इलेक्ट्रॉनिकी एवं सूचना प्रौद्योगिकी मंत्रालय
MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY
सत्यमेव जयते

AI IMPACT SUMMIT भारत 2026 INDIA

सर्वजन हिताय | सर्वजन सुखाय
WELFARE FOR ALL | HAPPINESS OF ALL

INDIAai

INDIA AI IMPACT SUMMIT 2026

COMPENDIUM

UN WOMEN

FOR ALL WOMEN AND GIRLS

Real-World Impact of AI and Gender Empowerment

Table of Contents

Foreword

The India AI Impact Summit 2026 is being convened at a pivotal moment when global discourse on Artificial Intelligence (AI) is transitioning from principles and commitments to demonstrable outcomes and real-world impact.

Building on the momentum of major international forums, the UK AI Safety Summit, AI Seoul Summit, and the France AI Action Summit, the India - AI Impact Summit aims to strengthen global cooperation and set a shared vision for AI that benefits all, aligning AI deployment with public good, sustainability and national development priorities, and signalling that economic growth and social inclusion are co-equal objectives.

The India-AI Impact Summit is anchored in three Sutras: People, Planet, and Progress. These themes are not merely aspirational slogans; they are the guiding principles for the Summit's seven thematic working groups, or Chakras, which include critical pillars such as 'Human Capital' and 'Inclusion for Social Empowerment'. Our objective is to translate global principles into actionable policy frameworks that can be adopted by countries at varying levels of digital maturity, positioning the Global South as a co-architect of the global AI order.

Gender equality is foundational to the vision of this summit. India continues to advocate for the democratisation of AI, ensuring that women—as creators, consumers, and citizens—are central to the AI revolution. We are not only inviting women innovators to build for the world but also ensuring that AI systems are designed to be gender-responsive by default. At the national level, the IndiaAI Mission has made significant progress in supporting women-led AI solutions in

priority areas. By fostering an ecosystem that supports women-led startups, we are demonstrating how AI can drive economic growth that is truly inclusive.

The *Casebook on AI and Gender Empowerment*, developed in partnership with UN Women, represents a key milestone in these efforts. This casebook serves as a comprehensive reference for policymakers, innovators, and researchers. By documenting high-impact interventions that address critical gender-related challenges, it offers a practical blueprint for replication. Our objective is to empower stakeholders to adapt these proven successes to their unique local contexts, thereby accelerating the scaling of solutions across the Global South and supporting the digital transformation of social systems.

It stands as an example of how international collaboration can deliver practical and scalable outcomes for developing economies. As India prepares to host the AI Impact Summit 2026, the Ministry of Electronics and Information Technology reaffirms its commitment to fostering an AI ecosystem that is inclusive, ethical, and firmly grounded in real-world impact.

We acknowledge with appreciation the leadership of the Ministry of Women and Child Development, Government of India, in advancing gender-responsive innovation, the steadfast partnership of UN Women, and the support of the innovator and research communities in advancing responsible AI for gender equality. Together, these collective efforts demonstrate that artificial intelligence can serve as a powerful instrument for equitable development—transforming lives while upholding the highest standards of safety, transparency, and human dignity.

Shri S Krishnan
Secretary
Ministry of Electronics and Information Technology
Government of India

Foreword

The India AI Impact Summit 2026 marks a critical juncture in the global effort to ensure that advances in Artificial Intelligence translate into tangible and equitable outcomes for societies. As AI systems increasingly shape access to services, opportunities and public goods, it becomes imperative to centre women and girls within this transformation not only as beneficiaries of technology, but as active agents whose needs, realities, and leadership inform how AI is designed and deployed.

For the Ministry of Women and Child Development, gender empowerment is understood through a life-cycle lens encompassing health and nutrition, education, safety, livelihoods, and social protection. Artificial Intelligence, when applied responsibly, has the potential to strengthen service delivery across each of these domains. A gender-responsive approach to AI is therefore foundational to ensuring that technological progress advances equity rather than entrenching disparities. Innovations supporting women's safety, expanding access to financial services, improving maternal and reproductive health outcomes, and strengthening climate resilience for women farmers illustrate how technology can be aligned with social priorities.

The *Casebook on AI and Gender Empowerment*, developed by IndiaAI Mission in collaboration with UN Women, brings together these emerging trends in a structured and accessible manner. It presents a curated selection of AI applications that demonstrate measurable impact, ethical integrity, and potential for replication. Each case documents not only technological design and

deployment models, but also institutional arrangements, safeguards, and lessons learned offering policymakers and practitioners practical insights for adaptation and scale.

As India hosts the India AI Impact Summit 2026, the Ministry of Women and Child Development reaffirms its commitment to advancing innovation that strengthens women's agency, dignity, and economic participation. We acknowledge the leadership of the Ministry of Electronics and Information Technology and the IndiaAI Mission and value the steadfast partnership of UN Women in advancing this shared agenda.

This casebook stands as a testament to what is possible when technology is guided by social purpose. It reinforces a collective responsibility to ensure that Artificial Intelligence serves as a force for empowerment bridging gender divides, strengthening public systems, and contributing to inclusive and sustainable development.

Shri Anil Malik
Secretary
Ministry of Women & Child Development
Government of India

Foreword

The India AI Impact Summit 2026, the fourth in the series of international AI summits and the first one in the Global South, marks an important milestone in the evolution of collective efforts to translate the potential of AI into demonstrable and measurable outcomes.

Gender equity forms a crucial anchor of this vision. AI-enabled platforms are dismantling barriers to financial independence by offering women in rural geographies access to credit through alternative data scoring. Machine Learning applications are providing real-time alerts and support systems for women in urban spaces to enhance their safety, while AI-driven personalised healthcare is addressing maternal health and nutritional gaps with unprecedented precision.

The IndiaAI Mission's emphasis on democratised access to AI infrastructure and resources reflects the commitment to building solutions that are participatory and inclusive-by-design. This commitment to gender-inclusive innovation is visible in our approach which provides a platform to women-led AI solutions that solve large-scale public challenges.

The *Casebook on AI and Gender Empowerment*, developed in partnership with UN Women, represents the realisation of this collaborative vision. This Casebook is intended to serve as a valuable reference for policymakers, innovators and researchers seeking to adopt and scale proven gender-transformative AI solutions.

This joint initiative invited contributions from researchers, innovators and institutions worldwide, resulting in an enthusiastic response of over 235 submissions across geographies. Submissions underwent a multi-stage screening and technical evaluation process, assessed for relevance, quality and alignment with the objectives of ethical AI adoption and gender equality. Selected contributors submitted full chapters detailing solution design, deployment models, impact, ethical considerations and lessons learned. Each chapter was reviewed by an expert Evaluation Committee with representatives from the Ministry of Women and Child Development, Government of India and UN Women, to secure the inclusion of the most impactful, evidence-based and replicable solutions. The resulting Casebook presents a curated set of 23 AI applications from India and other regions, with use-cases ranging from innovations in women's safety to tools enhancing financial inclusion and climate resilience for women farmers.

We extend our sincere appreciation to the Ministry of Women and Child Development, Government of India for their leadership, and to UN Women for their steadfast partnership in advancing responsible and gender-responsive AI. Together, these collective efforts reaffirm that AI, when developed and deployed responsibly, ethically and inclusively, can serve as a powerful instrument for bridging the gender divide and advancing global equity.

Shri Abhishek Singh
CEO, IndiaAI Mission
Additional Secretary, Ministry of Electronics and Information Technology
Government of India

Note from Knowledge Partner: UN Women

India's leadership in convening this knowledge initiative, together with the Ministry of Electronics and Information Technology's commitment to gender-responsive innovation across the Global South, offers a compelling vision of what is possible when technology serves both development and equity. UN Women is honored to partner with the Government of India to advance an artificial intelligence (AI) ecosystem that supports national priorities.

This *Casebook on AI and Gender Empowerment*, launched jointly by the IndiaAI Mission, Ministry of Electronics and Information Technology, Government of India and UN Women at the India AI Impact Summit 2026, brings together practical examples of AI solutions advancing gender equality across diverse contexts. Drawing on 235 submissions from innovators working in more than 50 countries, this casebook reflects a deliberate effort to highlight context-driven and actionable approaches, and leverages UN Women's global, regional and country networks. This casebook provides practical examples of implementation grounded in lived realities.

UN Women's partnership with the IndiaAI Mission on this casebook aligns with India's vision for women-led development, in which women's economic participation and technological leadership are recognized as essential drivers of national progress. The solutions documented here demonstrate that gender equality and economic growth are not competing objectives; they are mutually reinforcing. When women participate as AI innovators, entrepreneurs and decision-makers, technologies become more effective, markets expand and development outcomes improve.

The cases featured highlight AI solutions that have moved beyond concept to real-world application, demonstrating measurable gender outcomes and responsible design. Spanning health, safety, education, economic empowerment, agricultural advisory services for women farmers and AI-supported access to justice, the 23 selected cases deployed across nine countries illustrate how gender-responsive AI can be utilized in diverse sectors and institutional settings. Contributors represent innovators from across Africa, Latin America and Asia, underscoring the leadership of the Global South in advancing practical, gender-responsive AI applications.

The casebook contributes to advancing Sustainable Development Goal 5 and reinforces UN Women's commitment to ensuring that digital transformation strengthens, rather than undermines, women's rights.

Collectively, these examples demonstrate that when grounded in lived realities and deployed with care, AI can deliver meaningful and replicable gains for women and girls. Each selected case underwent multi-stage screening and double-blind technical review and was required to demonstrate documented deployment, measurable outcomes and clear ethics, governance and safeguarding mechanism. In addition, cases were required to document data protection measures, bias mitigation approaches and accountability safeguards to support responsible, rights-based deployment.

This casebook was developed under the leadership of the Ministry of Electronics and Information Technology (MeitY) and the IndiaAI Mission. UN Women acknowledges the guidance provided by:

- Shri S. Krishnan, Secretary, Ministry of Electronics and Information Technology
- Shri Abhishek Singh, Chief Executive Officer, IndiaAI Mission
- Smt. Kavita Bhatia, Chief Operating Officer, IndiaAI Mission
- Shri Mohammed Y. Safirulla K., Director, Ministry of Electronics and Information Technology
- Smt. Shikha Dahiya, Joint Director, Ministry of Electronics and Information Technology

UN Women also acknowledges the support of the Ministry of Women and Child Development (MWCD) to this casebook, under the stewardship of Shri Anil Malik, Secretary, MWCD, the nodal ministry for gender issues in India.

Under the overall guidance of Ms. Kanta Singh, Country Representative a.i., UN Women India Country Office (ICO), Ms. Sudeshna Mukherjee, Head of Communications at UN Women ICO and member of the UN in India Digital Task Team, led the technical coordination and curation of the casebook. She was supported with inputs from Ms. Radhika Jhalani, Consultant, IndiaAI Mission, and Mr. Emad Karim, Advocacy, Campaigns and Innovation Lead, UN Women Regional Office for Asia and the Pacific.

UN Women appreciated the detailed assessment and final selection of the featured case studies by the evaluation committee, chaired by Shri Daya Shankar, Director, Ministry of Women and Child Development, and comprising of Smt. Shikha Dahiya, and Ms. Kanta Singh.

UN Women acknowledges the expert external evaluators who conducted the double-blind review of abstracts and the first round of case study assessments:

- Ms. Supriya Kulkarni, The Ethics Company
- Ms. Megan Bentley, Victor Pineda Foundation
- Ms. Kay Kim, GIZ GmbH
- Mr. Himanshu Joshi, COHUMAIN Labs
- Ms. Taylor Wong, LSE Generate

The casebook further benefited from the contributions of UN Women colleagues across offices. Ms. Carlotta Aiello and communications colleagues at UN Women Headquarters, together with Ms. Montira Narkvichien and the communications team of the Regional Office for Asia and the Pacific (ROAP), provided editorial guidance and supported outreach efforts. Mr. Rodrigo Herrera, UN Women Regional Office for the Americas and the Caribbean, supported the socialization of the global call for submission of abstracts. Ms. Federica Tornincasa, UN Women ROAP, and Ms. Urvashi Mitra, UN Women ICO, provided review and coordination support, while Ms. Himanshi Saini, UN Women ICO, contributed communications support. The UN Women country operations team provided procurement and publication support.

By foregrounding evidence, safeguards and lived experience, this casebook demonstrates that gender-responsive AI, when central to design, is implementable, scalable and measurable. It offers policymakers, technologists and development partners practical models for embedding gender equality within AI governance, innovation and deployment so that no woman or girl is left behind.

Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
DPI	Digital Public Infrastructure
GenAI	Generative Artificial Intelligence
GoI	Government of India
ICT	Information and Communication Technology
IndiaAI	India Artificial Intelligence Mission
LLM	Large Language Model
MeitY	Ministry of Electronics and Information Technology
MWCD	Ministry of Women and Child Development
NLP	Natural Language Processing
SDGs	Sustainable Development Goals
STEM	Science, Technology, Engineering and Mathematics
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women

Case Study 1

EDUCATION

AtenIA: Blending Physical and AI-Based Mentorship to Reduce Gender Gaps in STEM

Author: Nataly Vasquez Alzamora, Peru

Overview

Background & Problem Statement

In Peru, despite the presence of economically vital and socially responsible sectors driving national development, a significant gender gap persists in STEM fields. Research indicates that gender stereotypes take root early; for example, nine out of ten young girls (aged 6–8) associate engineering with masculine skills (UNESCO Regional Chair on Women, Science and Technology in Latin America, 2019). These self-limiting beliefs are reinforced by cultural stereotypes that frame key economic industries as masculine environments. As a result, the country fails to fully leverage approximately half of its potential talent pool. Traditional non-AI educational tools, such as static textbooks, have proven insufficient to provide the personalized and interactive mentorship required to challenge entrenched biases and inspire the next generation of women leaders.

Solution Summary

This 'phygital' solution (an approach that integrates physical materials with digital interactions) supports engagement with key economic sectors by combining an illustrated book with AtenIA, a specialized AI assistant. Using natural language processing, the solution transforms reading into an interactive mentorship experience, enabling students to engage in dialogue with role models and democratizing access to knowledge through advanced digital tools.

The AI Solution

Solution Description

This solution is centered on AtenIA, a customized large language model designed to function as a continuous mentor in STEM. Unlike general-purpose AI models, AtenIA operates within a closed-loop architecture to ensure child safety and data privacy. The system employs voice-to-text functionality, enabling verbal interactions and supporting inclusivity among young users with varying literacy levels. This feature is particularly important for bridging digital divides among children. Trained on a curated dataset focused on gender equality, innovation and

Large-scale vocational workshop held in the mining regions of Tacna and Moquegua. Female high school students (ages 14–17) participate in a live, interactive demonstration of the AtenIA virtual assistant, helping to bridge digital gaps in education. Photo: STEMLAB.

local context, the model creates a personalized narrative experience that actively reshapes users' perceptions of STEM fields.

Deployment & Implementation

Stage of Deployment

The initiative is currently implemented across 11 regions in Peru (Ancash, Arequipa, Cajamarca, Cusco, Huancavelica, Junín, Lambayeque, Lima, Moquegua, Pasco and Tacna) and has been under development since March 2025. It was developed exclusively by STEMLAB, with support from non-profit organizations and social enterprises. The solution operates on a hybrid infrastructure model in which a physical book serves as the primary engagement tool, while a mobile

Strategic presentation of the book and the AtenIA assistant to key stakeholders from civil society and the private sector. Engaging industry leaders supports the partnerships required to scale the initiative nationwide. Photo: SNMPE.

device connects users to the AI service hosted in the cloud. The system is currently integrated into school programmes as a technology-enabled educational resource for teachers and is also used in vocational talks. National institutions play a promotional role, with the initiative being advanced in collaboration with allies from the Peruvian government.

Users & Scale

Primary users are school-aged children (6–17 years), with a specific focus on girls in rural communities, as well as teachers. The initiative currently reaches more than 1,500 students and 120 teachers across approximately 200 communities. User engagement follows a hybrid model in which children colour the physical book and scan a QR code or link to interact with AtenIA. This approach helps bridge digital divides in areas with intermittent connectivity.

Impact & Outcomes

Measured Gender Impact

The intervention produced marked gender-related outcomes. Participants' intention to pursue a STEM career increased from 9 per cent to 76 per cent following the intervention, representing an 8.4-fold increase. Exposure to a broader range of role models also expanded significantly: prior to the intervention, 85 per cent of participants identified their mother as their only female role model, while after participation, 95 per cent also identified female engineers featured in the book as role models. In addition, 93 per cent of users reported recognizing women as active contributors in sectors critical to national development, and 92 per cent demonstrated an increased understanding of the importance of STEM fields, helping to challenge perceptions of STEM as an exclusively male domain.

Ethics, Governance & Safeguards

The system incorporates a set of ethics, governance and safeguarding measures designed to protect users and ensure responsible deployment. It operates within a closed-loop environment in which no data from minors is sold or used for external training, and all user interactions are anonymized to protect children's privacy. Bias mitigation is embedded in the design, with AtenIA trained to identify and reject gender-biased prompts and to actively redirect interactions towards highlighting women's competence and leadership in STEM fields. Oversight mechanisms are provided through content curation by STEMLAB education experts and validation by the women featured in the book, ensuring authenticity, accuracy and safety. Inclusivity is further supported through voice-to-text functionality, which helps ensure that literacy barriers do not prevent access to learning materials or role models.

Lessons & Replicability

What Worked

The phygital approach proved effective in rural Andean contexts by addressing connectivity constraints. The physical book served as a trusted, tangible anchor, while the AI component enabled scalable and personalized mentorship. Featuring real-life protagonists fostered strong emotional connections, and government support for reducing digital divides facilitated rapid expansion and diversification.

Challenges Encountered

The initiative encountered a small number of implementation challenges. While the book itself functions offline, interaction with the AI component requires

internet connectivity, which limited real-time engagement in highly remote areas. In addition, some cultural resistance was observed in more traditional settings, where STEM education for girls was initially perceived as unnecessary, requiring sustained community sensitization efforts to address these attitudes.

Key Lessons for Replication

Integrating physical materials with AI can effectively overcome infrastructural barriers in underserved regions. Trust-building through tangible entry points, scalable digital mentorship and institutional support are central to success. Replication requires adaptation to local contexts, strong partnerships and strategies that explicitly address digital divides.

Conclusion & Future Directions

This case demonstrates how generative AI can help address gender gaps in STEM education. Building on early results, the phygital model is being expanded to sectors such as construction, finance and health. By adapting AtenIA to these domains, the initiative aims to replicate gains in girls' interest in STEM and support more inclusive participation across the entire economic spectrum.

References

Inter-American Development Bank. (2019). *The future of work in Latin America and the Caribbean*.

RPP Noticias. (2024). *Interview with Nataly Vásquez: STEMLAB launches book inspiring girls in engineering* [Video]. https://www.youtube.com/watch?v=cjrhG1pRRCg

STEMLAB. (2024). *AtenIA Virtual Assistant Demo: Live interaction with users* [Video]. https://youtube.com/shorts/WoL4FVIFT70

STEMLAB. (2025). *Internal impact monitoring data of the "Mujeres STEM en Minería" programme* [Unpublished raw data].

UNESCO. (2021). *UNESCO science report: The race against time for smarter development*.

World Economic Forum. (2025). *The future of jobs report 2025*.

Case Study 2

SAFETY

HELPSTiR: Hyperlocal AI for Benefits Access

Author: Alice Luke, India

Background & Problem Statement

Women and children from marginalized communities in India face layered structural barriers to accessing essential social protection, including healthcare, shelter, protection from gender-based violence, government schemes and welfare benefits. Gender norms restrict phone ownership and mobility, while documentation gaps and low awareness further limit access to available schemes. Survivors of violence and trafficking often lack safe and trusted reporting channels. At the same time, welfare institutions operate in silos, civil society actors have limited visibility across cases, and frontline responders are unable to detect emerging vulnerabilities in real time.

Evidence from the HELPSTiR pilot showed that out-of-school children were predominantly girls, underscoring how gendered vulnerability remains unseen in the absence of proactive systems. These gaps pointed to the need for a gender-responsive, AI-enabled infrastructure capable of surfacing, classifying and connecting beneficiaries to hyperlocal support.

Solution Summary

HELPSTiR is an AI-powered platform that enables civil society actors to raise hyperlocal help requests on behalf of vulnerable women and children and automatically match them with nearby local NGOs, shelters, healthcare providers and welfare linkage organizations. By removing digital access barriers, the system accelerates gender-responsive benefits delivery across multiple domains.

The AI Solution

Solution Description

HELPSTiR uses AI and geospatial intelligence to process and route hyperlocal help requests. A civil society actor – such as a volunteer, NGO worker, teacher, Accredited Social Health Activist (ASHA) worker, social worker or passerby – identifies a vulnerable girl, woman or child and, with consent where possible, submits a help request. Exceptions to consent requirements apply in cases involving gender-based violence, trafficking, child safety or education-related exploitation, including child marriage.

Requests are submitted using minimal fields, including age, approximate location, need category (e.g., education, healthcare, shelter, gender-based violence protection or welfare schemes) and, where appropriate, an optional photo. Machine-learning models classify requests, while geolocation algorithms broadcast them to verified local NGOs operating in the relevant domain and vicinity. The first qualified local NGO accepts the request and delivers support. This mediated design removes the requirement for beneficiaries to own phones, be digitally literate, possess documentation or have safe mobility – barriers that disproportionately affect marginalized women and girls.

100

registered users facilitated support for more than 200 children, who were routed back into education

Deployment & Implementation

HELPSTiR is currently at the pilot stage of deployment, and is being implemented in Delhi, India, with future prospects for expansion to Tier 1, Tier 2 and Tier 3 cities. Platform development and NGO onboarding took place between January and April 2025, followed by pilot implementation in informal settlements from May to September 2025. A public launch is planned for November 2025.

