenya Federation of Employers. "Youth Employment." Accessed 2026. https://www.fke-kenya.org/policy-issues/youth-employment

King, Anthony. "Digital Targeting: Artificial Intelligence, Data, and Military Intelligence." Journal of Global Security Studies 9, no. 2 (2024): 1–16. https://doi.org/10.1093/jogss/ogae009

King's College London. "Artificial Intelligence Under Nuclear Pressure: First Large-Scale King's Study Reveals How AI Models Reason and Escalate Under Crisis." February 27, 2026. https://www.kcl.ac.uk/news/artificial-intelligence-under-nuclear-pressure-first-large-scale-kings-study-reveals-how-ai-models-reason-and-escalate-under-crisis

Kooli, Chokri, and Rim Chakraoui. 2025. Review of AI-Driven Assistive Technologies in Inclusive Education: Benefits, Challenges, and Policy Recommendations. Sustainable Futures 10 (July): 101042. https://doi.org/10.1016/j.sftr.2025.101042.

Kumar, Sachin. "Problems with Synthetic Data." Aitude, March 1, 2023. https://www.aitude.com/problems-with-synthetic-data/

Law, Matt. "OpenAI Plans 1GW India Data Centre in Stargate AI Push." Data Centre Magazine, September 2025. https://datacentremagazine.com/news/openai-plans-1gw-india-data-center-in-stargate-ai-push

Li, Haoran, and Xin Wei. 2025. Review of Factors Influencing Older Adults' Adoption of AI Voice Assistants: Extending the UTAUT Model. Frontiers in Psychology 16 (November). https://doi.org/10.3389/fpsyg.2025.1618689.

Li, Jiaan, Yulu Hu, Shihan Fu, Zhiqing Wu, Pan Hui, Hailiang Wang, and Mingming Fan. 2024. Review of "It's like Having a Friend": Exploring the Opportunity of Large Language Models (LLMs) in Assisting Retirees In Their Daily Lives. Proceedings of the Twelfth International Symposium of Chinese CHI, November, 463–75. https://doi.org/10.1145/3758871.3758911.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models [Review of Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models]. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2304.03271

Liu, Mingzhou, Caixia Wang, and Jing Hu. 2023. "Older Adults' Intention to Use Voice Assistants: Usability and Emotional Needs." Heliyon 9 (11): e21932. https://doi.org/10.1016/j.heliyon.2023.e21932.

Manthey, Ewa. 2025. "How Rare Earths Became the Next Trade War Weapon." ING Economics, October 30. https://think.ing.com/articles/how-rare-earths-became-the-next-trade-war-weapon/.

Musser, Megan. Adversarial Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications. Washington, DC: Center for Security and Emerging Technology, 2023. https://cset.georgetown.edu/publication/adversarial-machine-learning-and-cybersecurity/

Negroponte, N. 1996. Being Digital. Knopf Doubleday Publishing Group.

Nieva, Richard. "A Growing Side Hustle For American College Grads: Fixing AI's Wrong Answers." Forbes, March 6, 2025. https://www.forbes.com/sites/richardnieva/2025/03/06/scale-ai-outlier-us/

Njoku, A., Agbalenyo, M., Laude, J., Ajibola, T. F., Attah, M. A., & Sarko, S. B. (2023). Environmental injustice and electronic waste in Ghana: Challenges and recommendations. International Journal of Environmental Research and Public Health, 21(1), 25.

OECD. OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas. Paris: OECD Publishing, 2016. https://doi.org/10.1787/9789264252479-en

OpenAI. "Five New Stargate Sites." 2025. https://openai.com/index/five-new-stargate-sites/

OpenAI. "Introducing OpenAI for Government and Our Structure." 2025. https://openai.com/global-affairs/introducing-openai-for-government/

OpenAI. "Our Agreement with the Department of War." February 2026. https://openai.com/index/our-agreement-with-the-department-of-war/

Panay, Panos. 2025. "Introducing Alexa+, the next Generation of Alexa." Amazon News, February 26. https://www.aboutamazon.com/news/devices/new-alexa-generative-artificial-intelligence.

Perrigo, Billy. "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic." TIME, January 18, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/

Privacy International. "Kenyan Content Moderators Sue Over Working Conditions." Accessed 2026. https://privacyinternational.org/examples/5127/kenyan-content-moderators-sue-over-working-conditions

Rani, Uma, and Morgan Williams. "Challenging the Myth of AI Autonomy: The Convenient Fiction of Autonomous Intelligence." Weizenbaum Journal of the Digital Society 6, no. 1 (2026). https://doi.org/10.34669/wi.wjds/6.1.6

Roush, Tyler. "OpenAI, Nvidia, Alphabet And More Sign AI Deal With Pentagon For Classified Military Use." Forbes, May 1, 2026. https://www.forbes.com/sites/tylerroush/2026/05/01/openai-nvidia-alphabet-and-more-sign-ai-deal-with-pentagon-for-classified-military-use/

Rudolph, Jürgen. "The Hidden Labour in AI: Big Tech's Dirty Secret and the Need for Critical AI Literacy in Higher Education." In Handbook of AI and Higher Education. Cheltenham: Edward Elgar, 2025.

Schmidt, Eric, Robert Work, Safra Catz, Eric Horovitz, Steve Chien, Andy Jassy, and Andrew Moore. National Security Commission on Artificial Intelligence. Washington, DC: NSCAI, 2021.

SEC. 2017. "Conflict Minerals Disclosure: A Small Entity Compliance Guide." U.S. Securities and Exchange Commission (SEC), July 14. https://www.sec.gov/resources-small-businesses/small-business-compliance-guides/conflict-minerals-disclosure.

Shah, S. (2026, June 3). AI Could Use as Much Water as 1.3 Billion People by 2030, U.N. Report Warns. TIME; Time. https://time.com/article/2026/06/03/ai-global-water-resources-un-report/

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. "AI Models Collapse When Trained on Recursively Generated Data." Nature 631, no. 8022 (2024): 755–759. https://doi.org/10.1038/s41586-024-07566-y

Sovacool, Benjamin K. "When Subterranean Slavery Supports Sustainability Transitions? Power, Patriarchy, and Child Labor in Artisanal Congolese Cobalt Mining." The Extractive Industries and Society 8, no. 1 (2021): 271–286. https://www.sciencedirect.com/science/article/pii/S2214790X20303154

Teeratanabodee, W. (2017). IP22039 | The Environmental Impact of Military AI. @RSIS_NTU. https://rsis.edu.sg/rsis-publication/idss/ip22039-the-environmental-impact-of-military-ai/

Tubaro, Paola, and Antonio A. Casilli. 2022. "Human Listeners and Virtual Assistants: Privacy and Labor Arbitrage in the Production of Smart Technologies." In Digital Work in the Planetary Market, edited by Mark Graham and Fabian Ferrari. The MIT Press. https://doi.org/10.7551/mitpress/13835.003.0014.