The platform was developed by HELPSTiR and is implemented in partnership with local NGOs, including Giggles, Hope On, MAD, Sankalp Society and Tare Zameen. The underlying infrastructure comprises an Android application for submitting help requests, a cloud-based backend, GPS-based matching functionality and NGO dashboards. Local NGOs manage cases through AI-powered corporate social responsibility and beneficiary dashboards. Within the institutional model, local NGOs act as delivery nodes, while government and corporate social responsibility stakeholders are envisioned as scale-phase partners within broader welfare, health and safety ecosystems.

Users & Scale

The primary platform beneficiaries are women and children, while the primary users are civil society actors who raise help requests on their behalf. This approach directly addresses barriers related to phone access, literacy and mobility. During the pilot phase, the platform reached 50 girls through five local NGO partners, with more than 200 civil society actors pre-registered prior to launch. In total, 100 registered users facilitated support for more than 200 children, who were routed back into education.

Impact & Outcomes

Measured Gender Impact

The pilot demonstrated that civil-society-mediated reporting can surface otherwise invisible gendered needs. In the education pilot, 100 per cent of identified out-of-school children were girls; 98 per cent lacked awareness of available services, and 90 per cent had documentation but were unable to access schemes. Similar patterns were observed across healthcare, documentation, welfare and safety support, where beneficiaries did not self-report due to stigma, coercion or mobility restrictions. HELPSTiR reduced case detection and matching times from weeks or months to minutes. The model contributes to SDG 5 (Gender Equality), SDG 3 (Health), SDG 4 (Education) and SDG 16 (Justice), with particular relevance to targets 5.2 (eliminating gender-based violence and trafficking), 5.b (ICT access) and 5.4 (recognizing and valuing unpaid care and domestic work).

Ethics, Governance & Safeguards

Consent is obtained wherever possible; however, gender-based violence, trafficking and education cases involving exploitation or safety risks may bypass explicit consent in order to prevent harm. Sensitive data fields, including religion, caste and exact address, are optional and encrypted, and photos are not required in cases of gender-based violence or trafficking. Role-based access controls limit visibility of location data, and local NGOs undergo verification prior to onboarding. Audit logs, case-status tracking and community feedback mechanisms support accountability. Women participated as contextual informants in designing user interfaces and safety flows. Safeguards are aligned with India's emerging data protection norms and international human rights principles.

Lessons & Replicability

What Worked

Civil-society mediation proved effective in removing digital and mobility barriers. The use of AI and geolocation enabled faster matching across education, healthcare, gender-based violence, trafficking prevention, shelter and welfare schemes, while verified local NGO networks strengthened trust and cultural sensitivity.

Challenges Encountered

Uneven capacity among local NGOs resulted in variability in fulfilment, and digital distrust required sustained sensitization efforts. In many cases, phone numbers

belonged to other household members who were unreachable, complicating follow-up (the use of alternate contacts helped mitigate this challenge). Photo capture was not always feasible due to cultural norms or risks related to gender-based violence, which limited verification and occasionally slowed triage.

Key Lessons for Replication

Successful replication depends on the availability of localized NGO networks, privacy-safe geospatial infrastructure and procedures sensitive to gender-based violence. Language and cultural adaptation are critical in contexts where gender norms restrict phone ownership. Institutional partnerships with welfare departments, shelters, police and health systems are essential to support scalability and long-term sustainability.

Conclusion & Future Directions

HELPSTiR demonstrates that AI-enabled hyperlocal help requests can make invisible gendered vulnerabilities visible and actionable in real time. The model is scalable across cities and applicable to healthcare, gender-based violence, trafficking, shelter and cash-transfer benefits. Future phases will focus on deeper institutional integration and multi-city expansion.

References

HELPSTiR. (2025). *AI-powered HELPSTiR pitch deck (Version 1)*. HELPSTiR Internal Publication.

HELPSTiR. (2025). *Geolocation-based help pilot report 2025*.

Case Study 3

AI LITERACY

YASHODA AI: Digital Rights on Paper, Barriers in Practice

Authors: Parul Madan, Alisha Butala, India (with pilot in Tanzania)

Background & Problem Statement

For many women across the Global South, going online is not empowering; it is uncertain, confusing and often unsafe. In rural and semi-urban India, many women still do not own smartphones. Others use them cautiously, unsure of what is real, whom to trust or what to do when something goes wrong. As the world moves rapidly towards an AI-driven future, these women are being left at the margins of transformation, excluded from opportunity and often exposed to new forms of harm.

Digital change has outpaced support systems. Scam calls with familiar-sounding voices, artificial images that appear real and pressured messages demanding urgent action are increasingly targeted at women with limited digital literacy. A single click can cost savings, dignity and peace of mind. When harm occurs, fear and stigma often silence women, while legal remedies remain distant and inaccessible.

Government initiatives (such as Digital India, PMGDISHA and Cyber Swachhta Kendra), public campaigns and pre-call warning messages have expanded access and awareness. Yet these efforts often fail to reach women in ways that reflect their emotional, linguistic and social realities; knowing that a right exists is not the same as knowing how to exercise it.

YASHODA AI begins at this gap. It engages women within their everyday realities, helping them recognize risk, build safe digital habits, understand their rights and respond with confidence. The focus is not on teaching technology for its own sake, but on restoring agency, safety and dignity in an increasingly digital world. By centring women's lived realities through practical AI literacy, digital rights awareness and everyday safety practices, the programme enables women not only to go online, but to do so with confidence, dignity and resilience.

The AI Solution

Solution Summary

YASHODA AI is a women-led AI literacy and digital safety programme delivering

tailored training for police officials, Anganwadi workers, women from self-help groups (SHGs), students and community leaders. It translates awareness into protection, action and agency through AI-enabled language, voice and visual tools. The programme builds skills to identify AI-driven harms, practise digital hygiene and to access legal remedies, helplines and reporting systems.

5,500

women across 29 cities in 12 Indian states and 4 Union Territories were served by Yashoda AI

Solution Description

YASHODA AI is a human-centered, blended AI solution that combines accessible AI tools with facilitated, in-person learning to support women's digital safety and understanding of AI-enabled risks. The system functions primarily as a learning and decision-support mechanism, designed to build awareness and confidence rather than to automate decision-making or enforcement.

The programme employs natural language processing, which enables computers to understand and respond to human language, to support explanations, interactions and demonstrations in local languages through chat-based and voice-supported tools. Machine learning, which allows systems to recognize patterns in data, and computer vision, which enables analysis of images and videos, are used in guided demonstrations to show how deepfakes, voice cloning, scam messages and impersonation attempts are created and detected. These demonstrations are presented alongside real-world examples so participants can recognize warning signs in everyday digital interactions.

The solution is gender-responsive in both design and delivery. Training modules are adapted to the roles and lived realities of different groups, and sessions are led by cyber experts, legal practitioners and women trainers, ensuring that technical knowledge is grounded in practical and legal contexts. The "AI Sakhi" mentor network extends this reach by training women from within the community to facilitate peer-to-peer discussions in familiar environments.

Explainability and trust are central to the approach. All AI-based examples are accompanied by clear explanations, visual comparisons and step-by-step walkthroughs. This emphasis on transparency supports informed judgement, helping women understand not only what may be risky, but why – and what options are available to them.

Deployment & Implementation

YASHODA AI is at an early scale stage of deployment. Following pilot development

in 2024, the programme expanded in 2025 to reach over 5,500 women across 29 cities in 12 Indian states and four Union Territories, spanning rural, semi-urban, institutional and community settings. Primary users are women and girls, accounting for approximately 90 per cent of beneficiaries. It was also piloted internationally through an online workshop in Tanzania.

The solution was co-developed and overseen by the National Commission for Women and Future Shift Labs, with local deployment through state women's commissions, cyber cells, NGOs and community partners. Infrastructure requirements are minimal, relying on smartphones, basic internet access, printed toolkits and familiar platforms such as WhatsApp. The programme integrates seamlessly into existing community, education and frontline service workflows, with the National Commission for Women providing institutional leadership, coordination and alignment with national digital inclusion and safety priorities.

Impact & Outcomes

Measured Gender Impact

Pre- and post-workshop assessments indicate a marked increase in digital awareness, with participants reporting an improved ability to identify scams, deepfakes and unsafe online practices. Women demonstrated greater confidence in using reporting mechanisms, helplines and privacy controls, indicating enhanced decision-making agency. While direct income effects have not yet been measured, participants, particularly SHG members and students, reported increased readiness to use digital tools for education, livelihoods and entrepreneurship. Social impacts include heightened online safety awareness and peer knowledge-sharing through AI Sakhi networks. Compared to baseline conditions of low AI familiarity and limited cyber awareness, the programme contributes to SDG 5 targets on women's empowerment, digital inclusion and equal participation in public and economic life.

Ethics, Governance & Safeguards

YASHODA AI follows a human-centered and privacy-conscious approach, embedding ethical safeguards into programme design and delivery. Participant information is collected only through voluntary pre- and post-workshop surveys, with clear explanations of consent, purpose and use provided in advance. No sensitive personal data is collected or shared externally; data are used solely for programme learning, improvement and limited follow-up communication. AI tools are used for learning and demonstration, not for surveillance, profiling or automated decision-making.

Bias risks are addressed through locally grounded content and continuous feedback from women participants and trainers. Training materials are adapted

to regional language and context, avoiding stereotypes and ensuring cultural relevance rather than applying a one-size-fits-all model. Governance and oversight are jointly led by Future Shift Labs and the National Commission for Women.

Women participate as trainers, AI Sakhi mentors and contributors to content refinement. Potential unintended effects, such as fear or over-reliance on tools, are addressed through an emphasis on human judgement, legal awareness and support pathways. By prioritizing agency, transparency and inclusion, the programme advances gender equality (SDG 5), strengthens trust in justice mechanisms (SDG 16) and reduces digital inequalities for underserved women (SDG 10).

Lessons & Replicability

What Worked

YASHODA AI's effectiveness emerged through continuous, iterative learning rather than fixed design. Training approaches evolved based on pre- and post-workshop surveys, real-time field feedback from trainers, expert inputs and participant reflections. Adapting sessions to socio-economic realities, maintaining a personal and trust-based approach and tailoring content across audiences proved central to sustained engagement and relevance.

Challenges Encountered

Implementation revealed wide variation in women's exposure to technology, particularly in rural settings where some participants lacked phones and relied on male family members for access. Fear, hesitation, language diversity, limited connectivity and involvement shaped by lived realities – where family responsibilities and survival needs often prevail – influenced the nature of participation and engagement. Sessions therefore required stronger visual aids, slower pacing, trust-building and continuous adaptation while maintaining coherence during scale-up.

Key Lessons for Replication

A central lesson is that AI adoption does not follow a single trajectory. Even within the same country, users engage from vastly different starting points, ranging from ethical use among advanced students to confidence-building for first-time users. Effective approaches require differentiated pathways, trust-led facilitation, low-tech entry points and institutional support. Governments should design layered strategies rather than uniform AI rollouts.

Conclusion & Future Directions

YASHODA demonstrates how starting from women's everyday realities can strengthen meaningful digital participation. As it expands through multilingual tools, community partnerships and youth-led outreach, the programme offers a practical, locally grounded model aligned with global calls for inclusive and gender-responsive digital ecosystems.

References

GSMA. (2023). *The mobile gender gap report*. https://www.gsma.com/gender-gap/

International Telecommunication Union. (2023). *Measuring digital development: Facts and figures*. https://www.itu.int

Ministry of Electronics and Information Technology. (n.d.). *Digital India*. https://www.digitalindia.gov.in

National Crime Records Bureau. (2023). *Crime in India: Cyber crime*. https://ncrb.gov.in

OECD. (2019). *Human-centered artificial intelligence*. https://www.oecd.org/ai

OECD. (2021). *Bridging the digital gender divide*. https://www.oecd.org

Reserve Bank of India. (2022). Frauds in digital payments. https://www.rbi.org.in

UN Women. (2022). *Online violence against women*. https://www.unwomen.org

UNESCO Institute for Statistics. (2022). *Digital literacy and skills*. https://uis.unesco.org

UNESCO. (2023). *AI literacy for inclusive societies*. https://www.unesco.org

World Economic Forum. (2023). *Global gender gap report*. https://www.weforum.org

Case Study 4

SAFETY

AI-Enabled Safeguarding in Grassroots Sport

Author: Leacky Ochieng Amoya, Kenya

Overview

Background & Problem Statement

Women and gender-diverse athletes in grassroots football face systemic exclusion driven by gender norms, discrimination, safety risks and limited access to health and reporting services. In the Global South, these barriers are intensified by informality in sport structures, low institutional accountability and data invisibility, leaving abuse, injury and exclusion underreported and unaddressed. Prior interventions relied on manual reporting, workshops and fragmented services that were often inaccessible, unsafe or slow to respond. The absence of real-time data, anonymous reporting and evidence-based decision-making limited effectiveness. Non-AI approaches could not scale, adapt or generate the actionable insights required to transform gender-inequitable sport ecosystems.

Solution Summary

The solution deploys an AI-enabled digital platform that uses machine learning and natural language processing to enable anonymous reporting, risk detection and access to health and protection services in grassroots football. Predictive analytics generate insights that improve safety, participation and leadership opportunities for women and gender-diverse athletes.

The AI Solution

Solution Description

The AI system operates through a mobile-accessible platform that collects anonymized reports and participation data from women and gender-diverse athletes. Machine learning identifies recurring risk patterns related to safety, health and exclusion, while natural language processing analyses text-based reports to flag urgent cases. Predictive analytics support early intervention by highlighting high-risk environments. The solution is gender-responsive through co-design with women athletes, use of gender-disaggregated data and bias checks during model training. Outputs are delivered through simple dashboards

and alerts, with clear indicators that explain why risks were flagged, supporting transparent and informed stakeholder decision-making.

500

women and gender-diverse athletes across five Kenyan communities have been reached by the AI platform

Deployment & Implementation

The AI platform is deployed in five urban and peri-urban communities in Kenya, reaching grassroots women and gender-diverse football clubs. Development and co-design took six months, followed by pilot testing in two clubs and phased scale-up over 18 months, with ongoing monitoring. KISPED Kenya, local clubs, AI developers and NGOs jointly manage the solution. Infrastructure requirements include smartphones or tablets, internet connectivity, digital reporting forms and dashboards. Integrated with existing health and protection services, the platform supports referrals, tracks participation and informs interventions. National institutions provide technical guidance, legitimacy and partial funding.

Users & Scale

Primary users are women and gender-diverse athletes in grassroots football clubs across five Kenyan communities. The platform currently serves approximately 500 beneficiaries and six partner clubs. Users access the system via a mobile application or through trained community workers, enabling report submission, receipt of alerts and engagement with health, safety and participation services.

Impact & Outcomes

Measured Gender Impact

The AI platform has reached approximately 500 women and gender-diverse athletes across five Kenyan communities. Anonymous reporting and early intervention increased access to health and protection services by 65 per cent compared to baseline, and club participation increased by 40 per cent. During early adoption, increased awareness of what constitutes harassment and growing trust in the platform led to an initial rise in reporting. As coordinated safeguarding measures, gender-norms education with male athletes and coaches, and faster accountability responses were implemented, harassment reports subsequently declined by 30 per cent. Users reported greater decision-making agency in sports and community activities. Digital literacy improved through platform engagement, and timely alerts reduced response times for safety concerns by 50 per cent. The intervention advances SDG 5 (Gender Equality) by promoting inclusion, safety and leadership for women and marginalized genders, demonstrating measurable social, health and empowerment outcomes.

Ethics, Governance & Safeguards

User data is anonymized and encrypted, with informed consent obtained prior to collection. The system incorporates gender-disaggregated datasets and bias checks to promote equitable outcomes. Women and gender-diverse athletes participated in co-design, governance and oversight, shaping ethical decisions and interface design. Accountability mechanisms include regular audits, community feedback loops and rapid response protocols for reported harm. Privacy, safety and fairness are embedded from inception rather than applied retrospectively. Potential unintended effects, such as misreporting or misuse of alerts, are monitored and mitigated through moderation and mandatory human review. All safeguards align with human rights, privacy and ethical standards for AI deployment.

Lessons & Replicability

What Worked

Success was driven by co-design with women athletes, strong community engagement and integration with local clubs and NGOs. Key design features included user-friendly mobile reporting, anonymized data handling and predictive alerts. Partnerships with county authorities and health services supported legitimacy and scale. Adaptation is feasible across diverse cultural contexts where similar stakeholder engagement and basic infrastructure are present.

Challenges Encountered

Key challenges included limited digital literacy, intermittent internet connectivity and initial scepticism from athletes and club managers. Cultural norms related to gender and privacy slowed early adoption, while resource constraints limited access to devices. These challenges were addressed through targeted training, local-language interfaces, community advocacy and partnerships with NGOs and county authorities to build trust and support scale-up.

Key Lessons for Replication

Early engagement with women athletes and community stakeholders is essential. Digital literacy and connectivity constraints require tailored training and offline or assisted access options. Adaptation must account for local cultural norms, language and infrastructure. Strong partnerships with clubs, NGOs and local authorities are prerequisites for success. Governments and organizations should prioritize co-design, trust-building and gender-responsive monitoring to support effective replication.

Conclusion & Future Directions

This AI solution advances gender equality by enhancing safety, participation and leadership opportunities for women and gender-diverse athletes. It can be scaled to additional communities and sports, with potential integration into broader health and protection systems. Future directions include expanded deployment, continued co-design and contributions to inclusive, evidence-driven AI ecosystems across the Global South.

References

Kenya National Bureau of Statistics. (2021). *Kenya demographic and health survey 2020–2021*. KNBS

United Nations Women. (2021). *Gender equality in sports: Promoting women's participation and leadership.* UN Women.

United Nations. (2020). *The sustainable development goals report 2020.* United Nations.

World Bank. (2022). *Women and digital technologies in Sub-Saharan Africa: Bridging the gender gap*. World Bank Group.

Case Study 5

LAW

NyayaSakhi-SWATI: AI-Supported Pre-Litigation Guidance for Survivors of Domestic Violence

Author: Dipali Dilip Awasekar, India

Overview

Background & Problem Statement

For many women experiencing domestic violence in India, the primary barrier is not the absence of law but limited access to legal knowledge. Financial dependence, fear of retaliation, stigma and family pressure further deter women from approaching lawyers or courts. When women do seek help, questions such as "What relief will I get?" and "How long will it take?" are typically answered based on individual lawyers' experience rather than evidence. Delays compound these challenges. India has approximately 47.3 million pending cases, including around 11 million civil matters, with about 58.5 per cent pending for more than one year (National Judicial Data Grid). Prior to NyayaSakhi-SWATI, no Protection of Women from Domestic Violence Act (PWDVA)-specific judgment corpus or data-driven tool provided early, personalized guidance at scale.

Solution Summary

NyayaSakhi-SWATI is India's first large language model- and retrieval-augmented generation-powered decision-support assistant for domestic violence survivors. Trained on NyayaDeepa, India's first real-world domestic-violence judgment corpus derived from Maharashtra courts, the system supports pre-litigation decision-making under the PWDVA, 2005. It provides estimates of likely statutory reliefs and approximate case duration, enabling safer, more informed and financially realistic decisions.

The AI Solution

Solution Description

NyayaSakhi-SWATI supports domestic violence survivors in understanding potential legal outcomes before filing a case. The system uses a large language model to process a woman's narrative and retrieval-augmented generation to retrieve and analyse relevant judgments from NyayaDeepa, generating responses grounded in similar cases. This approach differs from existing tools in India, which generally provide generic legal information or lawyer referrals without offering

insights into likely reliefs or timelines. The system is gender-responsive in that it is trained on women's domestic-violence cases, uses bias-aware prompts, protects user privacy and explains outputs by referencing patterns identified in retrieved judgments. This design supports transparent, evidence-based decision-making for survivors.

Deployment & Implementation

NyayaSakhi-SWATI is deployed in India, with an initial focus on Maharashtra, serving women from low-income and marginalized communities. Work began in 2022 with LAMP (Legal Analytics Model for Prediction and Prescription, a legal AI research and prototyping initiative) and the creation of the NyayaDeepa PWDVA judgment corpus. Early support of INR 100,000 from UN Women contributed to initial dataset creation and validation. The system subsequently progressed to cloud-hosted testing and was integrated into SWATI as the user-facing conversational interface. A closed pilot began with 41 users and expanded to 5,141 users across districts, including Solapur and Latur.

Users & Scale

The primary beneficiaries are domestic violence survivors, predominantly women and adolescent girls. In the current launch-preview stage, more than 2,114 women

and girls have used NyayaSakhi-SWATI. Access is link-based, enabling private, on-demand use via smartphones, with assisted access available where needed. Users receive statute-aligned guidance, insights on likely reliefs and estimates of case duration.

2,114

women and girls have been reached across Maharashtra during the launch-preview stage by NyayaSakhi-SWATI

Impact & Outcomes

Measured Gender Impact

NyayaSakhi-SWATI has reached more than 2,114 women and girls during its launch-preview stage across Maharashtra. User feedback indicates significant behavioural change: 83 per cent of users reported improved understanding of potential legal reliefs and greater clarity on whether to resolve their situation early or proceed through the courts. Women reported increased confidence in deciding whether to pursue litigation, negotiate or seek protection. Many avoided unnecessary legal costs by clarifying options before consulting a lawyer. The system reduced the time spent searching for information and lowered the fear associated with legal processes. Socially, women reported feeling safer and more respected. These outcomes support SDG 5 by strengthening women's legal awareness, autonomy, and access to justice.

Collecting user feedback from community members

Ethics, Governance & Safeguards

NyayaSakhi-SWATI protects user privacy through anonymous access, encrypted storage and explicit consent prior to interaction. No personally identifiable information is required to receive guidance. Bias mitigation measures include training on women-specific domestic-violence judgments and the use of safety-constrained prompts to prevent harmful or discriminatory outputs. Ethical and accountability safeguards are embedded through retrieval-grounded responses, audit logs and human review of flagged outputs. The research team monitors system behavior and investigates any reported harm. Women survivors contributed to usability testing, language design and safety feedback, ensuring that the system reflects lived experiences and remains respectful, protective and aligned with women's rights and dignity.