In DR Congo, UNICEF supports efforts to help child labourers return to school. (2012, June 13). Unicef. https://www.unicef.org/childsurvival/drcongo_62627.html

United Nations University Institute for Water, Environment and Health (UNU-INWEH). Environmental Cost of AI's Energy Use: Carbon, Water and Land Footprints. Richmond Hill: UNU-INWEH, 2026. https://unu.edu/inweh/news/environmental-cost-of-AIs-Enrgy-use-carbon-water-and-land-footprints

Valle Escolano, Raquel. 2023. "Artificial Intelligence and Rights of People with Disabilities: The Power of Algorithms." Revista Española de Discapacidad 11 (1): 29–50. https://doi.org/10.5569/2340-5104.11.01.02.

Vieira, Alessandro Diogo, Higor Leite, and Ana Vitória Lachowski Volochtchuk. 2022. Review of The Impact of Voice Assistant Home Devices on People with Disabilities: A Longitudinal Study. Technological Forecasting and Social Change 184 (November): 121961. https://doi.org/10.1016/j.techfore.2022.121961.

Watch.dig. "Human Workers in AI Training Raise Privacy Concerns." Accessed 2026. https://dig.watch/updates/human-workers-ai-training-raise-privacy-concerns

WORKING GROUP 3

Implications of Artificial Intelligence and Emerging Technologies for Youth Wellbeing

Facilitated by Dunola Mgbachi (ITU)

Authors:

Alissar Azzam
Anusha Simkhada
Anyu Jiang
Asztrik Benjamin Herbak
Diana Nalyvanna
Dorcas Asare
Ezekiel Olaniyan
Giulia Bonsanti
Laura Rodriguez
Marilyn Conde
Pranjali Thakur
Uzair Khan
Wendpanga Boua

Summary

This paper examines how artificial intelligence and emerging technologies affect the wellbeing of young people aged 18–35, with particular attention to how disparities in digital access shape both the opportunities and the risks these technologies present. It reviews AI's potential to support mental health, education, and civic participation; assesses the socio-psychological, privacy, and equity risks associated with its use; and proposes a set of recommendations for governments, international organisations, and the private sector.

Introduction

Artificial Intelligence is reshaping how young people learn, communicate, seek help, and understand the world around them; often before regulators, educators or even young people themselves have had time to weigh its consequences. This brief paper examines the implications of AI and emerging technologies for the wellbeing of youth, understood here as those aged 18 to 35, and defined broadly to encompass mental, social, and psychological health rather than the absence of illness alone.

The stakes of this inquiry are not evenly distributed. Access to the internet, reliable connectivity, and digital literacy vary sharply across regions, and these disparities shape not only whether young people can benefit from AI-enabled tools, but also how exposed they are to the risks such tools introduce. A young person navigating an AI chatbot on a stable broadband connection in a well-regulated market experiences a fundamentally different set of opportunities and vulnerabilities than one accessing the same technology through a shared device and an intermittent mobile network. Any assessment of AI's impact on youth wellbeing must therefore account for this unevenness rather than treat "youth" as a single, homogeneous population.

This raises the central question guiding this paper: How does AI technologies affect youth wellbeing in regions with different levels of digital access? In addressing this question, the paper proceeds in three parts. Section I considers the ways AI can promote youth wellbeing, from mental health support and personalized education to safety-enhancing wearables. Section II turns to the risks in terms of privacy erosion, dependence, misinformation, and digital inequality that accompany these same technologies. Building on this dual assessment, section III proposes a set of recommendations organized around three reinforcing levels. The first one being setting a common standard, the second one being building the capacity to meet it and lastly, ensuring youth have a genuine role in shaping both.

1. Opportunities: How AI Can Promote Youth Wellbeing

Artificial intelligence has the potential to improve youth wellbeing in several areas, particularly by expanding access to healthcare, strengthening education and employability, and creating new opportunities for participation and inclusion. While AI presents important challenges, it can also become a valuable tool for supporting young people when implemented responsibly.

1.1 Mental Health Support

One of the most promising applications of AI is in mental health support. AI-powered conversational agents and chatbots can provide 24/7 assistance, helping young people access information, emotional support and self-management tools at any time. A recent systematic review found that AI-based conversational agents can reduce psychological distress and improve wellbeing, particularly when they are designed to complement rather than replace professional healthcare services. These tools may be especially valuable for young people who face barriers to accessing traditional mental health services, such as cost, stigma or geographical distance (Li et al., 2023).

1.2 Education and Employability

AI also has considerable potential to improve education, skills development and youth employability. AI can support personalised learning, identify individual skill gaps and provide tailored recommendations for education and career development. According to the World Bank, AI can improve youth employment by enabling better job matching, generating personalised skills assessments and providing real-time labour market information. These applications can help young people make more informed educational and career decisions while preparing them for rapidly changing labour markets (Singh S. et al., 2020).

1.3 Participation and Social Inclusion

Beyond education and employment, AI can also strengthen participation and social inclusion. Recent research argues that AI systems can empower young people by improving access to information, supporting civic engagement and facilitating participation in digital communities when they are developed in an ethical and inclusive manner. At the same time, researchers emphasise that these opportunities depend on transparent governance, responsible implementation and equal access to digital technologies (Fitas R., 2025) and (Smith et al., 2023).

Overall, AI should not be viewed solely as a technological innovation but as a tool that can enhance youth wellbeing across multiple dimensions. To maximise these benefits, governments, international organisations and technology developers should promote human-centred AI that improves access to services, strengthens digital skills and ensures that young people can actively participate in the digital transformation.

2. Risks: How AI May Undermine Youth Wellbeing

2.1 Socio-Psychological Concerns

The use of generative AI, chatbots and companions are found to be undermining youth's socio and psychological well-being. It encourages emotional dependence through confirmation-biased interactions and manipulative engagement strategies, such as guilt-inducing farewells, which have been linked to crisis-level attachment and worsening mental health outcomes (Negreiro & Anaç, 2026). Excessive reliance on AI for communication and emotional support may also reduce face-to-face interaction, weaken interpersonal skills, and increase loneliness, particularly among youth with limited social networks (Klimova & Pikhart, 2025). While AI can complement mental health services, it lacks the empathy and contextual understanding required for complex emotional needs and should not replace professional diagnosis and treatment (Miner et al., 2017; Vaidyam et al., 2019). Further, AI anxiety and unemployment anxiety among youth is closely tied to fears of job displacement, with a moderate positive correlation found between the two, among university students (Ucar et al., 2024). Beyond mental health, over-reliance on AI-generated responses may increase laziness, weaken critical thinking, decision-making, and analytical reasoning by reducing opportunities for peer interaction, independent thinking, learning and reflection (Zhai et al., 2024; Ahmad et al., 2023; Bai et al., 2023).

2.2 Privacy, Data, and Harmful Content

AI systems also present important risks related to privacy, online safety, and harmful content. Young people frequently share sensitive personal information with AI applications, creating risks of data misuse, privacy breaches, and inadequate protection of confidential information (American Psychological Association, 2019). At the same time, AI-generated misinformation and deepfakes exploit people's tendency to trust realistic images, audio, and text, making false information increasingly convincing. Repeated exposure to manipulated content and algorithmically amplified misinformation can weaken critical thinking, reinforce echo chambers, and erode trust in reliable sources, increasing psychological distress and confusion among adolescents (Kang & Valadez, 2025).