Lessons & Replicability

What Worked

A key success factor was the development of NyayaDeepa, India's first PWDVA-focused judgment dataset, which enabled evidence-based and trustworthy outputs. Grounding large language model responses through retrieval-augmented generation proved effective in generating guidance on reliefs and case duration. Combined legal and technical expertise was critical in ensuring accurate statute mapping. For replication, adaptation to local languages, court practices and socio-cultural contexts is essential.

Challenges Encountered

Accessing and structuring multilingual, unstructured domestic-violence judgments from Indian courts posed a significant challenge for model training. Variations in legal language and formatting increased data-preparation complexity. Digital literacy gaps, limited internet access and trust in AI also affected adoption, while regulatory and ethical constraints required careful handling of sensitive survivor data.

Key Lessons for Replication

Future deployments should invest early in structured legal datasets and multilingual coverage. Models must be adapted to local laws and court practices. Reliable internet connectivity, cloud infrastructure and combined legal–technical capacity are prerequisites. Government support for dataset creation and ethical oversight can enable safe and scalable adoption of gender-focused legal AI systems.

Conclusion & Future Directions

NyayaSakhi-SWATI advances gender equality by providing women with early, data-driven clarity on likely reliefs and timelines, reducing uncertainty and dependence. The approach has the potential to scale across India and extend to other women-focused legal frameworks. Planned next steps include multilingual expansion and broader statutory coverage.

References

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science, 2*, e93. https://doi.org/10.7717/peerj-cs.93

Awasekar, D. D. (2024). *Empowering women and girls: Assessing the impact of an online webinar on legal rights awareness and knowledge of the DV Act 2005 in India*. *Journal of Engineering Education Transformations, 37*(Special Issue 2).

Awasekar, D. D., & Lobo, L. M. R. J. (2025). *ProtectJC 1.0: Building India's first statute-aligned legal corpus to train machine learning models for relief prediction in domestic violence cases.* In *Proceedings of the International Conference on Advanced Networking and Applications Technologies (ICONAT 2025)*

Awasekar, D., & Lobo, L. M. R. J. (2024). *Artificial intelligence for legal assistance: A prescriptive analytics model integrating social emotional learning for assisting victims of domestic violence in India*. In S. Mishra, A. Kothiyal, S. Iyer, S. Sahasrabudhe, A. Lingnau, & R. Kuo (Eds.), *Proceedings of the International Conference on Technology 4 Education 2024*, (Vol. 1, pp. xx–xx). Springer. https://doi.org/10.1007/978-981-96-5761-2

Awasekar, D.D., & Lobo, L.M.R.J. (2025). *Empowering women through AI: A comprehensive chatbot for domestic violence awareness and legal support in India*. In M.E. Auer, TRüütmann. (Eds) *Futureproofing engineering education for global responsibility*. (Lecture Notes in Networks and Systems, Vol. 1281). Springer. https://doi.org/10.1007/978-3-031-83520-9_19

D. D. Awasekar and L. M. R. J. Lobo, *SWATI AI: Advancing AI models from rule-based frameworks to NLP-driven prescriptive analytics to assist victims of domestic abuse.* In *2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICAET63349.2025.10932279.

Protection of Women from Domestic Violence Act, 2005, No. 43, Acts of Parliament, 2005 (India). https://www.indiacode.nic.in

Malik, V., Bhattacharya, I., & Pandey, G. (2021). *ILDC for CJPE: Indian Legal Documents Corpus for court judgment prediction and explanation*. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics* (pp. 4046–4062). Association for Computational Linguistics.

Ministry of Law and Justice. (2019). *NyayaBandhu: Pro bono legal services*. Government of India. Retrieved from https://probono-doj.in

Nigam, S. K., Patnaik, B. D., Mishra, S., Shallum, N., Ghosh, K., & Bhattacharya, A. (2025). *TathyaNyaya and FactLegalLlama: Advancing factual judgment prediction and explanation in the Indian legal context*. arXiv, https://arxiv.org/abs/2504.04737

Nigam, S. K., Patnaik, B. D., Mishra, S., Shallum, N., Ghosh, K., & Bhattacharya, A. (2025). NyayaAnumana and INLegalLlama: *The largest Indian legal judgment prediction dataset and specialized language model for enhanced decision* analysis. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 11135–11160). Association for Computational Linguistics.

Srivastava, P. K., Raj, U., Patel, P., Shallum, N., & Bhattacharya, A. (2025). IBPS: Indian bail prediction system. arXiv preprint arXiv:2508.07592.

Surden, H. (2019). Artificial intelligence and law: *An overview. Georgia State University Law Review, 35*(4), 1305–1336.

Case Study 6

SAFETY

VoiceValor: Survivor-Led Auditing of AI Moderation

Authors: Dridhata Silwal, Himanshu Panday, Hameeda Syed; South Asia

Overview

Background & Problem Statement

Online gender-based violence and misogynistic hate speech undermine women's and gender-diverse people's ability to participate safely online. In South Asia, where language, caste, religion and region shape how abuse is expressed and experienced, harms are often highly intersectional and complex. Yet commercial AI moderation systems are typically trained on datasets that under-represent languages with limited digital footprints and training data, particularly where meaning depends heavily on cultural and social context. This leads to algorithmic failures: harmful comments are missed, while survivors' responses are sometimes incorrectly flagged.

Before VoiceValor, most responses relied on manual reporting or opaque platform enforcement approaches that are slow at internet scale and frequently misaligned with survivor safety and evolving perpetrator behavior.

Solution Summary

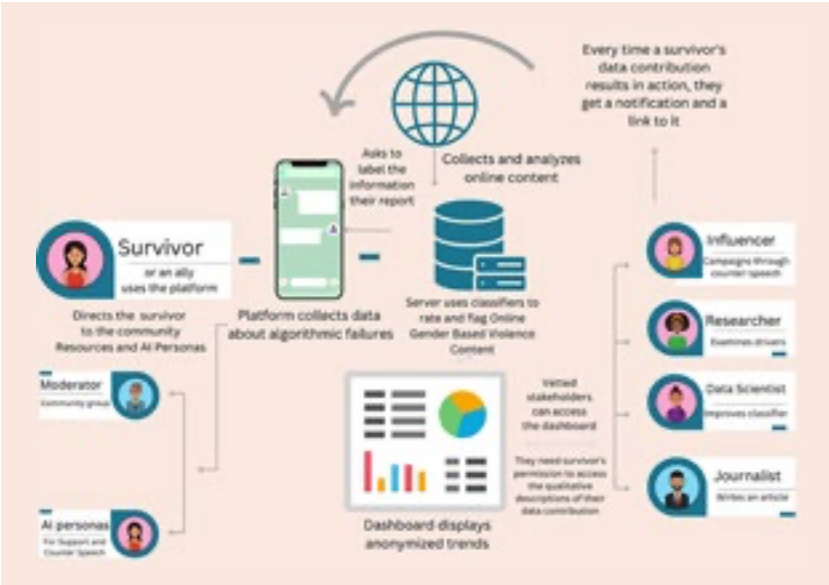
VoiceValor is a web-based auditing tool designed to uncover blind spots in social media content moderation. It uses community-led reporting and cultural-context analysis to identify forms of online toxicity, specifically misogyny and hate speech, that automated systems typically miss. Marginalized communities use VoiceValor to document and flag these algorithmic failures, while researchers and advocates can use the platform's insights to challenge bias and demand more equitable digital safety standards.

The AI Solution

Solution Description

VoiceValor enables users to audit and correct how AI toxicity classifiers interpret local expressions of misogyny. Using widely deployed commercial classifiers, the platform analyses survivors' digital experiences across common harm indicators used in trust and safety policies. It then presents transparent scores that show where algorithms fail.

Users can flag missed harms, down-rank false positives and provide rich contextual insights grounded in lived experience. Outputs are explainable at the individual comment level, allowing clear documentation for policymakers and digital platforms. The design and implementation operationalize design justice by introducing a new method through which survivors of technology-facilitated gender-based violence can exercise agency in algorithmic decision-making.



VoiceValor Visual Map

Deployment & Implementation

VoiceValor launched as a web-based tool in 2024, supported by iterative survivor-led testing and multilingual expansion. Users access the platform through a lightweight browser interface by pasting a content URL, analysing the material and submitting consent-based corrections. The system relies on stable connectivity and standard devices.

Implementation is led by Dignity in Difference, with community partners supporting onboarding and safeguarding. Survivor corrections generate structured datasets that can improve moderation pipelines, such as retraining commercially available AI toxicity classifiers. Where feasible, the tool aligns

with existing moderation workflows by translating lived experience into machine-readable feedback for model improvement. Where this is not possible, it supplements datasets used within existing technology infrastructure.

10,000

have been engaged and over 500,000 comments have been processed for analysis and review

Users & Scale

Primary users are women and gender-diverse digital users. To date, the platform has engaged more than 10,000 participants and processed over 500,000 comments for analysis and review. Access is free through the platform, and engagement is self-directed and consent-based.

Impact & Outcomes

Measured Gender Impact

VoiceValor's core contribution is the creation of an evidence base that survivors control. During early scaling, users processed more than 500,000 comments and contributed over 10,000 survivor inputs documenting misclassifications and missed harms, including more than 5,000 distinct algorithmic failure cases linked to cultural context.

This approach enhances women's decision-making agency by shifting their role from passive reporters to active auditors of moderation logic. Reported social impacts include improved safety planning and greater confidence among content creators who routinely face harassment, supported by safer-response guidance embedded in the platform. The work aligns with Sustainable Development Goals 5.1 and 5.2 by addressing discrimination and violence through strengthened detection and response capacity in digital public spaces.

Ethics, Governance & Safeguards

VoiceValor minimizes risk by collecting no direct personal identifiers and limiting stored data to what is necessary for model feedback and monitoring. Consent is explicit at the point of contribution, with clear warnings and opt-out options. Bias mitigation is embedded through survivor review cycles.

Community members regularly test indicators, flag gaps and guide updates, ensuring that definitions of harm are not determined solely by external

developers. Safeguards include trauma-sensitive moderation practices, exposure-reduction design and supportive pathways for users who encounter distressing content. Accountability is maintained through internal oversight by the civil society organization that hosts and operates the system, documented change logs and clear mechanisms for reporting harms or requesting removal of contributed material.

Lessons & Replicability

What Worked

Success depended on treating survivors as experts. Explainability was the most effective design choice: showing how models scored individual comments increased user engagement and confidence in correcting errors. Lightweight web access reduced participation barriers, while community outreach helped build trust and sustain engagement.

Challenges Encountered

Key challenges included user interface design, survivor safety and fatigue associated with revisiting harmful content, uneven connectivity and the difficulty of standardizing context across dialects and evolving slang. Platform resistance and limited transparency from large technology companies slowed the integration of community corrections into production moderation systems.

Key Lessons for Replication

Replication requires a strong digital community approach and an engaging user experience grounded in safety and consent. Tools should be paired with community governance rather than being dominated by a narrow set of institutional or technical actors. Governments and organizations should invest in survivor-led benchmarks and multilingual evaluation frameworks.

Conclusion & Future Directions

VoiceValor advances gender-inclusive AI by operationalizing survivor-led accountability. Those most affected by online abuse define what harm looks like, where systems fail and what safer detection requires. The approach is scalable across resource-poor languages and digital platforms. Next steps focus on improving user experience design, expanding multilingual coverage, strengthening institutional partnerships and developing sustainable pathways beyond single-provider moderation application programming interfaces.

References

Ananny, M., & Crawford, K. (2016). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability". *New Media & Society, 20*(3), 973-989. https://doi.org/10.1177/1461444816676645

Barocas, Solon and Selbst, Andrew D., *Big data's disparate impact*. California Law Review, 104(3), 671–732 http://dx.doi.org/10.2139/ssrn.2477899

Boyd, Danah & Crawford, Kate. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*. 15(5), 662–679. https://www.researchgate.net/publication/281748849_Critical_questions_for_big_data_Provocations_for_a_cultural_technological_and_scholarly_phenomenon

Citron, D. K. (2014). Introduction. In *Hate crimes in cyberspace*. Harvard University Press.

Costanza-Chock, S. (2018). Design Justice: towards an intersectional feminist framework for design theory and practice. In C. Storni, K. Leahy, M. McMahon, P. Lloyd, & E. Bohemia (Eds.), Design as a catalyst for change: DRS International Conference 2018 (pp. xx–xx). Design Research Society. https://doi.org/10.21606/drs.2018.679

Cotton, I., & Darragh, S. (2022). *New horizons in digital anthropology: Innovation in understanding humanity* (MOST/IGC/2022/013). UNESCO & LiiV Center. https://unesdoc.unesco.org/ark:/48223/pf0000382647.

Daniels, J. (2018). The algorithmic rise of the "alt-right." *Contexts, 17*(1), 60–65. https://doi.org/10.1177/1536504218766547

Downing, J. (2009). Digitizing race: Visual cultures of the internet by Lisa Nakamura [Review of the book *Digitizing race: Visual cultures of the internet, by L. Nakamura*]. *Journal of Communication*, 59(3), xxx–xxx. https://doi.org/10.1111/j.1460-2466.2009.01461.x.

Ginsburg, F. (2008). Rethinking the digital age. In P. Wilson & M. Stewart (Eds.), *Global indigenous media* (pp. 287–306). Duke University Press. https://doi.org/10.1215/9780822388692.

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge* (pp. xx–xx). Springer. https://doi.org/10.1007/978-94-007-4587-2_7

Horst, H. A., & Miller, D. (Eds.). (2012). *Digital anthropology* (1st ed.). Routledge. https://doi.org/10.4324/9781003085201

Jane, Emma. (2016). "Misogyny Online: A Short (and Brutish) History." 10.4135/9781473916029.

Jane, E. A. (2016). Misogyny online: A short (and brutish) history. In J. Daniels, K. Gregory, & C. T. McMillan (Eds.), *Digital sociologies*. SAGE Publications. https://doi.org/10.4135/9781473916029

Keller, J., Mendes, K., & Ringrose, J. (2018). Speaking "unspeakable things": Documenting digital feminist responses to rape culture. *Journal of Gender Studies, 27*(1), 22–36. https://doi.org/10.1080/09589236.2016.1211511

Lynn Dombrowski, Ellie Harmon, and Sarah Fox. (2016). "Social Justice-Oriented Interaction Design: Outlining Key Design Strategies and Commitments". In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Association for Computing Machinery, New York, NY, USA, 656–671. https://doi.org/10.1145/2901790.2901861

Munk, A. K., Olesen, A. G., & Jacomy, M. (2022). "The Thick Machine: Anthropological AI between explanation and explication." *Big Data & Society, 9*(1). https://doi.org/10.1177/20539517211069891

Nissenbaum, Helen Fay. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford Law Books. https://archive.org/details/privacyincontext0000niss

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

Rettberg, J. W. (2022). *Algorithmic failure as a humanities methodology: Machine learning's mispredictions identify rich cases for qualitative analysis. Big Data & Society, 9*(2). https://doi.org/10.1177/20539517221131290

UN Women. (2022). *Accelerating efforts to tackle online and technology-facilitated violence against women and girls.* https://www.unwomen.org/en/digital-library/publications/2022/10/accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls.

Pew Research Center. (2021, January 13). *The state of online harassment*. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/.3

Snodgrass, J. (2013). [Review of the book *Ethnography and virtual worlds: A handbook of method*, by T. Boellstorff, B. Nardi, C. Pearce, & T. L. Taylor]. American Anthropologist, 115, 000–000. https://doi.org/10.1111/aman.12038_2.

Tacchi, J., Foth, M., & Hearn, G. (2009, March 31). "Action research practices and media for development. *International Journal of Education and Development Using Information and Communication Technology, 5* (2). http://ijedict.dec.uwi.edu/viewarticle.php?id=560.

Case Study 7

SAFETY

AI-Enabled Safeguarding from Digital Harm

Authors: Mohit Kumar, Taniya Agrawal, India

Overview

Background & Problem Statement

Women and girls in India face escalating technology-facilitated gender-based violence – online harms that include non-consensual intimate imagery (NCII), sextortion, phishing, impersonation and social media-enabled harassment. Fraudulent online schemes disproportionately target women seeking to supplement household income, particularly those outside the formal workforce. Studies and law enforcement advisories indicate that women constitute a significant proportion of victims of online financial fraud in India. Emerging research and media investigations also show that AI-generated deepfake abuse is becoming an increasingly common vector of gender-based online extortion globally. When such scams result in financial loss, women may face secondary harms, including family conflict or domestic violence. In the Global South, limited awareness, fragmented reporting mechanisms and delayed response systems compound these risks, rendering non-AI approaches insufficient.

Solution Summary

Digital Safe India, an initiative of Arya Shield Digital Foundation, employs a hybrid model that combines AI-supported workflows with expert human review to address NCII, phishing, impersonation and gender-based online abuse. Automated processes support rapid triage, evidence preservation and structured takedown reporting, while human oversight ensures accuracy, contextual sensitivity and survivor safety. This approach improves response speed, reliability and access to justice for women affected by online harm across India.

The AI Solution

Solution Description

The Digital Safe India platform follows a hybrid operational model that integrates manual investigation with AI-supported workflows. Most NCII identification and verification is conducted through human review to ensure survivor safety and contextual accuracy, while automated tools support structured takedown

reporting to online platforms. Gender-responsiveness is embedded through survivor-informed case handling, prioritization of harms that disproportionately affect women and sustained human oversight to reduce misclassification risks. Where requested, NCII cases are handled by trained female volunteers in order to respect survivor comfort and cultural sensitivities. The solution maintains gender-disaggregated tracking of reported cases and incorporates feedback from women users to refine response workflows.

Deployment & Implementation

The solution operates primarily in India, supporting women and girls across multiple states through online reporting and response channels. Limited cross-border support has been provided in exceptional cases involving women from the Indian diaspora, including an NCII incident reported from Canada. Development began in 2025, followed by phased deployment combining manual review with AI-supported workflows and automated reporting. To date, the initiative has supported the removal of more than 1,000 NCII items, disrupted over 2,000 malicious websites and facilitated the identification of 560 mobile malware samples.

1,000

women experiencing online abuse, including NCII threats and digital blackmail, have been directly or indirectly supported by Digital Safe India

Arya Shield Digital Foundation develops and operates the system, which requires basic internet connectivity, secure web interfaces and external threat intelligence inputs. The initiative is currently self-funded by the founder.

Users & Scale

Primary beneficiaries are women and girls affected by NCII, online fraud and digital harassment, with women representing approximately 99 per cent of NCII-related cases supported. Secondary users include NGO responders and cybersecurity volunteers. Women access the solution through online reporting channels. For non-NCII cybercrimes, AI-based detection is used to identify potentially malicious domains and applications, which are disrupted only after human verification.

Impact & Outcomes

Measured Gender Impact

Digital Safe India has directly or indirectly supported more than 1,000 women experiencing online abuse, including NCII threats and digital blackmail. AI-assisted

triage reduced response times from days to hours, improving survivor safety and evidence preservation. Survivors reported increased confidence in reporting and reduced psychological distress due to faster intervention. The initiative has contributed to safer online participation and increased awareness of digital rights among women. Compared with pre-intervention manual reporting, the solution demonstrates improved efficiency and scalability, aligning with SDG 5 targets on eliminating violence against women and girls, including technology-facilitated abuse. Feedback from survivors indicates that timely takedown support helped reduce fear of exposure and provided emotional relief during NCII incidents.

Ethics, Governance & Safeguards

User privacy is protected through data minimization, secure storage and informed consent processes. Personally identifiable information is anonymized wherever possible. Human oversight ensures accountability for AI-supported outputs. Gender-sensitive handling protocols, including trauma-informed workflows and escalation mechanisms, reduce risks of misclassification or retraumatization. Women participate through reporting and feedback mechanisms that inform ongoing system improvements. The solution aligns with human rights principles and responsible AI practices, without relying on opaque automated decision-making.

Lessons & Replicability

What Worked

Key success factors included survivor-centered design, AI-assisted prioritization, partnerships with cybercrime stakeholders and public awareness efforts. Combining automation with human judgment proved critical for effectively addressing sensitive forms of gender-based harm.

Challenges Encountered

Challenges included underreporting due to stigma, linguistic diversity, evolving attacker tactics and coordination delays with online platforms. Limited digital literacy and trust barriers initially affected adoption among vulnerable users.

Key Lessons for Replication

Replication requires strong local partnerships, ethical AI safeguards and sustained survivor trust-building. Solutions must adapt to cultural, legal and linguistic contexts rather than rely on copy-paste deployment. Capacity-building is essential for sustainable impact.

Conclusion & Future Directions

Digital Safe India demonstrates how AI can advance gender equality by protecting women from online harm at scale. The solution shows potential for expansion across regions and for integration with national cybercrime response frameworks, contributing to a more inclusive and safe digital ecosystem.

References

Kumar, A. (2025). *The silent exploitation: How Instagram and Telegram are becoming hunting grounds for women's private images. Medium*. https://medium.com/@bugtest252/the-silent-exploitation-how-instagram-and-telegram-are-becoming-hunting-grounds-for-womens-8e567a8e9710.

National Crime Records Bureau. (2025). India's crime surge driven by cyber offences, violence against women and children. https://www.indiatracker.in/story/indias-crime-surge-driven-by-cyber-offences-violence-against-women-and-children.

Rudra, A. (2023). *A socio-legal study on cyber crimes against women in India. IOSR Journal of Humanities and Social Science, 28*(7), 26–30. https://www.iosrjournals.org/iosr-jhss/papers/Vol.28-Issue7/Ser-7/D2807072630.pdf.

Case Study 8

AGRICULTURE

FarmerChat: AI Advisory Services for Women in Agriculture

Authors: Namita Singh, Jona Repishti, Tetyana Zelenska; India, Kenya, Ethiopia, Nigeria, Brazil

Overview

Background & Problem Statement

Women make up approximately 43 per cent of the agricultural workforce in the Global South, yet they are consistently less likely than men to access agricultural advice. Even when they do, advice often fails to reflect factors such as their crops, constraints or decision-making roles. Structural barriers, including limited mobility, lower phone ownership, limited literacy in accessible languages and formats, and restrictive social norms, contribute to a gender productivity gap of 11–45 per cent, Agricultural extension systems are essential for sharing knowledge and improving practices. However, traditional models struggle to operate in a gender-inclusive manner and, at scale, to provide personalized and timely advice. In some countries, extension worker–to–farmer ratios exceed 1:5,000.

FarmerChat, an AI advisory service, aims to address these gaps by providing instant and localized agricultural advice via voice, text and images. Using natural language processing, FarmerChat enables women farmers to access advisory services in their own languages and in formats suited to their literacy levels, thereby reducing gender gaps in information access and decision-making.

Solution Description

FarmerChat operates as a mobile application that allows farmers to ask questions in multiple formats and languages. User queries are routed through a central system that translates local languages, interprets intent and retrieves relevant answers. When local, vetted content exists, the system uses retrieval-augmented generation, ensuring responses are grounded in trusted, location-specific information rather than generic AI outputs. The system also integrates weather forecasts and market prices. Output is provided in local languages and highlights the need to consult human experts in critical scenarios, supporting explainability and human decision-making. Gender-responsiveness is embedded through co-design with women farmers, bias mitigation in training data and continuous monitoring of gender-disaggregated usage.

Pilot Deployment and Field Validation

FarmerChat is currently at a mature, scaled stage, deployed across Brazil, Ethiopia, India, Kenya and Nigeria. Development began in May 2023, with pilots launched in October 2023 through Ministry of Agriculture extension workers in India and Kenya, before expanding access directly to farmers via a mobile application. The solution was developed and is operated by Digital Green and deployed in partnership with governments, integrating into existing extension systems by complementing extension workers. Infrastructure requirements are intentionally lightweight, relying on basic smartphones and low-bandwidth connectivity, with zero-rated data partnerships in some countries to reduce cost barriers for women.

Users & Scale

By December 2025, FarmerChat had reached approximately 830,000 users across six countries through partnerships with government agricultural and rural livelihood agencies in Ethiopia, India and Kenya. At least 35 per cent of users are women, who access FarmerChat using their own smartphones or through women's groups, frontline workers and shared household devices.

Impact & Outcomes

Measured Gender Impact

Across three studies, FarmerChat shows consistent positive outcomes for women farmers. In a 60 Decibels study in India, women reported a higher Net Promoter Score than men (87 versus 66), with 82 per cent applying information on their farms, indicating increased decision-making. In Kenya, women were more likely to be meaningful users (72 per cent versus 58 per cent for men) and to report that FarmerChat met all their farming needs (51 per cent versus 37 per cent), reflecting improved access to relevant advice. In Ethiopia, FarmerChat reached 40 per cent of women through Development Agents, strengthening locally relevant advisory support. Together, these outcomes align with SDG 5 by expanding women's access to digital advisory services.

Ethics, Governance & Safeguards

Ethical safeguards are embedded by design. FarmerChat follows principles of informed consent, minimal data collection and restricted access to sensitive user information. Gender-specific risks are addressed through participatory design, bias monitoring and reinforcement learning from human feedback, including input from women users and agricultural experts. The system includes feedback mechanisms to support accountability. Accountability is shared among

implementing organizations and partners, with mechanisms to flag harmful or inaccurate responses through user feedback. Women participate as co-designers, testers and feedback providers, influencing both product features and governance decisions. The system aligns broadly with human rights and data protection principles.

Lessons & Replicability

What Worked

FarmerChat's effectiveness stems from intentional gender-inclusive design, voice-first and multilingual interfaces, zero-rated data access and onboarding through trusted women's networks. Strong partnerships with governments, researchers and mobile operators enabled credibility and scale. Co-design with women farmers, continuous feedback loops and ongoing gender-disaggregated monitoring ensured that women's needs shaped product evolution and scalability across contexts.

Challenges Encountered

Designing an intentionally gender-inclusive AI system involved social barriers such as restrictive norms around women's phone use, digital literacy gaps and limited data affordability. Infrastructural and technical challenges included limited connectivity, low-end devices, voice recognition limitations and the complexity of a voice-first interface. Ensuring that advice aligns with women's time, mobility and financial constraints remains an ongoing challenge.

Key Lessons for Replication

Key prerequisites include basic connectivity, trusted local partners and an understanding of local contexts. Embedding Gender-responsiveness and user research throughout deployment and monitoring can support adaptation to local needs and limitations, language requirements and delivery models. Governments should integrate AI tools within existing extension systems to strengthen accountability and trust.

Conclusion & Future Directions

FarmerChat demonstrates that intentionally gender-inclusive AI systems can advance gender equality in agriculture. Inclusion and innovation can reinforce each other. By reducing costs to reach farmers and enabling effective scaling, the model provides a framework for expansion with local adaptations, contributing to a growing ecosystem of gender-inclusive AI for development.

References

Food and Agriculture Organization of the United Nations. (2011). The *role of women in agriculture* (ESA Working Paper No. 11-02). http://www.fao.org/4/am307e/am307e00.pdf.

Ragasa, C., Berhane, G., Tadesse, F., & Taffesse, A. S. (2012). *Gender differences in access to extension services and agricultural productivity* (ESSP Working Paper No. 49). International Food Policy Research Institute. https://cgspace.cgiar.org/items/45f6d3ec-d8b2-4dd3-a538-d70d1dd2c621.

Deka, M. B., Saikia, P., & Saikia, R. M. (2018). An analysis of gender gap in access to extension services. *Indian Journal of Extension Education*, 54(3), 138-141. https://epubs.icar.org.in/index.php/IJEE/article/view/143988.

Singh, B., Kaur, K., & Tiwari, D. (2025). Agricultural extension as a tool for empowering rural women. *International Journal of Agriculture Extension and Social Development*, 8(7), 499-504. https://www.extensionjournal.com/article/view/2168/8-7-58.

Mahajan, K. (2019). Back to the plough: Women managers and farm productivity in India. *World Development*, 124, Article 104633. https://doi.org/10.1016/j.worlddev.2019.104633.

Mugisha, J., Sebatta, C., Mausch, K., Ahikiriza, E., Okello, D. K., & Njuguna, E. M. (2019). Bridging the gap: Decomposing sources of gender yield gaps in Uganda groundnut production. *Gender, Technology and Development*, 23(1), 19-35. https://doi.org/10.1080/09718524.2019.1621597.

Anteneh, A., & Beyene, A. M. (2024). ICT-based agricultural extension and advisory service in Ethiopia: A review. *Cogent Food & Agriculture*, 10(1), Article 2391121. https://doi.org/10.1080/23311932.2024.2391121.

Sumanth, Nalina, Sanjay, Pradnya, Kaveri, Kalyan Roy, Moitreyee, Sai Sree, & Saravanan Raj. (2020). *Agricultural extension and support systems in India: An agricultural innovation systems (AIS) perspective* (Discussion Paper No. 20). National Institute of Agricultural Extension Management (MANAGE). https://www.manage.gov.in/publications/discussion%20papers/MANAGE_Discussion%20Paper%2020.pdf.

Digital Green & 60 Decibels. (2025, June). *Digital Green FarmerChat: User experience report (Bihar)*.

Digital Green & 60 Decibels. (2025, September). *Digital Green Farmer.Chat study (Kenya)*. https://60decibels.com/insights/farmers-ai-recommendations/

Digital Green & Fund for Innovation in Development (FID). (2025, October). Leveraging the transformational potential of AI to empower small-scale farmers in Ethiopia: Pilot study: Testing FarmerChat in Ethiopia.

Case Study 9

SAFETY

From Secure Reporting to Effective Remedies

Author: Tina Pirnovar, Slovenia

Overview

Background & Problem Statement

Sexual harassment of young women in the Global South remains persistently under-criminalized and under-prosecuted, reflecting deep gender inequalities, weak evidentiary standards and limited institutional capacity (Grove, 2015; World Justice Project, 2022; Red Dot Foundation, 2018). Survivors face stigma, unsafe reporting channels and the rapid loss of digital evidence, making access to justice difficult (Fahmy et al., 2014; HarassMap, n.d.; Red Dot Foundation, 2023). Existing systems have lacked secure intake mechanisms, timely verification processes and tools to detect coercive, workplace and online abuse (Karlekar & Bansal, 2018; HarassMap, n.d.).

Non-AI solutions, including manual review, paper-based reporting and traditional investigations, have been too slow, fragmented and resource-constrained to address the scale and complexity of digital harassment (World Justice Project, 2022; Flinkman, 2019). AI-enabled reporting, evidence preservation and structured prosecutorial workflows address these gaps by strengthening investigations and improving survivor-centered outcomes (Red Dot Foundation, 2023; Karlekar & Bansal, 2018).

Solution Summary

The AI-driven solution enhances access to justice for young women by enabling secure reporting, natural language processing-based detection of harassment, deepfake verification and automated case triage (Karlekar & Bansal, 2018; Red Dot Foundation, 2023). Drawing on natural language processing, machine learning and forensic analytics, the system strengthens evidence collection and prosecutorial decision-making (World Justice Project, 2022; HarassMap, n.d.). The primary gender outcome is improved access to justice and survivor-centered remedies (Red Dot Foundation, 2018; Fahmy et al., 2014).

The AI Solution

Solution Description

This AI system uses natural language processing to detect harassment in chat logs

and social media content, computer vision techniques to identify deepfakes, and machine learning models to triage cases and guide trauma-informed interviews (Karlekar & Bansal, 2018; HarassMap, n.d.; SasaTang, n.d.). It enables secure, anonymous reporting and structured evidence intake, preserving digital traces before they are lost (Red Dot Foundation, 2018; Red Dot Foundation, 2023).

Gender-responsiveness is embedded through survivor co-design, the use of gender-disaggregated data and safeguards to mitigate bias (Fahmy et al., 2014; Red Dot Foundation, 2018). Outputs are generated through transparent algorithms with human-in-the-loop oversight to validate decisions (World Justice Project, 2022; Karlekar & Bansal, 2018). Explainability is supported through audit trails, chain-of-custody protocols and clear criteria for case prioritization, enabling ethical and accountable use in prosecutorial contexts (World Justice Project, 2022; Red Dot Foundation, 2023; SasaTang, n.d.).

Deployment and Implementation

The AI-assisted criminal law tools are being piloted within specialized prosecution units in selected Global South cities (Red Dot Foundation, 2018; World Justice Project, 2022). The geographic scope draws on pilots inspired by Safecity in India and HarassMap in Egypt, integrating community-based reporting with prosecutorial workflows (Red Dot Foundation, 2023; HarassMap, n.d.; SasaTang, n.d.).

Implementation progressed from doctrinal design to prototype development, field testing and evaluation (Karlekar & Bansal, 2018; Fahmy et al., 2014; HarassMap, 2014). Deployment is supported through partnerships between justice ministries, women's rights organizations and local technology developers, with national institutions providing oversight, data access and funding (World Justice Project, 2022; Red Dot Foundation, 2018). The system operates through secure mobile and web interfaces, integrates with police intake and digital evidence systems and requires stable connectivity, privacy-by-design infrastructure and human-in-the-loop validation (Red Dot Foundation, 2023; Karlekar & Bansal, 2018).

Users & Scale

Primary users include young women reporting harassment, alongside prosecutors, investigators and specialized units (Red Dot Foundation, 2018; HarassMap, n.d.). Pilot deployments have reached several hundred survivors and two justice institutions drawing on community platforms such as Safecity and HarassMap (Red Dot Foundation, 2023; Karlekar & Bansal, 2018). Women access the system through secure mobile and web-based reporting tools, community facilitators and integrated government intake services (Red Dot Foundation, 2018; World Justice Project, 2022).

Impact & Outcomes

Measured Gender Impact

Pilot deployments reached approximately 300–800 young women. Reporting through secure and anonymous channels increased by approximately 50 per cent compared with baseline levels, while case attrition declined by about one third, reflecting improved evidence preservation and prosecutorial follow-through (Red Dot Foundation, 2023; World Justice Project, 2022). Automated triage reduced intake and prioritization time by approximately 40 per cent, accelerating referrals to support services (Karlekar & Bansal, 2018; Red Dot Foundation, 2023). Survivors reported higher satisfaction and perceived safety, and prosecutions in pilot units progressed at a notably higher rate than in comparable units. Economic impacts were indirect, including reduced time lost to repeated reporting (Fahmy et al., 2014; World Justice Project, 2022). These outcomes align with SDG 5 by improving access to justice and reducing barriers to participation (Red Dot Foundation, 2018; World Justice Project, 2022).

Ethics, Governance & Safeguards

Ethical safeguards are embedded throughout the system life cycle. Data protection measures include consent-based, opt-in reporting, AES-256 encryption at rest, TLS encryption in transit, strict role-based access controls, automatic pseudonymization, retention limits (e.g., 90 days for non-evidentiary data), and auditable chain-of-custody procedures for evidentiary materials (Red Dot Foundation, 2018; Red Dot Foundation, 2023; World Justice Project, 2022).

Bias mitigation includes the use of gender-disaggregated datasets, balanced sampling, fairness metrics, adversarial testing and continuous monitoring with corrective retraining (Karlekar & Bansal, 2018; Red Dot Foundation, 2023). Additional safeguards include privacy-by-design principles, minimal data collection, trauma-informed user interfaces, mandatory human-in-the-loop review for automated flags and model documentation outlining limitations (Red Dot Foundation, 2018; Fahmy et al., 2014).

Governance mechanisms include a multi-stakeholder oversight board comprising women's organizations, prosecutors and technologists, independent external audits, public transparency reporting and a survivor-accessible grievance and remediation process (Red Dot Foundation, 2023; World Justice Project, 2022). Risk management measures include incident response protocols, opt-out mechanisms and periodic impact assessments to mitigate risks of surveillance, retraumatization and wrongful prioritization (Red Dot Foundation, 2023; HarassMap, n.d.).

Lessons & Replicability

What Worked

Survivor co-design, legal reforms addressing digital and coercive offences and the establishment of specialized prosecution units, combined with AI-supported evidence preservation and automated triage, proved most effective (Red Dot Foundation, 2018; World Justice Project, 2022). Key enablers included partnerships with women's organizations, law enforcement and prosecutors, mobile reporting tools, funding for training and sustained community outreach. Localization, offline functionality, cultural translation and continuous monitoring supported adaptation across contexts (Red Dot Foundation, 2023; HarassMap, n.d.).

Challenges Encountered

Technical challenges included multilingual and informal text, limited labelled data affecting natural language processing accuracy and low-resolution media constraining forensic analysis. Connectivity limitations, shared device use and low smartphone access restricted reporting. Institutional constraints, including outdated legal frameworks and weak case management systems, alongside stigma, distrust and low digital literacy, further impeded uptake. Funding shortfalls threatened sustainability and increased legal uncertainty (Red Dot Foundation, 2023; Flinkman, 2019).

Key Lessons for Replication

Successful replication requires multilingual and culturally adapted systems capable of handling informal and code-switched text, alongside investments to address data scarcity. Reliable connectivity, secure access and institutional reform are critical enablers. Sustained funding, legal clarity and strong partnerships with women's organizations are essential to ensure scalability, trust and long-term impact (Red Dot Foundation, 2018; World Justice Project, 2022).

Conclusion & Future Directions

AI-assisted reporting and forensic tools can advance gender equality by strengthening young women's access to justice, preserving digital evidence and improving prosecutorial outcomes. The model shows potential for adaptation across Global South contexts and for addressing related gendered harms (Red Dot Foundation, 2023; World Justice Project, 2022).

Next steps include expanded pilots, formal legal adoption, capacity-building initiatives and independent audits. The approach advances gender-inclusive AI through survivor co-design and accountable, privacy-first practices (Red Dot Foundation, 2023; World Justice Project, 2022).

References

Fahmy, A., Abdelmonem, A., Hamdy, E., & Badr, A. (2014). *Towards a safer city: Sexual harassment in Greater Cairo — The effectiveness of crowdsourced data* (Research report). HarassMap, Youth and Development Consultancy Institute. https://www.researchgate.net/profile/Amel-Fahmy/publication/273440790_Toward_A_Safer_City/links/6106b1f5169a1a0103cd2ddf/Toward-A-Safer-City.pdf

Flinkman, N. (Ed.). (2019). Mapping sexual harassment in Egypt. In *This is not an atlas: A global collection of counter-cartographies*. https://www.notanatlas.org/wp-content/uploads/2019/06/Mapping-Sexual-Harassment-in-Egypt.pdf

Grove, N. S. (2015). The cartographic ambiguities of HarassMap: Crowdmapping security and sexual violence in Egypt. *Security Dialogue, 46*(4), 345–364. https://doi.org/10.1177/0967010615583039

HarassMap. (2014). *Towards a safer city: Sexual harassment in Greater Cairo: Executive summary*. https://harassmap.org/storage/app/media/uploaded-files/Towards-A-Safer-City_executive-summary_EN.pdf

HarassMap. (n.d.). *Studies and reports*. https://harassmap.org/en/studies-and-reports.

Karlekar, S., & Bansal, M. (2018). *SafeCity: Understanding diverse forms of sexual harassment* [Dataset]. GitHub. https://github.com/swkarlekar/safecity/blob/master/README.md

Red Dot Foundation. (2018). *Annual report 2017–2018 (RDF annual report)*. https://reddotfoundation.in/uploads/2019/01/RDF-Annual-Report-2018.pdf

Red Dot Foundation. (2023). *Safecity data dashboard: Outside India (Introduction)*. https://reddotfoundation.in/uploads/2023/11/Safecity_Data_Dashboard-Outside-India.pdf

SasaTang. (n.d.). *SafeCityIndia*. GitHub. https://github.com/SasaTang/SafeCityIndia

World Justice Project. (2022). *Red Dot Foundation; The Urban Vision — Safecity: a crowdmap for sexual and gender-based violence*. https://worldjusticeproject.org/world-justice-challenge-2022/red-dot-foundation-urban-vision

Case Study 10

HEALTH

Project Saathi: Scalable AI for Gender-Responsive Health

Authors: Madhavi Kaivalya Kandalam, Sailaja Raman, India

Overview

Background & Problem Statement

Women in the Global South face systemic barriers to accessing healthcare. Their needs are often deprioritized due to unpaid care responsibilities, limited service availability, requirements for family consent and financial constraints, leading to delayed care. More than 1.2 billion women experience at least one micronutrient deficiency. Existing health-focused AI systems reflect male-dominated medical research and gender bias in underlying datasets. For example, AI models are significantly more likely to miss liver disease in women than in men, contributing to underdiagnosis. In low- and middle-income countries, these gaps are compounded by language barriers and the digital divide. Non-AI solutions are unable to provide continuous, scalable symptom triage at home, with travel remaining a limiting factor.

Solution Summary

Project Saathi is a vernacular, voice-based AI system designed to support women in low- and middle-income countries in interpreting symptoms, taking preliminary action and accessing appropriate care. The initiative focuses on creating women-centric health datasets to mitigate algorithmic bias and provides end-user training to bridge the digital divide. It is designed to scale to additional domains, including financial literacy.

The AI Solution

Solution Description

Project Saathi offers a health-first, vernacular, voice-based AI framework built around women's real-world constraints, drawing on Indic language models. Central to the approach is the ethical creation of women-centric health datasets. This gender-disaggregated architecture enables the system to learn women's symptom patterns, incorporating vernacular cues and physiological markers, such as atypical signs of micronutrient deficiencies.

Outputs are designed to be simple and actionable, delivered through guided conversational interactions that explain underlying reasoning rather than providing automated diagnoses. End users receive direct training on the intuitive user interface, supporting digital inclusion among rural women. The framework is designed to be extensible to other domains, such as financial literacy, supporting a scalable and implementation-ready approach.

Deployment & Implementation

WizLearnr is developing and piloting Project Saathi, with inputs from domain experts and female community leaders through early pilots and community validation activities using a train-the-trainer model. Initial pilots are underway in select districts of Maharashtra and Telangana, engaging women from low-income and informal work settings. State-wide pilots are planned for 2026, with pan-India deployment targeted for 2027. The solution is a lightweight conversational application designed for basic smartphones with low-bandwidth requirements, with potential integration into the Ayushman Bharat Digital Mission to support continuity of care. Saathi is designed to complement, rather than replace, existing healthcare workflows.

Users & Scale

Primary beneficiaries are adult women in low-income, informal or rural settings in low- and middle-income countries. Initial pilots target 500 households, with women trained to use the native-language mobile application through WhatsApp, supported by voice and text interactions from home. Local women entrepreneurs (‘Saathis’) act as training intermediaries, helping to overcome trust and technology barriers.

Impact & Outcomes

Measured Gender Impact

Drawing on evidence from comparable pilots in India, Project Saathi aims to increase early care-seeking among women by 30 per cent. Interviews highlighted patterns of delayed diagnosis, with conditions such as gallbladder disease and appendicitis frequently misinterpreted as routine stomach pain. Evidence showing more than 80 per cent agreement between reproductive health symptom checkers and general practitioner assessments indicates strong potential to reduce diagnostic delays through early, bias-aware AI-supported triage. Saathi enables at-home triage, saving an estimated two to four hours per visit and reducing travel-related costs. By providing AI-supported guidance before women engage families or health providers, the system strengthens women’s health-related agency and aligns with SDG targets 3.8 and 5.b.

Screenshots of chats in English, Hindi and Malayalam

Ethics, Governance & Safeguards

Ethical considerations are central to Saathi's design and implementation. Safeguards include a privacy-first architecture with secure, anonymized data handling and access controls; explicit consent protocols that clearly explain data purpose and use; and bias mitigation measures that address gender data gaps through dataset segregation during pre-processing and intentional enrichment of training data with female-specific studies and markers. Inclusive access is supported through native-language and voice-based interfaces on low-resource devices. A clear red-line protocol governs high-risk scenarios, while feedback loops with women users and continuous output review support human-in-the-loop governance. Oversight is provided by a female-led steering committee to manage unintended consequences, such as technology-induced anxiety, and guide responsible scaling.