2.3 Inequalities, Exclusion, and Gender

The benefits and risks of AI are unevenly distributed, with existing social inequalities shaping who can safely and effectively use these technologies. Limited AI literacy, inadequate digital skills, and unequal access to quality digital infrastructure increase young people's vulnerability to misinformation, privacy violations, online harms, and manipulative AI systems (Yasmine et al., 2025; Negreiro & Anaç, 2026). These disparities extend beyond access to meaningful participation and opportunities in education, employment, and social life. Gender further compounds these inequalities. Studies have found widespread gender bias in AI systems, with large language models often reproducing sexist stereotypes (UNDP, 2025; UN News, 2026). Girls and young women are also disproportionately affected by AI-enabled online harassment, including deepfakes and hate speech, while women remain more exposed to AI-driven labour market disruption due to occupational segregation (ILO, 2026; UNESCO, 2025).

3. Recommendations

The following recommendations are addressed to stakeholders involved working in the intersection of technology and youth well being, such as: international organizations, national governments, non-governmental organizations, policy makers, researchers, civil society and young adults.

3.1 Strengthen AI literacy and digital competencies

- Develop AI literacy curricula and toolkits and distribute them across educational programs. Integrate age-appropriate AI literacy into formal, informal, and non-formal education, and regularly update educational materials and content to reflect the rapidly evolving AI landscape.
- Expand AI literacy programs to parents and caregivers, recognizing that adult digital behaviors and awareness influence youth attitudes and technology use, and that a whole-of-community, multistakeholder approach is critical.
- Emphasize within digital competency education the principles of ethical AI use, critical thinking, privacy, transparency, data protection, and algorithmic bias and discrimination, with a focus on inclusivity and reaching vulnerable communities through partnerships with relevant stakeholders.
- Establish standards and guidelines, in collaboration with teachers, pedagogues, youth, and AI experts, for the use of AI in learning in ways that support critical thinking, decision-making, and analytical reasoning, and promote the implementation of these standards and guidelines.

- Increase ongoing support for youth in addressing career and employability anxiety by equipping them with knowledge, resources, and skills through free courses, training, and innovative solutions that help them adapt to changing labor markets and technological innovation.
- Develop an AI, mental health, and well-being course, as well as a global alliance of psychologists, in partnership with relevant UN institutions, and disseminate these resources to local communities and educational institutions.

3.2 Invest in research and meaningful youth participation

- Invest in interdisciplinary research on AI's impact on youth mental health and well-being, as well as in community-led projects and youth initiatives.
- Ensure research includes and is led by diverse populations, including women, persons with disabilities, marginalized communities, refugees, Indigenous Peoples, and conflict-affected youth, among others, to better understand their needs and how AI is affecting them.
- Build on and expand the work and reach of existing UN-led youth groups and innovation hubs, enabling young people to actively participate in AI governance and policymaking.
- Develop Youth AI and Well-being Programs in regional offices and local communities to collect evidence and generate data and guidance for regulators, policymakers, and other stakeholders on the safe and responsible use of AI.

3.3 Strengthen governance, transparency, and platform accountability

- Support the development of national regulatory frameworks that strengthen protections for youth and adolescents against AI-related risks while promoting responsible, inclusive, and human rights-based technology use.
- Strengthen context-specific and collaborative AI governance through capacity building, international partnerships, and inclusive policymaking that upholds human rights.
- Encourage AI developers to provide transparent and accessible explanations of algorithms, recommendation systems, and data collection practices.
- Promote regular assessments of algorithmic bias and discriminatory outcomes, along with mitigation measures and direct engagement mechanisms through which young people can raise concerns and provide feedback.
- Prevent the unauthorized use of the private data of youth and adolescents and strengthen reporting mechanisms for harmful AI-generated content, including deepfakes and AI-facilitated abuse.
- Introduce direct mental health support in high-risk situations by connecting affected youth to local lifelines or 24/7 mental health support services through platform-based interventions (e.g., pop-up alerts).
- Promote the integration of crisis intervention mechanisms into digital platforms so that young users who are exposed to, or engage with, content associated with self-harm or suicide are immediately directed to trusted local mental health services, crisis helplines, and emergency support resources.

3.4 Protect privacy, safety, and equitable access

- Limit the collection and secondary use of adolescents' personal data and prohibit exploitative targeted advertising and the sale of personal data.
- Promote clear, age-appropriate informed consent and strengthen protections for sensitive biometric and behavioral data.
- Encourage AI systems that provide health-related information to clearly communicate their limitations and encourage users to seek guidance from trusted adults or qualified professionals.
- Address structural digital inequalities through multilingual and accessible AI tools, technical assistance, knowledge sharing, and investments in digital skills, recognizing that meaningful participation depends not only on infrastructure but also on education, social capital, and cultural context.

References

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10, Article 311. https://doi.org/10.1057/s41599-023-01787-8

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X*, 1(3), e30. https://doi.org/10.1002/brx2.30

Fitas, R. (2025). Inclusive education with AI: supporting special needs and tackling language barriers. *AI and Ethics*, 5, 5729–5757. https://doi.org/10.1007/s43681-025-00824-3

International Labour Organization. (2026). Gen AI, occupational segregation and gender equality in the world of work. https://www.ilo.org/publications/gen-ai-occupational-segregation-and-gender-equality-world-work

J., Aryee, L. M. D., Bang, H., Prajogo, S., Choi, Y. K., Hoch, J. S., & Prado, E. L. (2023). Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: Systematic review and meta-analysis. *JMIR Mental Health*, 10, e43066. https://doi.org/10.2196/43066

Klimova, B., & Pikhart, M. (2025). Exploring the effects of artificial intelligence on student and academic wellbeing in higher education: A mini-review. *Frontiers in Psychology*, 16, Article 1498132. https://doi.org/10.3389/fpsyg.2025.1498132

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and wellbeing. *npj Digital Medicine*, 6, 236. https://doi.org/10.1038/s41746-023-00979-5

Negreiro, M., & Anaç, Ö. D. (2026, May). The spread of AI companions and the challenges they generate (PE 789.299). European Parliamentary Research Service, Members' Research Service.

UN Women. (2026, June 21). AI is getting women wrong as gender bias persists, data reveals. *UN News*. https://news.un.org/en/story/2026/06/1167776

UNESCO. (2025, September 22). Tackling gender bias and harms in Artificial Intelligence (AI). https://www.unesco.org/ethics-ai/en/articles/tackling-gender-bias-and-harms-artificial-intelligence-ai

Uçar, M., Çapuk, H., & Yiğit, M. F. (2025). The relationship between artificial intelligence anxiety and unemployment anxiety among university students. *Work*, 80(2), 701–710. https://doi.org/10.1177/10519815241290648

United Nations Development Programme. (2025, October 5). AI, gender bias and development. https://www.undp.org/eurasia/blog/ai-gender-bias-and-development

World Bank. (2020). Three ways artificial intelligence can help boost youth employment. *World Bank Blogs*. https://blogs.worldbank.org/en/jobs/three-ways-artificial-intelligence-can-help-boost-youth-employment

Yasmine, D. I., Colombijn, F., van Deursen, A. J. A. M., & van Ingen, E. (2025). Youth digital wellbeing: The role of digital skills and positive and negative digital outcomes in youth's subjective wellbeing. *Computers in Human Behavior Reports*, 20, Article 100796. https://doi.org/10.1016/j.chbr.2025.100796

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-7

ITU One-Page Summary

Companion brief to the Working Group 3 report, prepared for the International Telecommunication Union (ITU)

Artificial Intelligence is reshaping the lives of young people, influencing how they learn, connect, work and seek support. As AI adoption accelerates, ensuring equitable access while addressing challenges related to wellbeing, trust and inclusion has become a global priority. Unequal digital access, connectivity and AI literacy mean these benefits and risks are not experienced equally.

Why This Matters

Globally, an estimated 79% of people aged 15–24 use the internet, making them the most connected age group.

75%+	2.2 B	13 M+
young adults view Generative AI as useful	people remain offline globally, limiting equitable access	use AI to support learning and homework

Opportunities vs. Emerging Challenges

Opportunities

- **Mental Health Support:** AI-enabled tools can expand access to timely emotional support.
- **Education & Employability:** Personalized learning and AI-enabled skills development.
- **Participation & Inclusion:** Improved access to information and digital collaboration.
- **Innovation for Social Good:** AI can accelerate youth-led innovation in health and education.

Emerging Challenges

- **Psychosocial Concerns:** Emotional dependence, reduced social skills, career anxiety.
- **Privacy & Harmful Content:** Data misuse, misinformation, and deepfakes.
- **Digital Inequalities:** Youth without digital skills face greater exposure to harms.
- **Gender & Online Safety:** 58% of young women and girls report online harassment.

Strategic Priorities for ITU

- **Standardize:** Global standards for ethical AI development, data privacy, and youth protection.
- **Strengthen:** Capacity-building for digital literacy and AI-related skills, and stronger regulation.
- **Empower:** Institutionalize youth engagement in AI governance and policymaking.

Our Vision

A future where Artificial Intelligence serves as a powerful enabler for all young people as partners, fostering their growth, protecting their wellbeing, and empowering them to thrive in the digital age.

Working Group 3 — Video Presentation

The following video accompanies the Working Group 3 report on AI, emerging technologies, and youth wellbeing.

Play video – click HERE

WORKING GROUP 4

AI & International Security: Technological Applications, Risks and Governance

Facilitated by Ioana Puscas (UNIDIR)

Authors:

Valentina Barbosa
Maximilian Huter
Timur Bekzhanov
Abdulhabir Karahanli
Shauna Cassin
Jomana Abunahla
Sandrine Chartray
Soufiane Talbi
Ananyaa Yadav
William Cunningham
Charlotte Yeung
Laila Ansari

1. Introductory Remarks

1.1 Analyzing AI Potential in Different Stages of the Conflict Cycle

Artificial intelligence (AI) has entered a period of rapid innovation, with research, industry, and governments partnering for its development and regulation. Its impact is no longer limited to civilian sectors as AI is increasingly relevant to international security, ranging from cybersecurity and critical infrastructure resilience to intelligence gathering, battlefield decision-support, autonomous systems, and post-conflict recovery.

In conflict and security studies, conflict is often analyzed using the conflict cycle framework, developing through different stages commonly centered around latent tensions to escalation, stalemate, de-escalation, and post-conflict peacebuilding (Wohlfeld, 2010).

AI is currently used across all stages of the conflict cycle. Across these stages, the applications, actors, and regulatory challenges associated with AI differ. Implications range from low-risk support functions like translation to active military uses like Lethal Autonomous Weapon Systems (LAWS). Furthermore, actors also vary between both state and non-state actors, as well as the growing role of private technology companies, data providers, and humanitarian organizations.

Furthermore, this emerging technology also comes with questions of accountability, human oversight, data reliability, and the gap between technological development and regulatory challenges.

This report delves into the technological innovations, risks and governance associated with the growing role of AI in international security. It moves through the five stages of conflict to provide an account of its current uses, implications and potential ways forward.

2. Latent Conflict Intention

2.1 AI Applications

2.1.1 Predictive Conflict Forecasting

ACLED (Armed Conflict Location & Event Data) built one of the world's largest datasets on political violence, pulling from local media, human rights reports, and monitors on the ground to log protests, battles, and attacks on civilians as they happen (ACLED, 2021). Its AI-powered CAST (Conflict Alert System) forecasts the likelihood of future violent events, achieving accuracy rates exceeding 90%. This enables policymakers and humanitarian organizations to identify areas at heightened risk before violence escalates (Panic, 2025).

2.1.2 Satellite Imagery and Military Activity Detection

In 2022, Ukraine partnered with Planet Labs along with UNOSAT and various investigators to capture pictures from space and utilise those as early warning signals to detect and monitor, in real-time, troops and military gear near Yelnya. These satellite images were also used to track battlefield infrastructure impact (International Panel on the Information Environment, 2025).

2.1.3 Social Media Monitoring and Information Environment Analysis

Alongside imagery, AI is also being used to monitor the information environment. AI and machine learning are increasingly geared towards analysing posts across platforms like X, TikTok, Facebook, and WhatsApp to identify emerging trends and potential risks. Build Up's Phoenix tool, for example, analyses public social media posts to flag areas where peacebuilding engagement may be urgently needed (Bernard, 2023).

2.2 Risks and Security Challenges

2.2.1 AI-powered conflict mapping (satellites, drones, crowd data)

A major concern is the possibility of misinterpretation caused by inaccurate AI assessments in limited context. Predictive models depend on historical data and statistical correlations, but conflict dynamics are highly political and context-specific. An AI system may incorrectly identify routine military exercises, political protests, social movements and movement near borders as indicators of imminent conflict but they cannot fully explain the social and political reasons behind these patterns. In the latent stage, where tensions are still developing, incorrect interpretation of mapped indicators may exaggerate threats or overlook underlying grievances. This can result in preventive responses based on visible symptoms rather than the actual drivers of conflict (Cederman & Weidmann, 2017).

2.2.2 AI monitoring of information environments (hate speech, polarization, disinformation)

Although AI-based monitoring of information environments can help detect early signs of social tension, it also creates risks. One major concern is the potential transformation of conflict-prevention tools into instruments of surveillance and political control. Systems designed to monitor hate speech, extremist narratives, or disinformation may also be used by governments to track political opposition, journalists, activists, or minority communities. In fragile contexts, this could reduce trust between citizens and authorities and contribute to the grievances that create instability (OHCHR, 2021; Freedom House, 2023). Another challenge is the difficulty of accurately interpreting online communication. AI systems may struggle to understand cultural context, local languages, political expression, humor, or sarcasm. As a result, legitimate criticism or activism could be incorrectly classified as harmful content, while genuinely dangerous narratives may remain undetected. These errors are especially concerning in politically sensitive environments where incorrect classification can influence security responses (UNESCO, 2023). Finally, the same AI capabilities used to detect harmful information can also be weaponized to manipulate the information environment. Actors may use AI to generate large-scale disinformation campaigns, create realistic deepfakes, or deploy automated accounts that amplify polarization. Therefore, AI becomes a dual-use technology: it can help identify emerging conflict risks, but it can also accelerate the spread of narratives that increase tensions before violence begins (UN Secretary-General, 2023; Europol, 2022).