Lessons & Replicability

What Worked

Early validation and secondary research indicate that vernacular-first, voice-enabled and home-based access is critical for inclusion. Women commonly reported relying on informal advice from peers before seeking formal care. The train-the-trainer model, leveraging existing community trust, proved effective in overcoming initial technology scepticism and supporting scale.

Challenges Encountered

Key challenges included wide variation in symptom descriptions, limited availability of women-specific health data for early modelling and high dialectal variation affecting multilingual speech-to-text performance. Shared device use, linked to the digital divide, complicated privacy considerations. Women also expressed anxiety about digital tools that emphasize worst-case diagnoses, underscoring the need for carefully calibrated AI outputs.

Key Lessons for Replication

Effective AI-supported triage must balance awareness with anxiety reduction. Voice-based, vernacular approaches are effective in low-literacy settings, but the digital divide must be actively addressed. Investment in women-centric datasets and training is essential to reach the most marginalized groups. Successful replication depends on contextual adaptation and sustained community engagement to build trust.

Conclusion

Project Saathi positions AI as a gender equalizer rather than a bias amplifier, supporting women's health despite persistent access, data and digital barriers. The initiative presents an implementation-ready, scalable and gender-responsive model that can extend to additional domains, such as financial literacy. By recognizing women's interconnected responsibilities, the framework contributes to more inclusive digital ecosystems across the Global South.

References

Bećirović, A. (2025). The intersection of gender data gaps and AI [Conference paper]. World Bank. https://thedocs.worldbank.org/en/doc/9fe224595c1ecfe9e508d7f3be205a8d-0080012025/related/S2-8-P-Azra-Becirovic-paper.pdf

GSMA. (2025). The mobile gender gap report 2025 (Version 1.0). GSM Association. https://www.gsma.com/r/wp-content/uploads/2025/06/The-Mobile-Gender-Gap-Report-2025.pdf

Microsoft Research. (2024). ASHABot: Empowering Frontline Health Workers through Generative AI. https://www.microsoft.com/en-us/research/publication/ashabot-an-llm-powered-chatbot-to-support-the-informational-needs-of-community-health-workers/

Murthy, N., Chandrasekharan, S., Perumal Prakash, M., Kaonga, N. N., Peter, J., & others. (2019). The impact of an mHealth voice message service (mMitra) on infant care knowledge and practices among low-income women in India. Maternal and Child Health Journal, 23(12), 1658–1669. https://link.springer.com/article/10.1007/s10995-019-02805-5

Peven, K., Wickham, A. P., Croft, J., Zhaunova, L., Peven, K., Bamford, R., Radovic, T., & Gilbert, S. (2023). Assessment of a digital symptom checker tool''s accuracy in triaging and identifying reproductive health conditions. JMIR mHealth and uHealth, 11, e46718. https://doi.org/10.2196/46718

Seyyed-Kalantari, L., Liu, G., McDermott, M., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs. Nature Medicine, 27(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0

Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health & Care Informatics. https://informatics.bmj.com/content/29/1/e100457

UN Women. (2024). Artificial intelligence and women's rights in the Asia-Pacific: Promises and perils (AP/C871). UN Women Regional Office for Asia and the Pacific. https://asiapacific.unwomen.org/sites/default/files/2024-05/ap-c871-ai-research-report-2024-full.pdf

World Health Organization. (2023). Accelerating efforts for preventing micronutrient deficiencies and their consequences, including spina bifida and other neural tube defects, through safe and effective food fortification (Executive Board document EB152(13)). World Health Organization. https://apps.who.int/gb/ebwha/pdf_files/EB152/B152(13)-en.pdf

Case Study 11

SKILLING

Sanmati: Participatory Bias-mitigation for Indic Language Models

Authors: Safiya Husain, Vivek Seshadri, Kavya Tadakaluri; India

Overview

Background & Problem Statement

Language is one of the most powerful ways in which people interface with technology, shaping who is seen and what is considered normative. As AI systems increasingly mediate everyday interactions, gender-bias risks are becoming both invisible and systemic. For many women, technologies reflect inequality and reproduce it by encoding stereotypes and exclusions into tools that operate at scale.

These dynamics are particularly pronounced in multilingual contexts such as India. Most large language models are trained on English-dominant text datasets and evaluated using Western linguistic benchmarks. As a result, AI systems can produce socially misaligned outputs that reinforce gendered assumptions (Hada et al., 2024). Prior approaches to addressing bias have remained English-centric, relying on translation or English-first taxonomies that fail to capture language-specific dynamics (Blodgett et al., 2020; Malik et al., 2022).

It was within this context that Project Sanmati and an Indian-language gender-bias benchmarking tool were conceived. Funded by the Gates Foundation and implemented by Karya, Sanmati responds to women's exclusion from the AI value chain. By engaging women as contributors, annotators and definers of bias, the initiative shifts how AI systems are built.

Solution Summary

Sanmati developed a benchmarking-oriented data and evaluation pipeline that centres women as co-creators in order to surface gender bias across six Indian languages (Bengali, Bhojpuri, Hindi, Malayalam, Marathi and Telugu). The intervention prioritizes representation and establishes foundations for bias-aware evaluation at the dataset level.

Solution Description

Sanmati uses a sequence of participatory language tasks that enable women to generate and evaluate sentences reflecting everyday experiences of gender

bias. These sentences are manually coded using a multidimensional framework that captures the presence and direction of bias, its explicitness, relevant linguistic markers and associated contextual harms.

20,000

women across six Indic languages and eight Indian states were engaged by Sanmati

The system produces a structured 'golden dataset' that will be used to benchmark language models and inform bias-aware dataset cleaning. Women's lived experiences inform task design, bias definitions and coding logic, ensuring that outputs remain culturally grounded and interpretable.

Deployment & Implementation

The project spans three phases: learning, data collection and benchmark tool creation (currently under development). During the data collection phase, Sanmati engaged 20,000 women across the six Indic languages and eight Indian states. All work was conducted through Karya's smartphone-based application, with participants contributing approximately four to five hours in total and receiving INR 2,000 for their participation.

Implementation was managed entirely by Karya. Women participants were onboarded through NGO partnerships and women's self-help groups, which proved particularly catalytic. State government partnerships in Kerala, Maharashtra and Telangana supported outreach and trust-building. Participation required women to have access to their own Android smartphones and internet connectivity. All outputs, including datasets, methodologies and the benchmarking tool, are designed to integrate into existing AI evaluation workflows and will be open-sourced.

Users & Scale

The primary users of Sanmati's outputs are technologists, AI developers and researchers seeking gender-intentional evaluation tools. The primary beneficiaries of the process are the women contributors whose linguistic knowledge and labour shape the datasets. Work was completed through an app and supported by remote training, enabling a scalable delivery model.

Measured Gender Impact

Sanmati engaged 20,000 women and generated more than 15,000 curated sentences. These contributions form the basis of a corpus used to develop the

final benchmarking tool through the creation of gender-bias-coded gold-standard sentences. The initiative created digital work opportunities and enabled women to contribute directly to AI systems that shape their linguistic communities. At this stage, the project prioritizes representation rather than claiming bias-reduction outcomes. The work aligns most directly with SDG 5 (Gender Equality) through participation and representation, and SDG 8 (Decent Work) through paid, dignified digital labour.

Ethics, Governance & Safeguards

Informed consent was obtained through the Karya application as part of a work-for-hire agreement. Data is stored on secure cloud servers, and personal identifiers were removed prior to analysis. In each language, 10–15 women, referred to as Karya Champions, participated in workshops to define local manifestations of gender bias, shaping the coding framework alongside literature review, as well as reviewing content and providing feedback. Unintended consequences included occasional household tensions, particularly where male family members were unfamiliar with phone-based work.

What Worked

Task sequencing was critical to eliciting authentic expressions of bias. Application-based delivery enabled scale without intensive in-person training, while partnerships with women's self-help groups built trust, collective learning and sustained participation.

Challenges

Not all collected sentences were suitable for coding, and achieving inter-coder reliability required iterative refinement. Linguistic nuance and schema complexity posed challenges, particularly across languages. Social dynamics, including household-level tensions, highlighted the need for responsive support systems alongside technical rigour.

Lessons for Replication

Gender-bias frameworks cannot be copied wholesale. Language-specific structures and cultural norms must inform task design and annotation logic. Successful replication requires trusted women's networks, fair compensation and investment in participatory design. For governments, this model offers a dual opportunity to advance gender-aware AI while creating dignified digital livelihoods.

Conclusion

Sanmati demonstrates that when women are meaningfully included in shaping AI systems, those systems can better reflect the societies they serve. By centring women's linguistic knowledge and lived experience, the initiative lays the groundwork for more gender-aware evaluation of language models in low-resource settings. This approach can be extended to additional languages and applied to domains such as agriculture and healthcare, contributing to a more inclusive and responsible AI ecosystem.

References

Ahuja, K., Diddee, H., Hada, R., Ochieng, M., Ramesh, K., Jain, P., & Sitaram, S. (2023, December). *Mega: Multilingual Evaluation of Generative AI*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 4232–4267)*. Association for Computational Linguistics

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). *Language (technology) is power: A critical survey of "bias" in natural language processing. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. https://aclanthology.org/2020.acl-main.485/

Hada, R., Husain, S., Gumma, V., Diddee, H., Yadavalli, A., Seth, A., & Bali, K. (2024, June). *Akal badi ya bias*: An Exploratory Study of Gender Bias in Hindi Language Technology. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1926–1939). Association for Computing Machinery.

Malik, M., Joshi, P., & Sharma, D. M. (2022). *Gender bias in Hindi language technologies. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 1149–1158). https://aclanthology.org/2022.naacl-main.76/

Case Study 12

SAFETY

NariRaksha: Gender-Responsive AI for Women's Safety

Authors: Vikhram S, Jeffin Gracewell J, India

Overview

Background & Problem Statement

Women in urban areas of the Global South continue to face significant safety challenges that restrict mobility, access to employment and education, and participation in public life. In Indian cities, underreporting of incidents, delayed institutional responses and fragmented safety systems limit the effectiveness of traditional interventions. Existing non-AI approaches rely largely on reactive mechanisms, such as helplines and manual surveillance, which are insufficient in dense and complex urban environments. The absence of integrated data analysis further constrains authorities' ability to identify high-risk locations and recurring patterns. These gaps create systemic barriers for women, underscoring the need for a proactive, scalable and gender-responsive approach that strengthens institutional capacity while improving women's confidence and safety in public spaces.

Solution Summary

NariRaksha is a deployed AI-based early warning and support system that combines geospatial analytics, computer vision and multilingual natural language processing to detect safety risks, support faster institutional responses and improve women's safety in urban environments.

The AI Solution

Solution Description

NariRaksha operates as a decision-support system for urban safety authorities while remaining accessible to women users. The system analyses location-based risk indicators, closed-circuit television feeds and user-generated inputs to identify emerging safety concerns. Machine-learning models detect anomalous behavioural patterns in public spaces, while natural language processing models interpret distress-related inputs across multiple local languages. Gender-responsiveness is embedded through the use of gender-disaggregated data, bias monitoring during model development and participatory co-design with women users and frontline officials. Outputs are presented through explainable dashboards, enabling transparent and accountable decision-making by non-technical users.

Deployment & Implementation

The solution was deployed between 2023 and 2025 across selected urban districts of Chennai and Bengaluru in India. Implementation involved collaboration between technology developers, municipal authorities and law enforcement agencies. The system integrates with existing surveillance infrastructure, command-and-control platforms and patrol-planning workflows without requiring major hardware upgrades. Standard urban connectivity, secure servers and dashboard access for officials are required. Local institutions play a central role in governance, monitoring and operational use of the system.

Users & Scale

Primary beneficiaries include women commuters and residents in urban districts; institutional users include city police units and urban safety administrators. The system supports decision-making across multiple districts, indirectly benefiting thousands of women through improved response mechanisms and preventive safety planning.

Impact & Outcomes

Measured Gender Impact

During deployment, the system processed more than two million surveillance frames and analysed 18,000 verified incident records. Comparisons with pre-deployment conditions showed reduced response delays, improved credibility of reported incidents and more targeted allocation of patrol resources. Authorities were able to identify high-risk micro-locations and implement preventive interventions, such as adjusted patrol schedules and infrastructure improvements. Women users reported increased confidence in reporting incidents due to visible institutional follow-up. Together, these outcomes contributed to safer public spaces, enhanced mobility and progress towards gender equality, aligned with Sustainable Development Goal 5.

Ethics, Governance & Safeguards

Ethical safeguards were embedded throughout the system life cycle. Personal data were anonymized, and only aggregated insights were used for decision-making. Consent mechanisms governed user-generated inputs, and strict access controls limited system use to authorized personnel. Bias mitigation checks were conducted during model training to reduce disproportionate impacts. Explainable outputs supported transparency, while institutional oversight committees and feedback from women representatives strengthened accountability and responsible deployment.

Lessons & Replicability

What Worked

Strong institutional partnerships, integration with existing safety infrastructure and participatory design with women users were key success factors. Explainable dashboards enabled adoption by non-technical officials and improved operational decision-making.

Challenges Encountered

Challenges included uneven data quality across locations, connectivity limitations in certain areas and initial trust barriers among users. Continuous engagement and capacity-building were required to address these constraints.

Key Lessons for Replication

Successful replication requires institutional ownership, reliable urban data infrastructure and early community engagement. Solutions must be adapted to local contexts rather than directly replicated. Ethical governance should be established from the outset.

Conclusion & Future Directions

NariRaksha demonstrates how ethically governed and gender-responsive AI can strengthen urban safety systems in the Global South. The solution shows strong potential for geographic expansion and adaptation to additional public service contexts, contributing to inclusive and accountable AI ecosystems.

References

Ministry of Electronics and Information Technology. (2024). *IndiaAI mission strategy and responsible artificial intelligence adoption*. Government of India.

Organisation for Economic Co-operation and Development. (2021). *Artificial intelligence, data governance and privacy*. OECD Publishing.

UN Women. (2023). *Gender equality and digital transformation in emerging economies*. United Nations Entity for Gender Equality and the Empowerment of Women.

World Bank. (2022). *Safe cities and gender-inclusive urban development*. World Bank Publications.

Case Study 13

EDUCATION

Sakhi: AI-Supported Peer Mentoring from Education to Work

Authors: Anjali G Sharma, Taniya Negi, Anisha Sharma, India

Overview

Background & Problem Statement

Girls from underserved backgrounds face critical drop-off points at key transitions – from school to college, college to first employment and the early years of work. In low-resource education and employment systems across the Global South, formal guidance at these stages is limited, forcing reliance on informal peer networks. When girls who successfully navigate these transitions move on, their lived experiences exit the system, creating a structural memory failure inherent to human-led mentoring, where success pathways cannot be retained or transmitted at scale. As participation grows, such knowledge becomes fragmented and unrecoverable through manual systems.

Maitri, a peer-led mentoring community supporting girls aged 15–23 from government and low-income schools in South Delhi, encountered this challenge as its reach expanded. Existing non-AI solutions were episodic, person-dependent and lacked continuity.

Solution Summary

Sakhi is a text-based, AI-supported peer mentor developed within the Maitri community and currently used in a controlled pilot. Using natural language processing, it surfaces relevant peer experiences to support reflection on education, work and personal growth, strengthening continuity of guidance and decision-making confidence for first-generation girls during key transitions.

The AI Solution

Solution Description

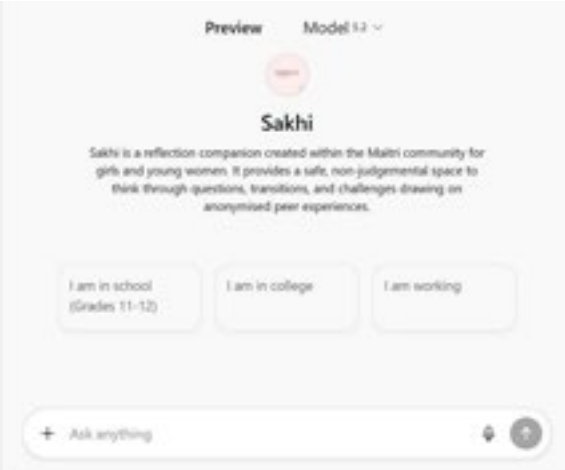
Sakhi preserves and retrieves collective learning across transition stages within the Maitri community. It performs semantic similarity matching across free-text reflections to identify overlaps in language, transition stage and concern. It then matches a girl's reflections or questions with anonymized peer experiences from others who have navigated similar educational, professional or personal transitions.

Manual curation became unworkable as narratives scaled, leading to recall bias; natural language processing enables retrieval at scale without human mediation. Sakhi functions one-on-one, emphasizes self-reflection over prescription and is gender-responsive by design, relying on consented peer narratives. Outputs remain explainable, as responses directly reference anonymized peer experiences. The system is text-based, uses simple language and is designed to be accessible across varying literacy levels, devices and learning needs.

Deployment & Implementation

Sakhi is deployed in a controlled pilot within the Maitri peer-mentoring community in Delhi. Conceptualization began in late 2025, followed by iterative testing alongside ongoing mentoring activities in order to ensure alignment with girls' language and concerns. The solution is developed, operated and maintained by the Maitri core team and accessed through an internal, text-based chat interface, with no public-facing release.

The pilot requires basic smartphone access and internet connectivity and operates as a standalone support layer, complementing rather than replacing existing peer- and facilitator-led interactions. While currently operated independently, the pilot is structured to generate learning relevant to potential adoption by NGOs or public education-to-work programmes.



Sakhi pilot interface. The screenshot illustrates a non-public, text-based reflection interface for users across school, college and early work transitions; no user data or personalized responses are displayed

Users & Scale

Primary users are girls and young women aged 15–23 from government and low-income schools participating in the Maitri community, many of them first-generation learners. The pilot includes approximately 150 contributors who also serve as active users, with additional cohorts across school, college and early work stages scheduled for near-term testing. Engagement occurs through one-on-one, text-based conversations focused on reflection.

Impact & Outcomes

Measured Gender Impact

As Sakhi is in a controlled pilot phase, impact evidence remains early -stage. The pilot includes approximately 150 girls and young women. Over half of active users returned to Sakhi more than once during periods of educational or career uncertainty, indicating demand for continuity rather than one-off advice.

Facilitators observed engagement clustering around high-risk transition moments, such as examination results, college entry and early job decisions. Users reported feeling safer articulating concerns due to the absence of judgement or prescriptive guidance, enabling clearer reflection on available options compared with episodic counselling or informal advice. A college-going participant noted, “It helped me understand my situation in a way that felt practical and reduced my fear”, suggesting strengthened decision-making agency aligned with SDG 5.

Ethics, Governance & Safeguards

Ethical safeguards are embedded by design. Sakhi is deliberately designed not to evaluate, rank or prescribe life choices; it functions solely as a reflective, non-therapeutic support mechanism. Peer experiences are anonymized and shared only with explicit, informed consent, and Sakhi does not store or process personal data from individual conversations. Responses are generated from consented peer narratives rather than automated scoring, supporting transparency and explainability. Access is restricted to Maitri participants through an internal interface. Governance is community-led, with facilitators overseeing use, addressing concerns and reviewing unintended harms where identified. This approach prioritizes privacy, accountability and trust, aligning with human rights-based and gender-responsive AI principles.

Lessons & Replicability

What Worked

Trust and relatability emerged as critical enablers. Engagement was driven by a

deliberate design choice to prioritize reflection over prescription, using simple, familiar language grounded in lived peer experiences. Keeping the interface text-based and avoiding ranking, recommendations or automated advice helped sustain emotional safety and trust. Embedding the system within an existing peer community and relying on consented narratives ensured relevance with minimal infrastructure requirements.

Challenges Encountered

A central challenge was resisting pressure to over-engineer the system or accelerate scale prematurely. Early testing showed that overly polished or directive responses risked undermining authenticity. Additional constraints included varying levels of digital literacy among users and the need for facilitator oversight to ensure responsible use, which limited the pace of iteration.

Key Lessons for Replication

This approach cannot be replicated through technology alone. Successful adaptation requires community governance, consented peer contributions and grounding in local language and context. For NGOs or public programmes, key prerequisites include an existing mentoring structure, facilitator capacity and clear consent protocols. Phased rollout and community stewardship should be prioritized over technical polish.

Conclusion & Future Directions

Sakhi demonstrates how AI can support gender-inclusive education and work transitions by preserving community knowledge rather than replacing human support. The pilot indicates a scalable model for NGOs operating within community-based mentoring contexts. Next steps include deepening consented peer memory and informing adoption by similar programmes. Sakhi offers a community-governed, non-prescriptive design pattern for gender-inclusive AI ecosystems.

References

UNESCO. (2022). *Global education monitoring report: Gender review*.

World Bank. (2018). *World development report: Learning to realize education's promise*.

Case Study 14

HEALTH

Thermalytix: Privacy-Preserving AI for Breast Cancer Screening

Authors: Janki Chitroda, Yashvi Chavda, Krishna Bhatt, India

Overview

Background & Problem Statement

Breast cancer is a leading cause of cancer mortality among women in the Global South, largely due to late-stage detection driven by socio-cultural barriers. In conservative societies, hesitation around disrobing and physical contact, often involving male technicians, limits access to mammography (which is also capital-intensive and unsuitable for rural deployment) (NIRAMAI Health Analytix, 2024). Earlier non-AI methods, such as manual thermography, suffered from subjective interpretation (NIRAMAI Health Analytix, 2024). In a clinical evaluation of 470 women, Thermalytix achieved 91.02 per cent sensitivity and 82.39 per cent specificity, with an AUC of 0.90, underscoring the need for a scalable, privacy-preserving screening solution (NIRAMAI Health Analytix, 2024).

Solution Summary

Niramai, an ISO-certified organization, utilizes Thermalytix, an automated diagnostic engine powered by machine learning and computer vision, to detect malignancies in thermal images (NIRAMAI Health Analytix, 2024). Holding 39 granted patents in thermal imaging and AI across multiple countries, the solution preserves privacy, eliminates physical touch, respects cultural modesty and increases screening uptake (NIRAMAI Health Analytix, 2024).