2.3 Governance

2.3.1 AI Monitoring of Information Environments (Hate Speech, Polarization, Disinformation)

Governance of AI-based information monitoring should establish clear separation between conflict early-warning functions and security surveillance activities. Systems developed to identify hate speech, disinformation, or polarization trends should operate with strict purpose limitations, where data collected for conflict prevention cannot be reused for identifying political opponents, tracking individuals, or monitoring lawful civic activity. Independent auditing bodies should regularly assess whether these systems comply with privacy and freedom of expression protections, particularly in fragile political environments (OHCHR, 2021; UNESCO, 2021). In addition, maintaining human-in-the-loop verification before AI-generated alerts influence conflict-prevention or security decisions is essential. Regional experts and conflict analysts should review AI outputs to prevent cultural misunderstandings, language errors, or the misclassification of legitimate political expression as harmful content (UNESCO, 2023).

2.3.2 Governance Principles for AI-Based Conflict Early Warning Systems

AI should inform decisions, not make them. A conflict-mapping system might flag unusual troop movement or a spike in hostile rhetoric, but that flag should never be the trigger for a diplomatic, military, or humanitarian response on its own. It needs to pass through human eyes first like regional experts, intelligence analysts, conflict researchers, and local stakeholders before anyone acts on it. The people closest to a conflict should have a say in whether an AI's findings actually hold up. Local NGOs, journalists, peacebuilders, and humanitarian workers on the ground should be involved in the decision making.

3. Conflict Emergence and Escalation

3.1 Technical Applications

3.1.1 AI-Enabled Intelligence and Surveillance

AI changes the traditional parameters of conflict in the context of emergence and escalation by introducing a shift in how intelligence and surveillance is conducted. Previously, while large amounts of data could be collected about a person, there remained the issue of organizing the data. Agentic AI streamlines this process by quickly integrating data together (Schneier and Penney, 2026). In short, any domain for which data exists can be used by AI systems to detect patterns and trends that humans might otherwise miss. Big data and disparate information streams can therefore be brought together into a single common output. These multimodal models (models which integrate text, images, and audio) can also immediately analyze this data for government officials (Neuberger, 2025). This is used in military decision-making, especially in information-denied environments. It also supports cognitive warfare capabilities by increasing the availability and processing of data and information (UNIDIR, 2025).

3.1.2 Generative AI

In addition to this, Generative AI usage has been growing (Wan Rosli, 2025). This AI creates original content for a wide range of applications including malware detection, automated hacking, code-adapting malware, phishing e-mails, and sock puppet accounts. Previously, it took significant amounts of effort to create sock puppet accounts, or fake accounts used for intelligence purposes (Ritu, 2023). These often took months or years to build as these accounts required regular posting, interactions with other users, and feasible profile pictures and backstories. Now, Generative AI can produce relatively realistic images as well as automate posting, commenting, and liking posts.

3.1.3 State-Sponsored Exploitation of Commercial AI Systems

In recent years, there has been emerging documentation from tech companies of their AI being used for various forms of espionage and surveillance. Typically, individual, state-affiliated actors ask different AI groups to do certain, specific tasks. In 2024, OpenAI released a report where they identified five state-affiliated actors using ChatGPT for cyber operations (OpenAI Security). These actors used OpenAI services for various tasks including basic coding tasks, retrieving publicly available information on intelligence agencies, researching how malware could evade detection, and drafting content for what OpenAI said could be used for phishing campaigns. In 2025, Google's Threat Intelligence Group reported an actor known as APT41 using Gemini for code obfuscation while another user, Muddy

Water, tricked Gemini into developing custom malware under the guise of being a student (Google Threat Intelligence Group). Microsoft documented an operation using generative AI to place operatives' images onto stolen identity documents (Microsoft Threat Intelligence, 2025).

3.1.4 Autonomous AI-Driven Cyberattacks

In 2025, Anthropic released a report that claims that state actors used Claude to launch a cyberattack campaign against thirty government agencies, large tech companies, financial institutions, and chemical manufacturing companies (Anthropic). In this attack, Claude inspected target organization systems and high-value databases before testing security vulnerabilities and then writing exploit code that could harvest credentials to extract large amounts of private data. This was a unique attack in that the cyberattack operation was autonomously conducted with little human direction once the campaign started. It is widely believed to be the first publicly known example of this kind of cyberattack using AI "agents," or where systems can be run autonomously and be capable of complex tasks with minimal to no human intervention (Covino, 2025).

3.2 Risks

3.2.1 Compressed Decision Timelines

Shortened processing and data-collection timelines enables faster decision-making. While AI outputs can reduce information asymmetries through data-driven assessments, including the detection of early warning signs of conflict, such as troop movements, signals, and rhetoric, these analyses depend on the quality of the data available. Noisy environments, as is often the case in conflict settings, can compromise the reliability of outputs. When used to support decision-making in conflict, this creates risks of escalation where an AI system might recommend more aggressive courses of action based on incomplete data inputs, or even hallucinations. These assessments also present automation risks where actors over-rely on AI recommendations without questioning or verifying them (UNIDIR, 2025).

Furthermore, in traditional conflict scenarios, waiting periods can play a role in mitigating escalation by allowing time to gather information, reflect, and decide on a course of action. AI compresses this time dimension by shortening timelines and automating stages of decision-making processes. When the information environment evolves too fast for humans to intervene or de-escalate, this can pose important risks. It can contribute to increased violence, attacks or suspicion by prompting faster responses between parties, which may be perceived as threatening or provocative as a conflict emerges, especially in contexts where speed undermines human agency and control (Johnson, 2025).

3.2.2 Scaling Disinformation

More broadly, the use of Generative AI can transform the information environment by producing disinformation at scale, especially as surveillance and espionage tools are being fine-tuned. This flooding of false and misleading digital content can have a destabilizing effect further feeding conflict emergence and escalation (UNIDIR, 2025).

3.3 Governance

3.3.1 Risk Classification

The shift from traditional, human-centered tools to AI-assisted decision-making tools exposes the intense risks inherent in autonomous AI capabilities. It is therefore important that all AI-based military technology be classified as high-risk and subject to AI auditing, as defined under the EU AI Act of 2024, in order to minimize the black-box effect that currently limits accountability over these systems. This is in addition to foundational cybersecurity laws and programs, such as CISA 2015 and the State and Local Cyber Grant, are maintained, and should fund research into the defensive use of AI, such as DARPA's AIxCyber Challenge (Covino, 2025).

3.3.2 Differential Access

Policymakers should also support and incentivize developers to implement differential access strategies that promote defender access to advanced AI capabilities (Ee et al., 2025). This includes prioritizing access for strategic defenders such as software developers, cybersecurity firms, and critical infrastructure operators, whose security impacts broader societal resilience. This policy should also be extended to stronger privacy and data protection laws that restrict or ban facial recognition and other forms of surveillance and tracking technology for non-strategic defenders.

3.3.3 Human Oversight

Finally, confidence-building measures (CBMs) should be developed to ensure human oversight remains central to escalatory decision-making and deterrence (UNIDIR, 2022). Currently, no universally binding legal rule mandates a minimum frequency of human-in-the-loop involvement before intelligence and surveillance conclusions are finalized. This report therefore recommends international binding standards, such as requiring two independent human confirmations, as a confidence-building measure to keep human judgment central to escalatory decision-making.

4. Full-Blown Conflict (Stalemate)

4.1 AI Applications

Current military applications of AI can be grouped into several broad categories. Decision-support systems (AI-DSS) propose targets and prioritise threats from sensor, signals and open-source data; expert submissions to the Secretary-General emphasise that they exert significant, under-examined influence over the pace of military decisions (Bo et al., 2025). Intelligence, surveillance and reconnaissance (ISR) applications fuse satellite imagery, communications intercepts and unmanned aerial vehicle (UAV) video at a scale beyond human analysts. Autonomous and semi-autonomous platforms, AI-enabled UAVs and loitering munitions have been operationally integrated in Ukraine, while low-cost UAVs have expanded airborne strike capability to non-State actors. Command-and-control integration coordinates multi-domain operations, including the process of regulating, coordinating, and deconflicting radio frequencies (electromagnetic-spectrum management). Non-weapons applications, including predictive maintenance, medical triage, translation of intercepts, and increasingly large language models for course-of-action generation, raise distinct reliability and confidentiality concerns (Grand-Clément, 2023).

4.2 Case Studies

Two contemporary developments illustrate two distinct governance problems raised by the use of AI in armed conflict: one arising from the integration of AI into battlefield systems such as UAVs, ISR, targeting support, and electronic-warfare adaptation; the other arising from AI-assisted target production and compressed human review in densely populated civilian settings.

4.2.1 AI-Enabled UAVs in Contemporary Conflict

The Russia–Ukraine conflict illustrates how AI-enabled UAVs are affecting active conflict, particularly through ISR, targeting support, and resilience against electronic warfare. The International Institute for Strategic Studies (IISS) notes that the war has demonstrated the growing role of UAVs in both scale and function, while the Centre for Strategic and International Studies (CSIS) argues that Ukraine is advancing AI-driven unmanned systems to reduce direct warfighter involvement and enhance combat effectiveness (International Institute for Strategic Studies, 2025; Bondar, 2024). This use case shows that the immediate impact of AI in a conflict scenario is not necessarily fully autonomous warfare, but the integration of AI into drones, ISR, targeting support, and electronic-warfare adaptation in ways that accelerate the battlefield decision cycle and reshape battlefield operations (Watling & Sylvia, 2025).

4.2.2 AI-Supported Targeting and Intelligence Processing

Israeli military operations in Gaza show another form of military AI use: not fully autonomous warfare, but algorithmic support for target production and intelligence processing. Investigative reporting has described Lavender as a system used to identify suspected individuals, while Habsora reportedly identifies buildings and infrastructure for attack (Abraham, 2024). Human Rights Watch (2024) warns that such tools rely on faulty data and inexact approximations, increasing the potential of civilian harm. The Gaza case therefore suggests that AI's immediate impact lies in accelerating and bureaucratising targeting: it can make violence appear more precise while shortening human review, expanding target lists, and normalising higher civilian-casualty thresholds.

4.3 Risks

Although AI offers military advantages, each of the applications shares a common danger. Starting with the decision-making, the more it gets handed to machines, the harder it becomes for a human to object. In command and control, military decision-making speeds up, placing pressure on human operators' capacity to carefully examine AI-generated outputs before acting on them (Grand-Clément, 2023).

The case studies demonstrate how AI has altered military operations, but it also raises significant concerns. In the Russia-Ukraine conflict, AI enabled drones and surveillance systems help militaries detect targets and respond more quickly, however these technologies reduce the time available for human judgement and can be affected by electronic warfare, cyberattacks, or inaccurate data, increasing the risk of mistakes. In contrast, during the Gaza war, concerns have been raised by AI-assisted targeting systems like Lavender and Habsora. Both are using algorithms to identify targets but may wrongly classify civilians or civilian objects, particularly when the data is inaccurate or insufficient. These examples show that while AI can enhance military efficiency, it cannot replace human judgement or guarantee compliance with international humanitarian law (Abraham, 2024; Human Rights Watch, 2024; ICRC, 2019; IISS, 2025).

AI is not completely objective, and it learns from the data given. Therefore, if the data is incomplete, biased, or inaccurate, the AI's decisions can also be unreliable, from decisions on who or what to attack and when to decisions about military strategy and the use of nuclear weapons. Furthermore, AI can sometimes produce results that are difficult to predict, making it risky to rely on these systems in armed conflict. Therefore, human judgment must remain central to decisions (International Committee of the Red Cross [ICRC], 2021).

4.4 Governance

Governance should focus on implementing existing law rather than creating new legal categories. Since Resolution 79/239 confirms that IHL and IHRL apply across the full life cycle of military AI, states should codify meaningful human control through minimum review-time standards, explainable outputs, and clear legal responsibility (United Nations, 2024). Article 36 reviews should apply to AI systems, including

AI decision-support tools used in targeting, with pre-deployment testing, verification, validation, and independent auditability (Human Rights Watch, 2024). Commercial providers should also face binding due-diligence obligations under the UN Guiding Principles on Business and Human Rights. Multilateral processes, including Resolution 80/58, REAIM, and the Global Commission on Responsible AI in the Military Domain, can support confidence-building measures on doctrine, transparency, and deployment notification.

5. De-escalation, Negotiation and Mediation

5.1 Technical Applications

5.1.1 Negotiation Modelling

AI is increasingly used in the preparation and simulation of conflict negotiations. The UN University Centre for Policy Research (UNU-CPR) developed the "Ask Abdalla" model which was trained to act like a member of the Sudanese Rapid Support Forces (RSF) and the "Ask Amina" model which simulated the perspective of a woman in a refugee camp in Chad. These models allowed for interactive negotiation preparation, enabling arguments and counterarguments to be tested in advance (UNU-CPR, 2025).

5.1.2 Inclusive Representation in Peace Processes

AI can meaningfully enhance inclusivity and engage diverse voices in peace processes. AI-assisted digital dialogues enable the views of previously excluded populations to be collected and analysed to ensure peace negotiations represent the needs of the whole population.

In 2021, the UN Department of Political and Peacebuilding Affairs (DPPA) partnered with the software company Remesh to launch a series of AI-enabled digital dialogues with citizens in Yemen, Libya, and Iraq. The platform enabled live conversations with up to 1,000 anonymous citizens, and AI analysed participants' responses in real-time, helping UN officials better understand the views of peace constituencies (UN, 2022).

5.1.3 Implementing Mediation and Negotiation

AI can improve multilingual communication in international negotiation and mediation processes through tools such as real-time translation and sentiment analysis. These technologies reduce language barriers, facilitate communication between parties, and make negotiation processes more efficient.

In cross-border commercial and civil disputes, AI-integrated Online Dispute Resolution (ODR) platforms use sentiment analysis to analyse parties' communications and emotions to identify escalating tensions (Vaishali; Vallimayil, 2025).