The AI Solution

Solution Description

Niramai's core solution, Thermalytix, integrates artificial intelligence with thermal imaging to support breast cancer detection (NIRAMAI Health Analytix, 2024). The system captures high-resolution heat maps of the chest area using a standard thermal camera (NIRAMAI Health Analytix, 2024). Advanced computer vision and machine-learning algorithms analyse these images to identify abnormal vascular patterns and temperature asymmetries indicative of malignancies (NIRAMAI Health Analytix, 2024).

The gender-responsive process is completely non-contact and privacy-preserving, requiring no physical touch or full disrobing, thereby addressing cultural modesty barriers (NIRAMAI Health Analytix, 2024). The output is an automated, explainable diagnostic report featuring a quantitative risk score and visual heat markers, supporting clinicians with objective decision-making inputs (NIRAMAI Health Analytix, 2024).

1

Input: Thermal Imaging

High-resolution thermal camera captures temperature distribution of the chest area from a distance (3-5 feet).

- Non-Contact & Radiation-Free

2

Input: Thermal Imaging

Data Processing: Data Upload Thermal images are encrypted and uploaded to the cloud (or processed offline on a laptop).

- Anonymized Data

3

AI Analysis: Thermalytix

Machine Learning algorithms analyze vascular patterns and thermal asymmetry to detect abnormalities.

- Automated Detection

4

Output: Diagnostic Report

Generates a quantitative "Breast Health Score" and a visual heat map indicating potential malignancy.

- Decision Support

Historical Stigma:

Traditional thermography failed in the 90s due to human error. Niramai fixed this with AI, but overcoming that historical skepticism takes time.

Proprietary Data:

Niramai owns the world's largest dataset of thermal-to-biopsy matched images. Competitors can buy cameras, but they don't have the data to train the AI.

Patents:

They have 39 patents on the specific algorithms that differentiate a 'cancer hot spot' from a 'normal infection hot spot'; which was the biggest failure point of previous technologies.*

Operational Workflow of the Thermalytix Solution

Deployment & Implementation

Deployed across 22 countries and 29 Indian cities, including rural Karnataka, Maharashtra and Punjab, Niramai has screened more than 100,000 women since 2017. The solution uses a portable thermal device that requires only basic internet connectivity for cloud processing, enabling operation in low-resource settings (NIRAMAI Health Analytix, 2024). Developed by Niramai Health Analytix, it operates through a B2B2C model, partnering with diagnostic chains such as

Apollo and with NGOs (NIRAMAI Health Analytix, 2024). Integration with state health departments (e.g., the National Health Mission) and corporate social responsibility initiatives facilitates mass village camps, where local health workers mobilize women and ensure last-mile delivery and government-backed referrals.

Users & Scale

Benefiting underserved women in rural and semi-urban India, Niramai has completed more than 100,000 screenings across 29 districts. Access is enabled through hospital and NGO partnerships and community-led screening programmes (NIRAMAI Health Analytix, 2024). Listing on the Government e-Marketplace further enables government procurement, supporting wider public-sector availability (NIRAMAI Health Analytix, 2024).

Impact & Outcomes

Measured Gender Impact

Niramai has screened more than 100,000 women across 29 Indian cities and rural districts, with clinical studies confirming 91 per cent sensitivity, which is higher than conventional mammography for women with dense breasts (Adapa et al., 2025). This intervention has driven significant behavioural change. The non-contact, privacy-preserving nature of Thermalytix has increased uptake by addressing modesty concerns that previously deterred screening (NIRAMAI Health Analytix, 2024; Manjunath, Venkataramani and Kakileti, 2022). Economically, the solution reduces screening costs by approximately 90 per cent compared to mammography, eliminating out-of-pocket travel expenses for rural women by bringing diagnostics to their communities through village camps.

Socially, this decentralization strengthens women's agency over health-related decision-making by reducing reliance on male family members to accompany them to distant urban hospitals. From a health equity perspective, Niramai narrows the urban–rural gap by enabling earlier detection (at Stage I rather than the more typical Stage III or IV diagnosis), which improves survival prospects and reduces treatment burdens (Manjunath, Venkataramani and Kakileti, 2022). The process is efficient, delivering automated reports in under 15 minutes and substantially reducing anxiety and waiting times. This impact directly advances SDG 3 (Good Health and Well-being) and SDG 5 (Gender Equality) by supporting equitable access to life-saving screening technologies for underserved women, regardless of geography or income.

Ethics, Governance & Safeguards

Niramai protects user privacy through data anonymization and encryption, aligning with GDPR standards and principles set out in India's Digital Personal Data

Protection Act (NIRAMAI Health Analytix, 2024). Informed consent is mandatory prior to screening, and data are stored in secure, access-controlled cloud environments. To mitigate algorithmic bias, the AI was trained on a diverse dataset of Indian women across a range of ages and body types, reducing reliance on Western-centric models (NIRAMAI Health Analytix, 2024). Oversight is provided by an internal ethics committee, while regulatory compliance is supported through United States Food and Drug Administration clearance for SMILE-100 and CE marking (NIRAMAI Health Analytix, 2024). Female technicians lead field operations, reinforcing gender-sensitive governance and building trust during community deployment (NIRAMAI Health Analytix, 2024).

Lessons & Replicability

What Worked

NIRAMAI Health Analytix was founded in 2016 to develop Thermalytix, a non-invasive, AI-based breast cancer screening solution for women (NIRAMAI Health Analytix, 2024). Success has hinged on a privacy-centric, non-contact design, strong NGO partnerships and portable, offline-capable hardware. Thermalytix is now deployed in hospitals and screening programmes with United States Food and Drug Administration clearance for SMILE-100 (NIRAMAI Health Analytix, 2024).

Challenges Encountered

Challenges included initial clinical scepticism regarding the accuracy of thermal imaging compared to standard mammography (NIRAMAI Health Analytix, 2024). Operational scaling faced barriers linked to intermittent rural power supply and connectivity, necessitating offline-capable hardware (World Economic Forum, 2021). In addition, deep-rooted cultural stigma surrounding cancer diagnoses and low health literacy required sustained community engagement to build trust and encourage participation in screening camps (NIRAMAI Health Analytix, 2024).

Key Lessons for Replication

Replication requires prioritizing government partnerships early, rather than relying solely on private markets, in order to build trust (NIRAMAI Health Analytix, 2024). Solutions must be adapted to local norms around cultural modesty and data protection, avoiding rigid 'copy-paste' deployment models (NIRAMAI Health Analytix, 2024). Essential prerequisites include offline-first technical design and investment in local female workforce capacity to support sustained community adoption (NIRAMAI Health Analytix, 2024).

Conclusion & Future Directions

Niramai exemplifies how privacy-centric AI can advance gender equality by overcoming cultural barriers that limit access to breast cancer screening (NIRAMAI Health Analytix, 2024). Validated in India, the model is now scaling to the Philippines and Kenya, with future plans to support accessible home-based screening (NIRAMAI Health Analytix, 2024). This experience establishes a benchmark for inclusive digital health, demonstrating that technology must be culturally responsive in order to genuinely democratize healthcare (NIRAMAI Health Analytix, 2024).

References

Adapa, K., Gupta, A., Singh, S., Kaur, H., Trikha, A., Sharma, A., & Rahul, K. (2025). *A real world evaluation of an innovative artificial intelligence tool for population-level breast cancer screening*. https://niramai.com/about/thermalytix/

Chaudhary, A., Mittal, A., Rangarajan, K., & Nair, N. (2021). Observational study to evaluate the clinical efficacy of Thermalytix for breast cancer detection. *Journal of Global Oncology*, 7, 1–10. doi:10.1200/GO.20.00305

International Journal of Research in Pharmacy and Review. (2022). *Barriers to breast cancer screening in urban India. International Journal of Research in Pharmacy and Review, 6*(6).

Manjunath, G., Venkataramani, K., & Kakileti, S. T. (2022). Feasibility and outcomes of using a novel artificial intelligence–enhanced breast thermography technique, Thermalytix, in screening for breast abnormalities at primary health centres at the community level in South India. *International Journal of Community Medicine and Public Health, 9*(12), 4634–4640. https://www.ijcmph.com/index.php/ijcmph/article/view/10481

NIRAMAI Health Analytix. (2024). Thermalytix technology overview. *npj Digital Medicine, 8,* Article 2. https://www.nature.com/articles/s41746-024-01368-2

World Economic Forum. (2021). *AI can make breast cancer screening more affordable*. https://www.weforum.org/agenda/2021/11/ai-breast-cancerscreening-more-accessible-and-affordable/

World Health Organization. (2023). *Breast cancer*. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Case Study 15

KNOWLEDGE

CheetahFem: Multilingual AI for Women's Information Access

Author: Ife Adebara, Nigeria

Overview

Background & Problem Statement

Across the Global South, women often struggle to access reliable information due to systemic barriers, including linguistic exclusion. Most digital systems are optimized for colonial languages, leaving speakers of indigenous African languages behind (Adebara, 2025; Adebara and Abdul-Mageed, 2022). This disproportionately affects women, undermining access to health information, education and economic opportunities. Traditional approaches, such as static translation, printed materials or reliance on human interpreters, are not scalable across hundreds of indigenous languages and dialects. As a result, a gap has persisted in AI systems capable of generating contextually accurate text across a wide range of African languages to support gender-inclusive access to information.

Solution Summary

CheetahFem adapts the research model Cheetah (an African languages natural language generation system trained across 517 African languages and language varieties) to expand women's access to information. It supports multilingual text generation and translation in domains relevant to women, helping to reduce language-based digital exclusion (Adebara et al., 2024; Adebara, 2024).

The AI Solution

Solution Description

CheetahFem builds on Cheetah, a large natural language generation model trained on text from 517 African languages and variants (Adebara et al., 2024). During training, the model learns patterns across languages and scripts and generates outputs by predicting the most likely next words or characters given an input context. It supports tasks such as translation, question answering and summarization. CheetahFem emphasizes gender-responsive applications by prioritizing languages commonly spoken by women and supporting informational tasks central to women's lives, including health, education, safety and

livelihoods. Outputs are text-based and designed to support transparency and interpretability in informational contexts.

Deployment & Implementation

CheetahFem is currently at the pilot stage, with testing conducted among a small number of users for research and validation purposes. Initial pilots are being carried out in selected communities and academic or civil society settings to assess usability, language coverage and contextual relevance. The system is deployed through lightweight web-based and mobile interfaces, such as chatbots or messaging platforms, supported by cloud infrastructure. Implementation prioritizes low-bandwidth environments and minimal device requirements. Local research partners and civil society organizations support testing, feedback collection and iterative refinement of language coverage and content relevance.

Users & Scale

Primary beneficiaries are women and girls, particularly in rural, multilingual and under-served communities. Users access CheetahFem through mobile-friendly services, community information systems and AI-enabled support tools that deliver information in indigenous languages. Early-scale implementations are intended to reach thousands of users through partner-operated digital services and community-facing programmes.

Impact & Outcomes

Measured Gender Impact

Early pilot use of CheetahFem indicates that providing information in women's preferred languages improves access to health, civic, and economic content, reducing reliance on English or French. Community partners report faster information flows and reduced dependence on intermediaries. Evaluations of the parent model Cheetah, which underpins CheetahFem, show that the system significantly outperforms other multilingual language models on most natural language generation tasks for African languages, achieving leading results across benchmarks and outperforming competitors on most test sets (e.g., higher task accuracy and fluency). These outcomes support more informed decision-making, time savings and strengthened agency, aligning with SDG 5 on gender equality.

Ethics, Governance & Safeguards

CheetahFem is designed with safeguards appropriate for informational and support settings. User data is minimized, anonymized where possible and

collected with explicit consent. Bias mitigation includes broad multilingual coverage and continuous evaluation for systematic errors that may disadvantage specific communities or dialects. Women participate as contributors, annotators, advisors and evaluators, supporting the assessment of linguistic and cultural appropriateness. The system is positioned as a decision-support and information-access tool, not as an automated authority for high-stakes decisions, reducing risks of harmful over-reliance. Governance is maintained through partner oversight, feedback mechanisms and documented processes for handling complaints and correcting outputs.

Lessons & Replicability

What Worked

Training on a large, diverse corpus of African languages proved essential for strong multilingual performance, as demonstrated in evaluations of Cheetah. Combining scale with careful data curation and community-informed validation improved relevance and trust. Text-first, low-bandwidth deployment enabled use in resource-constrained settings, while a focus on women-relevant domains supported practical impact.

Challenges Encountered

Key challenges include uneven data availability across under-resourced languages, ensuring cultural nuance in generated outputs and addressing digital literacy gaps. Infrastructure constraints, such as intermittent connectivity and limited access to devices, can restrict adoption. Trust-building remains essential in communities with limited prior exposure to AI systems.

Key Lessons for Replication

Effective replication requires local adaptation rather than direct transfer. Strong partnerships with local organizations and co-design with women users are essential to ensure linguistic and cultural relevance. Ongoing quality monitoring across languages, user guidance and integration into existing health, education and social services strengthen sustainability, trust and accountability.

Conclusion & Future Directions

CheetahFem demonstrates how linguistically inclusive AI can advance gender equality by reducing language barriers to information access. Future directions include expanding deployments across regions and sectors, strengthening

community co-design, improving evaluation in women-centered contexts and integrating with national digital inclusion strategies to support gender-responsive AI ecosystems.

References

Adebara, I., Elmadany, A., & Abdul-Mageed, M. (2024). *Cheetah: Natural language generation for 517 African languages*. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics

Adebara, I. (2024). *Towards Afrocentric natural language processing (T)*. University of British Columbia. https://open.library.ubc.ca/collections/ubctheses/24/items/1.0440415

Adebara, I. (2025). *AI and language data flaring in Africa: Addressing the lowresource challenge (Policy Brief No. 216).* Centre for International Governance Innovation. https://www.cigionline.org/static/documents/no.216_Adebara.pdf

Adebara, I., & Abdul-Mageed, M. (2022). *Towards Afrocentric NLP for African languages: Where we are and where we can go*. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 3814–3841)*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.271

Case Study 16

HEALTH

MAAP AI: Community-Led AI for Malnutrition Action

Author: Romita Ghosh, India

Overview

Background & Problem Statement

In many communities across the Global South, routine childhood growth monitoring is irregular, equipment is broken or miscalibrated, and paper logs are error-prone, delaying the detection of undernutrition and disproportionately burdening mothers and frontline women care workers. MAAP AI addresses these gaps by replacing height boards and manual entries with smartphone-based anthropometry and WHO-aligned analytics that function in homes, Anganwadis and schools, including settings with limited connectivity or clinical staff. The approach is community-led: caregivers capture measurements and generate reports, creating a reliable and auditable data stream for programme managers.

Solution Summary

MAAP AI estimates a child's height from a single photograph, logs a tamper-proof, geo-tagged record, computes Z-scores and provides personalized nutrition guidance without additional hardware. Women frontline workers use the app for mass screening and timely referrals, while community nutrition managers access secure, role-based dashboards.

The AI Solution

Solution Description

MAAP (Malnutrition Assessment & Action Plan) is an artificial intelligence (AI) smartphone application that estimates a child's height from a single photograph and converts it into actionable growth insights. It uses computer vision to analyse body proportions, combined with machine-learning and predictive analytics to flag early growth faltering. Women frontline workers co-designed the workflows, and the models are trained on representative, gender-disaggregated data to reduce bias.

The app functions offline, recording tamper-proof geo-tags and outputting WHO-aligned indicators and Z-scores (indicating how far a child's measurements

deviate from healthy norms). Results are presented through simple traffic-light charts, referral prompts and diet guidance that mothers can act on.

Deployment & Implementation

MAAP reports that approximately 300,000–320,000 children have been screened to date across programmes implemented with NGOs, schools and government partners, alongside ongoing validation studies. Frontline workers capture photographs during home or centre visits, with data synchronizing to a secure cloud when connectivity is available. Flagged cases automatically trigger referral prompts. Dedicated school modules support older children and school meal programmes. The system operates on Android smartphones, enables offline data capture with deferred synchronization, and relies on cloud GPU and on-premise resources for model training and updates. MAAP is designed for interoperability with national digital platforms, including Poshan Tracker 2.0.

Users & Scale

Primary users of MAAP are Accredited Social Health Activists (ASHAs) and Anganwadi workers, who are predominantly women and conduct community-level screenings. Secondary users include primary- and community- health centre teams, Nutrition Rehabilitation Centre staff, paediatricians, schools monitoring meal programmes, and programme managers overseeing implementation. The solution provides access to de-identified datasets covering around one million children. The access model centres on community workers and schools, enabling women caregivers to receive nutrition guidance at or near the household level.

Impact & Outcomes

Measured Gender Impact

MAAP's gender impacts are direct and multi-layered. By making growth risks visible at an early stage, the solution interrupts intergenerational cycles of undernutrition and enables timely referrals and targeted nutrition support. Given that India's Anganwadi and ASHA workforce is predominantly female, the single-photo and paperless workflow reduces manual burdens, upgrades skills and creates paid roles in screening and supervision.

Mother-centered data literacy and gamified nutrition coaching support shifts towards diverse, safe, climate-resilient, low-waste and affordable household diets, improving adherence and children's attentiveness. Communities have reported practical change: 5,000 schoolchildren were counselled alongside teachers, and healthcare workers reported mother-led diet improvements. Across partner

programmes, earlier referrals for suspected growth delays were documented, while governments and NGOs cite time and cost savings from digital measurement and reporting.

Women caregivers receive personalized meal plans rooted in local foods, including millet-based staples and nutrient fortification using culturally familiar ingredients such as moringa, making guidance immediately actionable. Training handbooks and a skilling pathway aim to empower more than one million frontline women as deployments scale with public and NGO partners, expanding paid roles in screening, supervision and data quality. For system learning, MAAP maintains population-scale datasets, enabling gender-disaggregated analysis and national integration. These outcomes advance SDG 5 and reinforce SDGs 1, 2 and 3 through early detection and support.

Ethics, Governance & Safeguards

MAAP embeds safeguards by design, including tamper-proof geo-tagging, secure cloud storage and role-based access controls. Data collection is minimal, anonymized where appropriate and purpose-driven, avoiding unnecessary identifiers. Explainability relies on standard growth charts and WHO-recommended metrics familiar to clinicians. Governance is shared with programme partners (NGOs, schools and government), supported by audit trails and validation studies to monitor performance.

Lessons & Replicability

What Worked

MAAP was adapted to local languages, dietary practices and frontline workflows, easing adoption. Co-created with underserved families and frontline women, the solution integrated into daily routines rather than disrupting them. Models were trained on consented, representative data. Intellectual property was safeguarded while enabling secure, standards-based interoperability and privacy, balancing innovation, trust and responsible scale.

Challenges Encountered

Variations in the field (e.g., in lighting, posture and position) required continuous model retraining and tester-led usability improvements. Ensuring consistent use and data quality across cadres requires refresher training and supervision, while legacy practices such as paper records and proxy logging persist in some sites.

Key Lessons for Replication

Design for the last mile by starting with deep field research to understand real workflows and pain points. The AI use case should follow the problem, rather than the reverse. Co-creation with end users is essential. Localized content and representative data help ensure relevance and fairness. Sustained investment in training, supervision and feedback loops is necessary to maintain quality, trust and long-term adoption.

Conclusion & Future Directions

Community-led, gender-responsive AI can close measurement gaps at the household level while strengthening public nutrition programmes. MAAP AI is positioned to scale responsibly through partnerships and to expand into new geographies and school nutrition settings. Next steps include system integrations, continued validation and further localization of content. By standardizing ethical, community-led tools, MAAP AI advances a gender-inclusive public health AI ecosystem.

References

RevolutionAIze Pvt. Ltd. (2026). *MAAP AI*.

World Health Organization. (n.d.). *WHO Multicentre Growth Reference Study (MGRS)*. WHO Child Growth Standards.

Case Study 17

HEALTH

AI for Transgender Women's Empowerment

Authors: Vijay Jeyakumar, Prema S, Nithiya R, India

Overview

Background & Problem Statement

This initiative addresses structural gender inequality affecting transgender women, including exclusion from education, employment, healthcare, identity documentation and social welfare systems. In the Global South, poverty, limited accessibility, digital illiteracy and social stigma further intensify these barriers. Prior to this AI solution, services were fragmented, unsafe, discriminatory and poorly adopted. Manual NGO support and basic policy portals lacked scalability, personalization, confidentiality and real-time guidance, making them inadequate to address the complex and intersectional needs of transgender women.

Solution Description

The proposed solution is an AI-enabled digital platform designed to support transgender women by providing secure access to welfare schemes, healthcare, mental health assistance, safety tools and entrepreneurship opportunities. The platform advances gender equality through natural language processing–driven guidance, machine-learning-based personalization and bias-audited recommendations, enhancing access, autonomy, economic inclusion and social empowerment.

The AI Solution

Solution Description

The AI system functions as a secure online assistant that provides personalized guidance to transgender women based on self-identified needs. It uses machine learning to generate tailored recommendations and natural language processing to understand and respond to user queries. Co-designed as gender-responsive, the platform incorporates bias-audited models and voluntary self-identification mechanisms. A privacy-by-design architecture ensures encrypted data storage, role-based access and consent-driven feature use, while restricting unauthorized access. Transparent and explainable AI logic supports eligibility and service recommendations, fostering trust, accountability and user confidence.