5.2 Risks

5.2.1 Bias and Exclusion

The outputs of AI-based mediation systems can be biased based on the data upon which the model is trained. This is a particular issue in conflict-affected countries, where vulnerable populations may not have the devices, connectivity, or digital literacy to participate in AI-enabled community engagement. Clustering models designed to find patterns in qualitative interview data may exclude marginalised and minority voices, exacerbating existing power imbalances (Belfer Center, 2026).

5.2.2 Escalation
In negotiation simulations, AI seeks escalation significantly more frequently (Rivera et al., 2024). AI systems operate exclusively according to rational, efficiency-based behavioral patterns and lack de-escalation dynamics. This often leads to different outcomes than to negotiations led by human mediators, applying aggressive pressure early and rarely recommending concessions.

5.2.3 Undermining Trust
The use of AI in mediation and negotiations may undermine the public's trust in the legitimacy of the process.

5.3 Governance

5.3.1 Human Mediation
Human mediation should remain at the centre of AI-assisted negotiation and conflict resolution processes. AI should function as a support tool, enhancing mediators' and negotiators' ability to analyse information and manage complex negotiations. Human mediators and interpreters remain essential to ensure contextual understanding, cultural sensitivity, ethical judgment, trust-building between parties, and the accurate interpretation of diplomatic nuances.

5.3.2 Transparency and Accountability
Transparency must be ensured both within a country and at the international level as a prerequisite for successful and ethical negotiations. Only by disclosing the underlying mechanisms can public trust be secured. Similarly, in international negotiations, the use of AI and the processing of the data involved should be transparently disclosed to lay the groundwork for successful cooperation. While the programs and datasets used can generally be disclosed, it remains particularly difficult to present the decision-making processes of AI systems themselves in a fully traceable manner.

5.3.3 Inclusive Peace Processes
When using AI-assisted tools to collect and represent views within a peace process, the data collection and analytical tools should be specifically designed to represent all population groups, including minority and marginalised voices. Specific support and training should be provided to areas of the population that lack the technology or skills to participate. Analytical algorithms should avoid overt clustering and grouping to ensure all views are represented in conclusions.

6. Peacekeeping and Justice

6.1 Technical AI Applications

6.1.1 Legal Mapping: AI-Supported Refugee Rights Analysis
RiMAP is UNHCR's online tool that shows, country by country, which rights refugees, asylum-seekers, internally displaced and stateless people have in law, and where protections are missing. In the peacekeeping and justice stage, this helps reveal legal gaps affecting safe return, documentation, housing, work and services (UNHCR, 2026). Its AI-powered Virtual Legal Assistant uses large language models and smart search to quickly locate and summarise relevant legal texts, which lawyers then review and refine. AI is therefore used to accelerate and scale rights-focused legal analysis after conflict (AI for Good, 2026).

6.1.2 Urban Recovery: AI Used to Classify Damaged Buildings

In Harasta, Syria, the project combines maps, satellite imagery and building-level information gathered by local volunteers to see which areas are destroyed and which remain habitable. This supports peacekeeping by guiding decisions on where families can safely return and where to prioritise reconstruction and services. AI comes in by training image-analysis programmes on photos so they can automatically sort buildings into categories like “destroyed” or “partly damaged” (Rifai et al., 2026). It thus turns fragmented local evidence into a coherent damage profile for recovery planning.

6.1.3 Arrival Prediction: AI Used to Anticipate Refugee Movements

In Uganda, the World Bank uses an AI model to predict refugee arrivals from the Democratic Republic of the Congo and South Sudan by examining patterns in conflict, climate, economic data and other signals. This is relevant to peacekeeping because accurate predictions allow Uganda to activate contingency funds and expand schools, clinics and water systems months before large inflows, easing pressure on host communities and improving support for refugees. The machine-learning model, tested on new data, can forecast short-term arrival trends with over 80 percent accuracy (World Bank Group, 2025), triggering early mobilisation of the national response fund. Here AI is used to plan ahead, making post-conflict support more proactive rather than purely reactive.

6.2 Risks

6.2.1 The ‘Black box’ Issue

In all three cases, the systems make decisions, but how they reach those decisions is not clear to either the people affected or the staff who use them. In Harasta, convolutional networks classify buildings using patterns in images that cannot be translated into simple, readable rules, while in Uganda the forecasting model combines more than 90 variables in ways that no one can fully trace or explain. As a result, neither side can say why a particular house is marked as damaged, why a service is denied, or why a movement is flagged as risky, which makes meaningful contestation impossible. In post-conflict conditions, this lack of clarity turns technical error, loss of housing, denial of services or even persecution into outcomes that are effectively unappealable.

6.2.2 Bias and Lack of Accountability

The data on which AI is trained in post-conflict settings are already incomplete and politically filtered, which produces distortions under the appearance of objectivity. RiMAP’s dashboards depend on reporting from UNHCR country offices, allowing politically sensitive states to under-report and shape the global picture in their favour. Harasta’s models inherit the bias of volunteer labellers, and Uganda’s forecast reflects the bias of variable selection. These distortions are almost never independently audited, turning biased outputs into instruments of selective pressure and legitimation without accountability for their consequences.

6.2.3 Data protection and cross-border flows

Legal records, damage assessments, and demographic forecasts are transferred between institutions and states without transparent rules on storage or reuse. This can lead to data reaching a party linked to the previous conflict, while data subjects are unable to withdraw consent or trace the path of their information. In 2021, UNHCR shared biometric data of Rohingya refugees with Bangladesh, which then passed it to Myanmar, the very state from which they had fled, turning data collected for protection into a tool for identifying a persecuted group (Human Rights Watch, 2021). Without mandatory standards for data minimisation, purpose limitation, independent oversight, and a ban on transfer to non-humanitarian actors such systems convert humanitarian infrastructure into a channel for surveilling the vulnerable.

6.3 Governance

6.3.1 Rule on explainability and the right to contest

“Any AI system whose outputs affect the legal status, property, or access to services of displaced persons must provide a human-readable justification for each decision and ensure an accessible mechanism for contesting it, with mandatory human review.” This can be embedded as a binding norm in the mandates of UN agencies (e.g. UNHCR, IOM, WFP) and in contracts with vendors, and as a legal requirement in the data protection laws of post-conflict states.

6.3.2 Rule on data protection and prohibition of reuse

“Data collected under a humanitarian or protection mandate may be used only for the originally stated purposes; its transfer to security, border, intelligence, or other non-humanitarian actors is prohibited, and cross-border transfer requires a documented risk assessment for the affected population.” The reason for this rule is that information given in trust for protection can easily become a tool of harm if it is reused or shared without strict limits. Strengthening UNHCR’s Data Protection Policy and the ICRC Handbook through such a clause would make it clear, in internal standards and bilateral agreements, that humanitarian data must not be handed over to authorities or contexts where it could contribute to surveillance, discrimination or persecution.