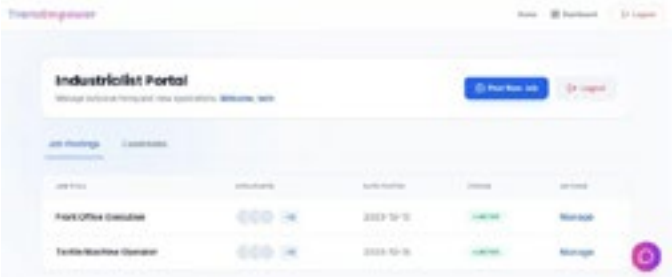
Operational Workflow of the Thermalytix Solution

Deployment & Implementation

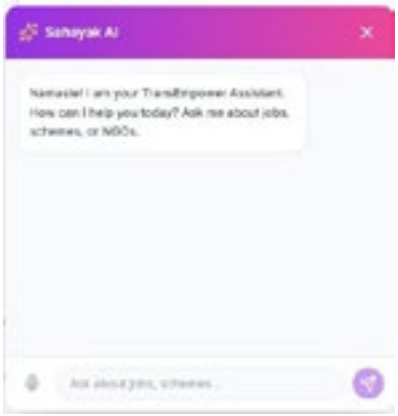
The initiative is currently in the pilot phase, having been tested with fewer than 1,000 beneficiaries at a single site for under 12 months. Its geographic focus is on urban and peri-urban areas of Tamil Nadu, India, with plans to scale to other states. The implementation timeline spans 12 months, with months 0–3 dedicated to co-design and ethics review, months 4–6 to platform development and testing, months 7–9 to pilot rollout and feedback, and months 10–12 to evaluation and scale readiness. The partnership model brings together an academic–technology team and transgender-led NGOs under a shared governance arrangement. The platform relies on smartphones and other web-enabled devices, low-bandwidth Internet connectivity, secure cloud storage and multilingual interfaces, and integrates with government portals, health directories and helplines through APIs (application programming interfaces). Institutional actors play a supporting role through policy alignment, data verification, funding and scale-up support.

Users & Scale

Primary users are transgender women, while secondary users include job providers, counsellors and community facilitators. The pilot targeted between 500 and 1,000 users in its first year. NGO outreach workers and community centres supported access to the platform through web- and mobile-based applications, providing inclusive and assisted onboarding as well as continued engagement.

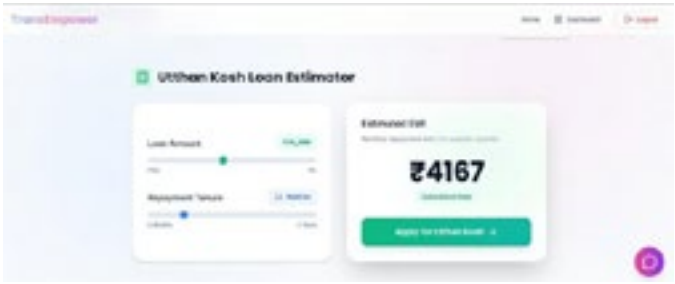


Employer Portal to Post Jobs

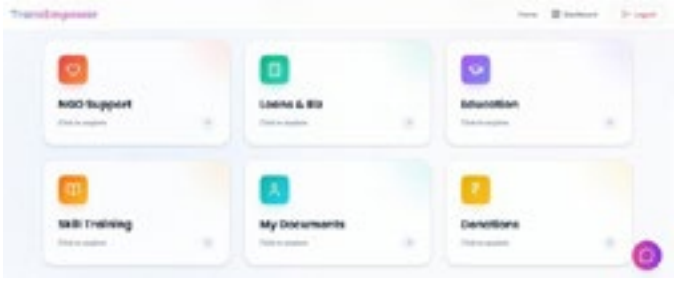


AI-assisted Support (Sahayak AI)

Community Connect Page



Loan Assistance through AI Support



Dashboard

Impact & Outcomes

Measured Gender Impact

The pilot reached 1,051 transgender women across urban and peri-urban areas. Compared with baseline NGO-only outreach, awareness of and access to welfare, healthcare and legal services increased to 55 per cent. Economic outcomes were also notable, with 30 per cent of participants accessing livelihood opportunities, skills programmes or micro-entrepreneurship platforms, contributing to improved income security. Social impacts included enhanced safety and confidence through anonymous access and emergency tools, which helped reduce stigma-related avoidance. Use of mental health assistance and digital literacy resources increased, particularly among participants engaging with formal digital services for the first time. In terms of efficiency, centralized, AI-assisted navigation reduced the time required to access services by approximately 30–40 per cent. Prior to the intervention, access was characterized by low continuity and fragmented, manual NGO mediation. Overall, the initiative aligns with SDG 5 (Gender Equality), SDG 8 (Decent Work and Economic Growth), SDG 10 (Reduced Inequalities) and SDG 3 (Good Health and Well-being).

Ethics, Governance and Safeguards

Privacy-by-design principles, including data minimization, encryption, role-based access and informed opt-in consent, safeguard user data. Bias reduction is supported through community co-design, gender-disaggregated analysis and continuous auditing of AI outputs. Accountability and ethical safeguards are embedded from the outset to ensure transparency, explainability and user control. Technical teams, transgender women representatives and partner organizations collaboratively review and audit the system, supported by grievance and redress mechanisms. Potential risks, including misuse or exclusion, are addressed through adaptive governance, user feedback loops and ongoing human-rights-based evaluation.

Lessons and Replicability

What Worked

Community co-design, privacy-first architecture and AI-enabled personalization were central to effectiveness. Collaboration with transgender-led NGOs and a low-bandwidth, mobile-first design supported trust-building, adoption and accessibility. Localized content and flexible governance, enabled by modular technology, allowed adaptation across diverse socio-economic and cultural contexts.

Challenges Encountered

Variation in digital literacy, shared or restricted device access, language diversity and initial trust posed challenges. These were addressed through assisted onboarding, community-based outreach and strong privacy assurances. Connectivity constraints, language differences and regulatory changes required adaptive design, multilingual support and continuous adjustment to user needs throughout deployment.

Key Lessons for Replication

Future iterations should expand multilingual support, strengthen digital literacy components and further streamline onboarding processes. Context-sensitive adaptation requires local partnerships, cultural awareness and flexible technology. Successful replication depends on reliable connectivity, sustained community engagement and institutional alignment. Privacy should remain a priority, beneficiaries should be involved in co-design, and governments and organizations should undertake ongoing monitoring to ensure comprehensive and inclusive impact.

Conclusion & Future Directions

This AI platform advances gender equality by empowering transgender women through improved access to services, economic opportunities and holistic health and mental well-being support. Scalable across regions and adaptable to other marginalized groups, next steps include expanding reach, refining AI support and strengthening community partnerships to foster an inclusive, safe and empowering AI ecosystem.

References

1. Thapa, A., Kumar, R., Kuppuswamy, R., Bairwa, M., & Das, A. (2025). *Lived experiences, health-seeking behaviour, quality of life, and self-esteem among transgender women: A mixed methods study from the Northern Himalayan region. BMC Public Health, 25*, Article 3836. https://doi.org/10.1186/s12889-025-25045-5
2. Umashankar, S., Sivakumar, G., Sundar, D. K., et al. (2025). *Health care access and utilization among transgender adults in Chennai: A cross-sectional study. International Journal for Equity in Health, 24*, Article 283. https://doi.org/10.1186/s12939-025-02556-6
3. National Human Rights Commission. (2025). *Transgender persons: Revamping spaces, reclaiming voices – Insights from Garima Greh shelters and beyond*.
4. Census of India. (2011). *Transgender in India: Census 2011 transgender population and literacy data*. Census 2011 Transgender Data Summary.
5. PubMed. (2025). Meta-ethnography on mental health of Indian transgender individuals. *Mental Health of Indian Transgender Individuals: A Meta-Ethnography*.
6. Ahuja, T. K., Goel, A. D., Gupta, M. K., Joshi, N., Choudhary, A., Suman, S., Taluja, K., Mittal, M., Ghuman, N. K., Suthar, N., & Bhardwaj, P. (2024) Health care needs and barriers to care among the transgender population: a study from western Rajasthan. (2024). *BMC health services research, 24(1), 989*. https://doi.org/10.1186/s12913-024-11010-2
7. Indian Express. (2024). *Systemic barriers in healthcare access for transgender and gender-diverse persons in India*.
8. Drishti IAS. (2025). *What are the key challenges faced by transgender persons in India?*.

Case Study 18

SKILLING

AI Skilling for Women in Green Careers

Author: Dr Rimjhim Agrawal, India

Overview

Background & Problem Statement

Women remain underrepresented in advanced clean energy and AI-enabled technical roles, particularly in renewable energy forecasting and analytics. In the Global South, structural barriers such as limited access to contextualized technical training, low exposure to operational energy systems and reduced confidence in engaging with data-driven tools constrain women's participation. While renewable energy deployment has expanded rapidly, workforce inclusion has not kept pace. Existing learning approaches, often static or classroom-based, do not adapt to individual learning needs or sufficiently simplify complex concepts. As a result, many women with STEM backgrounds struggle to transition into AI-driven clean energy roles, underscoring the need for adaptive, accessible and gender-responsive skilling solutions.

Solution Summary

This case study presents a prompt-packaged conversational AI Skill Assistant designed to support women learners in clean energy and renewable forecasting. Using natural language processing and adaptive learning flows, the solution strengthens technical understanding and confidence, enabling greater participation in AI-enabled climate and clean energy career pathways.

The AI Solution

Solution Description

The AI Skill Assistant is a structured conversational learning system delivered through a conversational interface. The Skill Assistant uses natural language processing enabled by a large language model and guided by predefined prompts and learning modules. It provides interactive explanations, responds to learner questions and guides users through short modules covering clean energy fundamentals, renewable forecasting, variability and introductory supervisory control and data acquisition concepts. The system adapts explanations based on learner responses, offering simplified descriptions or follow-up prompts

as needed. Gender-responsiveness is embedded through inclusive language, confidence-building feedback and women-relevant examples. Explainability is ensured through step-by-step reasoning, plain-language definitions and clear positioning of the assistant as a learning support tool rather than a decision-making system.

Deployment & Implementation

The solution was developed and piloted over approximately four weeks through mentorship-led engagement with a small cohort of women learners, including engineering students and women mentees from STEM support programmes. Participants accessed the assistant via a conversational AI interface on personal devices with standard internet connectivity. No specialized hardware, institutional system integration or access to sensitive data was required. Feedback collected during the pilot informed refinements to prompts, learning flow and content clarity. The solution is currently operated by the author and designed for replication through women-focused learning and mentorship networks.

Users & Scale

Primary users were women engineering students and early-career STEM learners. The pilot engaged a small, intentionally selected cohort of women participants to validate learning effectiveness and usability. Participants used the assistant in short, self-paced learning sessions. The conversational access model supports flexible, low-barrier participation without requiring prior specialization in energy or AI.

Impact & Outcomes

Measured Gender Impact

Despite its pilot scale, the solution demonstrated clear early gender-specific outcomes. Participants reported increased confidence in understanding clean energy and renewable forecasting concepts following interaction with the AI assistant. Informal pre- and post-pilot feedback indicated improved comprehension of forecasting logic, renewable variability and data-driven decision-making. Several participants noted that the conversational format reduced the intimidation associated with technical subjects and enabled open questioning. Increased awareness of career pathways in clean energy, energy analytics and AI-enabled climate roles was also observed. By lowering cognitive and confidence barriers to advanced technical learning, the pilot contributed to SDG 5 (Gender Equality) and SDG 7 (Affordable and Clean Energy).

Ethics, Governance & Safeguards

Ethical and governance safeguards were embedded by design. The assistant does not collect or store personally identifiable information, and users were informed that interactions were strictly for learning purposes. Bias mitigation was addressed through inclusive language, avoidance of stereotypes and iterative review of responses during the pilot. Women participants provided feedback that informed prompt refinement and content improvement. Clear disclaimers communicated that the system was not intended for operational or safety-critical decision-making. Oversight remains with the developer, with mechanisms in place to correct inaccuracies or unintended outputs through ongoing monitoring.

Lessons & Replicability

What Worked

Adaptive, conversational learning improved engagement and reduced anxiety around complex energy topics. Leveraging trusted mentorship networks enabled rapid pilot deployment, high participation and meaningful feedback within a short implementation period.

Challenges Encountered

Differences in digital literacy and baseline technical knowledge required careful simplification of explanations. Ensuring relevance across varied educational backgrounds remained a key design and content challenge during the pilot.

Key Lessons for Replication

Successful replication requires localized examples, low-infrastructure delivery and integration with trusted women-focused networks. Gender-responsive AI skilling initiatives should prioritize confidence-building alongside technical depth, rather than assuming prior expertise.

Conclusion & Future Directions

This pilot demonstrates the potential of conversational AI to support gender-inclusive skilling in the clean energy and climate domains. With minimal infrastructure requirements and high adaptability, the solution can be scaled geographically and extended to additional renewable technologies, strengthening inclusive AI ecosystems in the Global South.

References

International Energy Agency. (2023). *World energy outlook 2023*.

International Renewable Energy Agency. (2022). *Renewable energy and jobs: Annual review 2022*.

National Renewable Energy Laboratory. (2020). *Wind power forecasting: A primer*.

UN Women. (2023). Gender equality and women's empowerment in the context of climate change.

World Economic Forum. (2023). *Global gender gap report 2023*.

Case Study 19

AI LITERACY

TAGS: Gender-Responsive AI for Women's Digital Empowerment

Author: B Spoorthi

Overview

Background & Problem Statement

Despite rapid digitalization in India, rural, semi-urban and tribal women continue to face systemic barriers to digital access and participation. These include limited device ownership, low AI literacy, heightened exposure to online risks and exclusion from emerging technology-driven livelihoods. Studies show that rural girls aged 14–18 are nearly half as likely as boys to own a computer or smartphone, reinforcing educational and economic inequities (Asian Development Bank [ADB], 2024).

In the Global South, where women constitute nearly half the population, such disparities undermine inclusive growth and leadership. Earlier digital literacy initiatives – often one-off and tool-centric – lacked scalability, cultural adaptability and responsiveness to linguistic diversity. Non-AI interventions failed to address gendered learning needs dynamically or to embed safety, agency and confidence-building as core outcomes.

Solution Summary

The Tech Ambassadors on Ground (TAGS) initiative leverages natural language processing- and machine learning-enabled, community-based AI to support women's digital empowerment. TAGS' multilingual, feedback-driven design strengthens digital confidence, leadership, entrepreneurial skills and safe technology use, with more than 75 per cent of participants reporting measurable gains in digital self-efficacy.

The AI Solution

Solution Description

TAGS integrates natural language processing and machine learning to deliver context-aware, multilingual digital learning experiences tailored to women's lived realities. These learning activities include accessing government services, engaging with livelihood platforms for marketing rural produce and products, and using educational resources for skills development.

Natural language processing supports the creation and processing of Telugu- and English-language prompts, voice inputs and instructional materials, while machine learning analyses anonymized participant feedback to adapt pacing, content complexity and facilitation strategies. Gender-responsiveness is embedded through participatory, co-designed workshops, gender-disaggregated data analysis and bias audits that alert facilitators to stereotypical assumptions. This approach enhances the relevance and effectiveness of AI-enabled learning for women's empowerment (ADB, 2024).

Deployment & Implementation

TAGS has been implemented through more than 20 workshops in collaboration with academic institutions, NGOs and local colleges. The initiative relies on smartphones, low-bandwidth modes and offline-capable open-source AI models (e.g., Hugging Face) to address connectivity challenges. TAGS is integrated into college extension activities, district women's self-help groups and community outreach programmes, achieving over 90 per cent participant completion rates through facilitator-led, trust-based delivery models.

Users & Scale

Primary users include women students (40 per cent), grassroots women leaders (35 per cent) and community members (25 per cent), with women constituting 82 per cent of all participants. To date, TAGS has reached more than 580 women directly and over 600 women indirectly through cascade training, with peer-led diffusion expanding engagement beyond formal workshops via mobile-enabled AI tools.

Impact & Outcomes

Measured Gender Impact

TAGS directly reached more than 580 women across rural and tribal communities. Evaluation data indicate that 75 per cent of participants reported increased digital confidence, 68 per cent adopted safer online practices, such as improved privacy settings, and 62 per cent increased their use of digital platforms for education or livelihood purposes, including online sales of rural produce and products. Participants reported a 30 per cent reduction in time spent seeking information, while 55 per cent joined local leadership or community decision-making groups. These outcomes align with SDG 5 (Gender Equality), particularly target 5.5 on women's leadership, and SDG 4 (Quality Education) on digital skills, demonstrating stronger outcomes than traditional, non-AI digital literacy models.

TAGS workshops with women. TAGS in the news

Ethics, Governance & Safeguards

TAGS follows a human-centered AI governance framework grounded in consent, transparency and data minimization. Participation is opt-in, with only anonymized, non-identifiable data collected and securely stored. Bias mitigation is addressed through gender-lens content reviews and fairness checks embedded in facilitator training. Women participants are actively involved in governance and feedback processes. Advancing cyber resilience and digital gender security requires a shift towards human-centered approaches that prioritize women's safety and agency in digital spaces (United Nations University, 2023). Offline access options and grievance mechanisms address unintended harms and access gaps.

Lessons & Replicability

What Worked

Women-led facilitation, culturally grounded examples and Telugu voice-based natural language processing fostered trust and participation. Strong NGO–academic, academic–government and industry–academic partnerships ensured contextual relevance, while continuous feedback loops contributed to an 85 per cent retention rate. Offline AI functionality proved effective across diverse community settings.

Challenges Encountered

Key challenges included intermittent internet connectivity, initial scepticism towards AI, multilingual content alignment and time constraints for participants. These challenges were mitigated through simplified interfaces, offline fallback options, flexible scheduling and facilitator mediation.

Key Lessons for Replication

Successful replication requires deep co-design with local women, linguistic adaptation and training of diverse women facilitators. Basic smartphone access and institutional partnerships are prerequisites. Earlier integration of offline-first AI models and alignment with women's self-help group structures can enhance sustainability and scale.

Conclusion & Future Directions

TAGS demonstrates how ethical, gender-responsive AI can advance women's digital empowerment in the Global South. With plans to scale to 2,000 women by 2026, future modules will address health, entrepreneurship and integration with India's Digital Public Infrastructure, contributing a replicable model for inclusive AI ecosystems.

References

Asian Development Bank. (2024). *Artificial intelligence and gender equality*. https://www.adb.org/sites/default/files/publication/1092626/adb-brief-357-artificial-intelligencegender-equality.pdf

United Nations University. (2023). *Empowering women in the AI generation: Toward advancement of the SDGs*. https://unu.edu/macau/blog-post/empowering-women-ai-generationtoward-advancement-sdgs

Case Study 20

SAFETY

Gender-Responsive AI Policing Against Deepfakes in Rural India

Authors: Pranav Dwivedi, Pranjal Dwivedi, India

Overview

Background & Problem Statement

The rapid proliferation of AI-generated deepfakes has intensified technology-facilitated gender-based violence globally. Evidence indicates that over 90 per cent of non-consensual synthetic media consists of sexually explicit content targeting women and girls, reinforcing gendered patterns of online abuse and coercion. The increasing accessibility of deepfake generation tools has further accelerated this harm, enabling the large-scale production of realistic manipulated media with minimal technical expertise.

In the Global South, these risks are compounded by structural vulnerabilities such as low digital literacy and limited survivor support systems. In India, the sharp rise in reported cybercrime – over 3.6 million complaints recorded in 2024 alone – underscores the scale of AI-enabled digital harm confronting public institutions.

In districts such as Shrawasti, Uttar Pradesh – where female literacy remains below 35 per cent and multidimensional poverty affects a majority of households – women face heightened exposure to reputational harm, coercion and social exclusion arising from deepfake misuse.

Solution Summary

This initiative introduced India's first structured, AI-enabled deepfake detection and response framework designed for rural policing. Implemented by the Shrawasti Police, with administrative support from the District Magistrate and the judiciary, and technical and curriculum inputs from Netratvshaala (a private limited company based in Uttar Pradesh), Inclusive AI and Future Shift Labs (implementation partner), the programme combined offline-capable forensic tools with targeted capacity-building. The approach strengthened gender-responsive institutional responses to AI-enabled harm against women.

The AI Solution

Solution Description

The solution centres on a structured, in-person capacity-building programme for district law enforcement, aimed at strengthening institutional responses to deepfake-related crimes against women. Through facilitated workshops, officers were trained to understand how AI-generated deepfakes are created, weaponized and disseminated, and how such cases intersect with existing legal provisions under the Bharatiya Nyaya Sanhita (BNS) 2023 and the Information Technology Act, 2000. A dedicated deepfake detection toolkit, developed by Pranjal Dwivedi, co-founder of Inclusive AI, supported training with AI-assisted techniques – including Reality Defender and other global tools – for identifying synthetic visual, audio and metadata anomalies. Officers undertook hands-on exercises and scenario-based simulations to identify manipulated media, assess veracity and document evidence in legally admissible formats, reinforcing both technical literacy and procedural confidence.

Gender-responsiveness was embedded through the positioning of women officers as lead trainees, the use of gender-specific case simulations and trauma-informed survivor engagement practices. Training emphasized that women are often victims of deepfake abuse and should not be blamed when such content is circulated. The programme also introduced methods to intervene and limit the spread of deepfakes across social media platforms and websites, including reporting mechanisms, content takedown requests and community awareness activities.

Deployment & Implementation

The programme is currently in its second phase in Shrawasti district, building on a six-month pilot conducted in 2024–2025 that combined workshops, simulated investigations and field application. Implementation is led by Superintendent of Police Ghanshyam Chaurasia, with administrative coordination from District Magistrate Kritika Sharma and ongoing engagement with the district judiciary under District Judge Devendra Singh. Early outputs from this phase have informed operational refinement, with officers reporting increased confidence in identifying synthetic media and applying deepfake evidence within investigative workflows (internal evaluation data).

Building on this momentum, the model is planned for replication across Uttar Pradesh, aligning with the state's broader adoption of AI and digital tools in policing. Curriculum design and technical inputs continue to be provided by Inclusive AI, Netratvshaala and Future Shift Labs. Trained officers are integrating these competencies into routine workflows, including First Information Report registration, digital evidence handling and survivor referral pathways.

Users & Scale

Primary users are police officers and district-level law enforcement staff. Across Shrawasti district, more than 200 officers were trained, with women comprising 35 per cent of participants and leading most community outreach initiatives. Through workshops and public engagement activities, the programme directly reached over 5,000 citizens.

Impact & Outcomes

Measured Gender Impact

The intervention generated measurable gender outcomes at both institutional and community levels. More than 200 officers were trained, with women constituting 35 per cent of participants. Women officers led 60 per cent of community outreach activities, reaching over 5,000 citizens with guidance on AI-generated deepfakes, digital coercion, social media policies and safe reporting channels.