7. Conclusion

This paper set out to trace how artificial intelligence is used across the conflict cycle, from early-warning systems in periods of latent tension to targeting tools at the height of escalation, negotiation-support platforms during de-escalation, and damage assessment or refugee prediction tools in the aftermath. It is present throughout, and its role tends to grow as a conflict progresses. The paper does not take a position on whether AI should be used in these contexts. That decision has largely already been made by the states and organizations using it. These technologies are already employed in conflict settings, including early identification of risks, faster analysis during crises, improved humanitarian planning, and broader participation in peace processes. The challenge, then, is not the existence of AI, but ensuring that its use does not create greater risks than the problems it aims to solve.

Across every phase examined, the same problems recur. Outputs are difficult to explain, the window for human review before decisions are acted on is shrinking, and accountability structures, particularly around commercial suppliers of military and security systems, have not kept pace with the technology. Addressing this does not necessarily imply new legal frameworks. International humanitarian law and international human rights law already apply to AI systems used in conflict. The gap thus remains mostly in implementation. Effective governance should focus on keeping human judgment at the center of AI-supported decisions. This means establishing minimum standards for human review, mechanisms to explain AI-generated outputs, and clear lines of legal responsibility when something goes wrong, including for the private companies developing or operating these systems. Additionally, AI adoption in conflict settings is moving faster than international consensus. States, militaries, and international institutions are making decisions about its use now, in real conflicts, with real consequences for real people. Its future role in international peace and security will depend on the governance choices made today. With appropriate safeguards, it can strengthen prevention, response, mediation, and recovery. Without them, it risks accelerating and transforming the nature of conflict itself.

References

Abraham, Y. (2024). 'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine.

AI for Good (2024). ITU. https://unaihub.aiforgood.itu.int/activity-details.html?id=928

Anthropic. (2025, November 13). Disrupting the first reported AI-orchestrated cyber espionage campaign. https://www.anthropic.com/news/disrupting-AI-espionage

Armed Conflict Location and Event Data. Annual Report 2021. https://acleddata.com/about-acled/annual-report-2021/#s6

Belfer Center for Science and International Affairs. (2026). Reprogramming Peace: AI and the Future of Conflict Resolution.

Bo, M., et al. (2025). Submissions to the UN Secretary-General on autonomous weapons systems.

Bondar, K. (2024). Ukraine's AI-driven unmanned systems. Center for Strategic and International Studies (CSIS).

Cederman, L. E., & Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, 355(6324), 474–476. https://doi.org/10.1126/science.aal4483

Covino, C. (2025, November 15). The emergence of autonomous cyber attacks: Analysis and implications. Institute for AI Policy and Strategy. https://www.iaps.ai/research/autonomous-cyber-attacks

Ee, S., Covino, C., Labrador, C., Krawec, C., Kraprayoon, J., & O'Brien, J. (2025). Asymmetry by design: Boosting cyber defenders with differential access to AI. https://arxiv.org/abs/2506.02035

Europol. (2022). Facing Reality? Law Enforcement and the Challenge of Deepfakes. https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

Freedom House. (2023). Freedom on the Net 2023: The Repressive Power of Artificial Intelligence. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

Gill, R. (2023, April 17). What are sock puppets in OSINT. SANS Institute. https://www.sans.org/blog/what-are-sock-puppets-in-osint

Google Threat Intelligence Group. (2025, January 29). GTIG AI threat tracker: Advances in threat actor usage of AI tools. Google Cloud. https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

Grand-Clément, S. (2023). Artificial intelligence beyond weapons: Application and impact of AI in the military domain. UNIDIR.

Human Rights Watch. (2021). UN Shared Rohingya Data Without Informed Consent. https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent

Human Rights Watch. (2024). Israel/Palestine: AI-facilitated targeting in Gaza.

International Committee of the Red Cross. (2021). ICRC position paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach. *International Review of the Red Cross*, 102(913), 463–479. https://international-review.icrc.org/articles/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913

International Institute for Strategic Studies. (2025). The war in Ukraine and the growing role of UAVs.

International Panel on the Information Environment. (2025). Artificial intelligence and peacebuilding: Opportunities and challenges (Technical Paper TP2025.3). IPIE. https://doi.org/10.61452/RNGW7145

Johnson, J. (2025). Artificial Intelligence and Nuclear Stability: Understanding AI's Impact on Military Escalation Dynamics and Strategic Deterrence. GC REAIM Expert Policy Note Series.

Microsoft Threat Intelligence. (2025, June 30). Jasper Sleet: North Korean remote IT workers' evolving tactics to infiltrate organizations. Microsoft Security Blog.

Neuberger, A. (2025, January 15). Spy vs. AI: How artificial intelligence will remake espionage. *Foreign Affairs*. https://www.foreignaffairs.com/united-states/spy-vs-ai

Office of the United Nations High Commissioner for Human Rights (OHCHR). (2021). The Right to Privacy in the Digital Age (A/HRC/48/31).
https://documents.un.org/doc/undoc/gen/g21/249/21/pdf/g2124921.pdf

OpenAI. (2024, February 14). Disrupting malicious uses of AI by state-affiliated threat actors.
https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors

Rifai, G., Bachir, H., & Hajismail, S. (2026). AI-Assisted Community Mapping for Post-Conflict Housing Recovery: A Pilot in Harasta, Rural Damascus.

Rivera, J.-P., et al. (2024). Escalation Risks from Language Models in Military and Diplomatic Decision-Making. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*.

Schneier, B., & Penney, J. (2026, July 6). AI surveillance is being supercharged – and it will chill social progress. *The Guardian*. https://www.theguardian.com/commentisfree/2026/jul/06/ai-surveillance-policy

T. Bernard. (2023, April). Social cohesion technologies and online projects. Tech Policy Press.
https://www.techpolicy.press/social-cohesion-technologies-and-online-projects/

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2023). Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multistakeholder Approach.
https://unesdoc.unesco.org/ark:/48223/pf0000387339

UNHCR. (2026). Rights Mapping and Analysis Platform. https://rimap.unhcr.org/home

United Nations Institute for Disarmament Research Security and Technology Programme. (2025). Artificial Intelligence in the Military Domain and Its Implications for International Peace and Security: An Evidence-Based Road Map for Future Policy Action. UNIDIR.

United Nations. (2022). Making People's Voices Matter.

United Nations Secretary-General. (2023). Information Integrity on Digital Platforms: Our Common Agenda Policy Brief 8. https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-informationintegrity-en.pdf

Wan Rosli, W. R. (2025). Waging warfare against states: The deployment of artificial intelligence in cyber espionage. *AI and Ethics*, 5, 47–53. https://link.springer.com/article/10.1007/s43681-024-00628-x

Wohlfeld, M. (2010). An Overview of the Conflict Cycle. In O. Grech & M. Wohlfeld (Eds.), *Human Rights and the Conflict Cycle* (pp. 13–32). Mediterranean Academy of Diplomatic Studies.

World Bank Group. (2025). AI-Powered Refugee Forecasting: Preparing for Refugee Movements Before They Happen. https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/ai-powered-refugee-forecasting-preparing-for-refugee-movements-before-they-happen

Vaishali, R., & Vallimayil, P. L. (2025). Role of Artificial Intelligence in Online Dispute Resolution (ODR): A New Era of Mediation. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 11.