Officers reported an 85 per cent increase in confidence in identifying and responding to synthetic media abuse, including recognition of facial, audio and metadata manipulation in real cases. Training emphasized ethical handling of evidence to protect women's dignity, underscoring that the circulation of deepfakes does not imply consent.

As part of community engagement, Shrawasti Police and the district administration, led by Superintendent of Police Ghanshyam Chaurasia, conducted workshops for female students across 15 schools. These sessions focused on digital literacy, safe online practices and early recognition of AI-enabled harassment. Post-workshop surveys indicated that 78 per cent of participants felt better equipped to report cyber harassment and understood available institutional support mechanisms.

These efforts received international recognition through the Silver Polaris Award for Anti-Disinformation, reinforcing the initiative's contributions to gender-responsive AI policing and progress towards Sustainable Development Goal 5.

Ethics, Governance & Safeguards

Ethical safeguards were embedded from the outset to establish responsible standards for AI use in policing. All AI outputs required mandatory human verification in order to mitigate automation bias. Oversight by district police leadership incorporated women officers into governance and review mechanisms, while trauma-informed practices reduced repeated exposure to harmful content. These measures reinforced accountability, transparency and survivor dignity, in line with human rights and data protection principles and emerging international norms on AI governance.

Lessons & Replicability

What Worked

District-level ownership, strong institutional partnerships and integration into routine policing workflows were central to success. Positioning women officers as lead trainees and community educators accelerated adoption. Offline-capable AI tools and vernacular delivery enhanced effectiveness in a low-resource rural context.

Deepfake Detection Workshop with police

Training sessions conducted for government officials in coordination with the District Magistrate of Shravasti

Deepfake Detection Campaign awarded the Silver Polaris Award, Tweet from UP Police official handle

Challenges Encountered

Challenges included low baseline digital literacy, initial scepticism towards AI, and emotional strain associated with handling gender-based digital abuse. Connectivity constraints necessitated offline adaptations, while aligning training schedules with policing duties required ongoing adjustment.

Key Lessons for Replication

Future deployments must combine foundational AI literacy with continuous upskilling as synthetic media technologies evolve. Programmes should adapt to local legal, linguistic and infrastructural realities while remaining technically extensible. National scale-up will require interoperable tools, strong institutional ownership and forward-looking safeguards.

Conclusion & Future Directions

This case demonstrates that, even in low-literacy and resource-constrained districts, AI-enabled capacity-building can materially strengthen institutional responses to digital abuse. Embedding gender-responsive AI within frontline policing advances SDG 5 and offers a credible pathway for the Global South to govern emerging AI harms responsibly. Building on Shrawasti's experience, scaling to Uttar Pradesh's incoming police cohorts and broader national adaptation presents a significant opportunity to embed gender-responsive AI at the earliest stages of policing.

References

B. Jain. (2025). India's cybercrime reporting systems logged 36 lakh fraud cases in 2024; ₹22,845 crore lost. *The Times of India*. https://timesofindia.indiatimes.com/india/indias-cybercrime-reporting-systems-logged-36-lakh-fraud-cases-in-2024-rs-22845-cr-lost-over-10000-arrested/articleshow/122841677.cms

Online Bureau, ET Government. (2024). *AI misinformation concerns: Deepfake detection workshop to protect electoral integrity organized in Shravast*. ETGovernment.com. https://government.economictimes.indiatimes.com/news/secure-india/ai-misinformation-concerns-government-initiates-deepfake-detection-workshop-to-protect-electoral-iintegrity/109117377

Parliamentary Standing Committee on Home Affairs. (2025). *Cyber Crime Reporting and FIR Conversion Rates in India*. Medianama. https://www.medianama.com/2025/08/223-firs-low-cyber-crimes-rise-parl-report/

Registrar General & Census Commissioner, India. (2014). *District census handbook: Shrawasti, Uttar Pradesh (Census of India 2011, Series 10, Part XII-A)*. Directorate of Census Operations, Uttar Pradesh, Government of India. https://censusindia.gov.in/nada/index.php/catalog/1288/download/4166/DH_2011_0950_PART_A_DCHB_SHRAWASTI.pdf

Reuters. (2026). *UK urges platforms to act on sexualised AI-generated images targeting women and minors*. Reuters. https://www.reuters.com/technology/uk-tells-musk-act-fast-groks-sexualised-ai-images-sky-news-reports-2026-01-06/

Sippy, T., Enock, F., Bright, J., & Margetts, H. (2024). *Behind the Deepfake: Public Exposure and Perceptions of Deepfakes in the United Kingdom, arXiv preprint*. https://arxiv.org/abs/2407.05529

United Nations. (2024). *Ending violence against women and girls: Report of the Secretary-General (A/79/500)*. https://knowledge.unwomen.org/sites/default/files/2024-10/a-79-500-sg-report-ending-violence-against-women-and-girls-2024-en.pdf

Uttar Pradesh Police [@Uppolice]. (2025, June 19). *From local roots to global recognition: Shravasti Police honoured with the Silver Polaris Award 2025 for its deepfake detection campaign* [Post]. X. https://x.com/Uppolice/status/1935708522848690401

Hawkins, W., Russell, C., & Mittelstadt, B. (2025). *Deepfakes on demand: The rise of accessible non-consensual deepfake image generators* (arXiv preprint). https://arxiv.org/abs/2505.03859

Case Study 21

LAW

AymurAI: Structuring Judicial Data on Gender-Based Violence

Authors: vana Feldfeber, Mailén Amalia García, Milagro Teruel; Argentina, Costa Rica, Chile

Overview

Background & Problem Statement

AymurAI addresses a structural gender inequality in how knowledge about gender-based violence is produced, controlled and made visible within justice systems. In Latin America, this problem is intensified by weak transparency and accountability in judicial institutions, contributing to public distrust and a narrow focus on feminicides. As Ciolfi Felice et al. (2025) note, "the absence of traceable, high-quality data regarding gender-based violence is a pressing issue"; when data systems focus narrowly on feminicide, they obscure the broader spectrum of gender-based violence that precedes it and affects many more women. Prior to AymurAI, relevant information remained embedded in unstructured judicial documents, accessible only through slow and unsustainable manual processes. AymurAI facilitates the responsible transformation of this data into usable knowledge for prevention and oversight.

Solution Summary

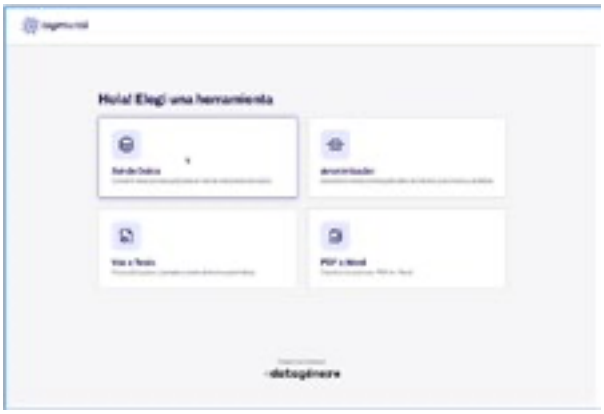
AymurAI is an open-source desktop application that uses natural language processing techniques to anonymize and structure judicial rulings on gender-based violence. Enabling and optimizing the responsible creation of anonymized justice datasets improves transparency, accountability and evidence-based policymaking to prevent gender-based violence and strengthen gender equality outcomes.

The AI Solution

Solution Description

AymurAI applies natural language processing techniques to Spanish-language judicial documents in order to support both dataset creation and anonymization. The system pre-fills structured form fields for expert validation, accelerating review processes and enabling the publication of open datasets. It also detects 30 types of sensitive information, which users review before finalization (Feldfeber et al., 2025). The resulting output is a copy of the original document in which sensitive elements are replaced with descriptive tags (e.g., "<NAME>").

The development process aligns with feminist epistemologies by rejecting claims to objective neutrality and instead centring situated knowledge (Ciolfi Felice et al., 2025). AymurAI is not designed as an automated or solutionist system – before any action is confirmed, data is reviewed by a person, preserving human oversight within legal processes.



AymurAI home interface with options to anonymize documents, create datasets, use speech-to-text tools and convert PDFs to Word documents.

Deployment & Implementation

AymurAI was developed in 2022 by DataGénero in collaboration with Criminal Court 10 of Buenos Aires, academic researchers and two private consulting firms, with funding from philanthropists and international organizations. The pilot launched in 2023 at Court 10 in Buenos Aires and reached the early scale stage during 2024, with stable deployment across more than 200 institutions in Argentina, Chile and Costa Rica.

The application has minimal infrastructure requirements and does not require internet access, operating on local court servers as well as on personal computers. A key feature of AymurAI is that judicial documents remain on the host computer at all times, reducing exposure to security breaches and supporting data protection within court environments.

Users & Scale

Primary users are paralegal employees who process court rulings on a daily basis. These users comprise approximately 2,000 court officials across Argentina, Chile and Costa Rica.

Impact & Outcomes

Measured Gender Impact

The impacts of AymurAI are centered on structural institutional change within the justice system. Through expanded deployment across judicial courts, the project shifts how cases of gender-based violence are documented, analysed and evaluated, strengthening the evidence base available for oversight and prevention.

20%

cases of physical violence have been reported, while sexual, economic, social, and environmental violence are reported less frequently

The project is creating a repository focused on gender-based violence to make judicial data systematically available for analysis, comprising more than 10,000 anonymized criminal court records from Latin America. Within this dataset, psychological and symbolic violence are the most prevalent forms, appearing in 28 per cent of records. Physical violence appears in 20 per cent of cases, while sexual, economic, social and environmental violence appear less frequently.

By making these patterns visible, the repository supports improved understanding of persistent challenges in addressing gender-based violence, strengthens institutional accountability and contributes to safer legal environments for women. The project aligns with SDG 5 (Gender Equality), SDG 16 (Peace, Justice and Strong Institutions) and SDG 9 (Industry, Innovation and Infrastructure).

Ethics, Governance & Safeguards

AymurAI embeds ethical, privacy and governance safeguards from design through deployment. Judicial documents are processed locally and anonymized prior to any training or dataset creation, and human validation is integral to the workflow, preventing automated exposure of sensitive information. Women's voices and feminist perspectives inform both development and ongoing governance, with accountability shared among DataGénero, court partners, academic researchers and technical stewards. Potential unintended consequences, including misclassification or misuse of outputs, are addressed through iterative review, transparency and contextual human judgement. The approach aligns with core human rights and data protection principles related to justice, dignity and non-discrimination.

Lessons & Replicability

What Worked

AymurAI's distinct contribution lies in positioning AI as a support tool for institutional development rather than automation. The system is embedded within everyday judicial routines, prioritizing usability over technical complexity. Development grounded in long-term relationships with courts enabled sustained use. Open-source architecture and gender-sensitive development approaches facilitated trust and contextual adaptation.

Challenges Encountered

Implementation revealed persistent tensions between technological innovation and judicial inertia. Fragmented documentation practices, limited staff capacity and uneven institutional commitment slowed adoption. Scaling required navigating heterogeneous and sometimes contradictory workflows across regions, while maintaining consistent ethical standards amid resistance to new data responsibilities and reluctance to adopt AI-based technologies.

Key Lessons for Replication

Replication requires framing AI as public-interest infrastructure rather than as a standalone product. Local legal expertise, gender-sensitive training and institutional buy-in are essential. Technical components must be adapted to local language, available infrastructure and resources, while capacity-building and governance design are as critical as the algorithms themselves.

Conclusion & Future Directions

AymurAI demonstrates how feminist AI can reconfigure justice data systems to advance gender equality. The approach engages across regions and legal domains where unequal access to justice persists for women, girls and LGBTIQ+ individuals. The project is expanding to include speech-to-text tools that support survivors during reporting processes, further strengthening feminist AI practices globally. AymurAI illustrates that alternative technological pathways are possible – human-centered, not-for-profit and explicitly gender-sensitive (Feldfeber, 2026).

References

Bercovich y Szulmajster, S., Feldfeber, I., García, M., & Quiroga, Y. B. (2021). *Datos con perspectiva de género y justicia abierta: La experiencia del Juzgado 10 (Informe final)*. ONU Mujeres. https://lac.unwomen.org/sites/default/files/Field%20Office%20Americas/Documentos/Publicaciones/2021/07/Datos%20con%20perspectiva%20de%20gnero%20thumbnail%20WEB.pdf

Ciolfi Felice, M., Feldfeber, I., Glasserman Apicella, C., Quiroga, Y. B., Ansaldo, J., Lapenna, L., Bezchinsky, S., Barriga Rubio, R., & García, M. (2025). *Doing the feminist work in AI: Reflections from an AI project in Latin America*. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713681

Feldfeber, I., & Quiroga, Y. (2025). *La anonimización de datos en la justicia*. https://drive.google.com/file/d/1oRAqbXGFiuEp1kM8EkDD4orvOCRSzrmJ/view

Feldfeber, I. (2026). AymurAI o cómo incorporar la IA para los datos de género a nivel judicial. In P. Ricaurte Quijano, M. Zasso & I. Polo (Eds.), *No son utopías: Presentes y futuros de la IA feminista en América Latina y el Caribe* (pp. 172–176). DataGénero.

Case Study 22

HEALTH

Smart Kojin: Generative AI and IoT for Menstrual Hygiene

Authors: Reshma Kulkarni, Deepak Kulkarni, Ajinkya Tamhane, India

Overview

Background & Problem Statement

Smart Kojin is a generative AI-enabled menstrual hygiene and safe disposal ecosystem designed to address persistent gaps in menstrual health education, dignified sanitary waste disposal, and infrastructure monitoring. Women and girls across India and the Global South continue to face stigma, misinformation and inadequate disposal facilities, with direct impacts on health, dignity, education and workforce participation. This integrated design links access to information with access to physical hygiene infrastructure.

The Smart Kojin ecosystem integrates a privacy-first, multilingual (English, Hindi and Marathi) AI-powered mobile application with IoT-enabled sanitary napkin disposal as well as dispenser machines. Through conversational AI (chat and voice), users receive reliable, confidential and expert-validated information on menstruation, hygiene practices and safe disposal. The application also enables users to locate nearby Smart Kojin sanitary napkin disposal machines, as well as nearby UPI-based (Unified Payments Interface) cashless sanitary napkin dispensers, supporting timely and dignified access to both hygiene products and disposal facilities.

Smart Kojin machines are SIM-enabled and transmit real-time operational data to a centralized dashboard. This allows continuous monitoring of machine health, operational status, service requirements and disposal volumes. Aggregated and anonymized data support efficient operations, predictive maintenance and data-informed decision-making by institutions and policymakers. This approach supports proactive maintenance, optimized uptime and evidence-based infrastructure planning.

By combining AI-enabled education with monitored physical infrastructure, Smart Kojin contributes to Sustainable Development Goals (SDGs) 3 (Good Health and Well-Being), 5 (Gender Equality), 6 (Clean Water and Sanitation) and 9 (Industry, Innovation and Infrastructure). The solution aligns with India's Digital Public Infrastructure vision and offers a scalable and replicable model for gender-inclusive sanitation systems.

The AI Solution

Background and Problem Statement

Menstrual health remains a critical but under-addressed public health and gender equality issue. Social stigma, limited access to accurate information and inadequate disposal infrastructure contribute to unsafe hygiene practices, environmental contamination and reduced participation of women and girls in education and work.

Conventional non-AI interventions, such as static awareness campaigns or standalone disposal systems are not well suited for personalization, scalability and real-time intelligence. They are limited in their ability to provide confidential support or generate the actionable insights required for effective sanitation governance.

App interface of Smart Kojin

The Smart Kojin AI-Powered Solution

Solution Description

The Smart Kojin system integrates a generative AI mobile application with IoT-connected sanitary napkin disposal machines. The AI application uses natural language processing to deliver confidential, judgement-free menstrual health information through chat and voice interfaces. It also enables users to locate nearby Smart Kojin disposal units and UPI-based cashless sanitary napkin dispensers. The system is gender-responsive by design, prioritizing privacy, anonymity, multilingual access and ease of use, while ensuring explainability through transparent system outputs.

Smart Kojin offers fully automatic, touch-free and smoke-free sanitary pad incineration solutions based on patented technology, supporting complete combustion, safe disposal and high standards of hygiene and environmental compliance.

Learn more at https://smartkojin.com

GenAI Mobile Application

Why Artificial Intelligence Was Necessary

Menstrual health is a deeply personal and stigmatized issue that requires privacy, contextual understanding and continuous availability of support. AI enables personalized, judgement-free, on-demand guidance; multilingual accessibility across diverse contexts; continuous learning from anonymized usage patterns; and integration of educational insights with infrastructure performance data. Non-AI systems cannot deliver these capabilities at scale, making AI a necessary rather than optional component of the solution.

Pilot Deployment and Field Validation

Developed and operated by SmartEnviro Systems Pvt. Ltd., Smart Kojin has been deployed across schools, residential societies and institutional campuses in Karnataka and Maharashtra since 2023. Key milestones include prototype development, field pilots, IoT dashboard integration and user feedback-driven design refinements.

Infrastructure requirements include SIM-enabled disposal/dispensing machines which can transfer data to servers using mobile connectivity. Real-time data is then available to users via the Smart Kojin app or a web link. The platform integrates physical sanitation infrastructure with a digital AI interface, enabling seamless operation without reliance on legacy sanitation systems.

Pilot Deployment and Testing Sites

Type of Institution	Name of Institution	Location
School & College	Sirsi Lions School & PU College	Sirsi, Karnataka – 5,814
School	Indo Scots Global School	Wagholi, Pune (Maharashtra) – 412,207
Residential Society	Hill View Residency	Kothrud, Pune (Maharashtra) – 411,038
Private Company	SmartEnviro Systems Pvt. Ltd.	Gaud Dara (Maharashtra) – 412,205
Residential School	MES Rani Laxmibai Mulinchi Sainiki Shala & Jr. College	Mulshi, Pune (Maharashtra) – 412,115
School	Symbiosis Primary and Secondary School	Deccan Gymkhana, Pune (Maharashtra) – 411,004

Users & Scale

Primary users include adolescent girls, women students, residents and staff members in schools and housing societies. Institutional administrators and facility managers use the dashboard for monitoring and maintenance oversight. Pilot deployments currently serve multiple institutions, with women and girls comprising over 90 per cent of end users; the remaining users are men and boys who primarily access the platform for awareness, facility management and institutional oversight purposes.

Impact & Outcomes

Measured Gender Impact

Pilot deployments demonstrated improved access to dignified disposal facilities and increased awareness of safe menstrual practices. Schools and residential sites reported higher use of designated disposal units, reducing unsafe disposal in toilets and open waste streams. Confidential AI-based access encouraged information-seeking behavior among adolescent girls without fear of stigma. Facility teams benefited from real-time monitoring, which reduced machine

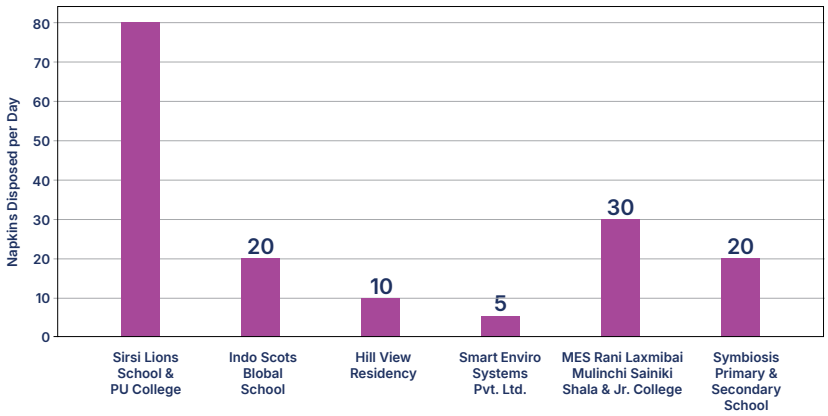
downtime and improved service reliability. The solution supports SDG 5 (Gender Equality), SDG 3 (Good Health and Well-being) and SDG 6 (Clean Water and Sanitation) by improving hygiene outcomes, reducing environmental impact and supporting uninterrupted participation of girls in education.

Objectives of Pilot Deployments

The pilots were conducted to:

- Validate complete, safe and low-emission incineration performance
- Assess usability and acceptance among students and residents
- Monitor machine reliability and safety during continuous operation
- Evaluate IoT-based data transmission and dashboard monitoring
- Understand real-world disposal patterns across diverse environments

Average Sanitary Napkin Disposal – Smart Kojin Pilot Sites



GenAI App Impact (Pilot Phase)

Measured Physical Impact

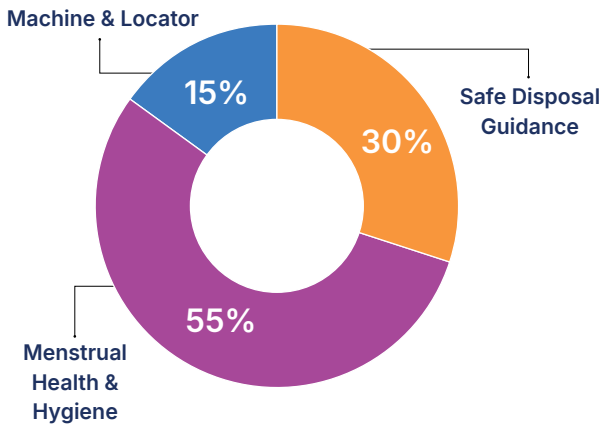
165

sanitary napkins disposed of per day (approximately 4,950 per month/59,400 per year)

120

AI interactions per day 100% anonymous & confidential usage multilingual access (English, Hindi, Marathi)

GenAI Smart Kojin App – User Query Distribution



Key Outcomes

- Consistent high-temperature disposal with minimal emissions
- High user acceptance due to contactless and hygienic operation
- Reliable real-time monitoring of machine health and usage
- Improved maintenance efficiency through predictive alerts
- Demonstrated suitability for schools, residential societies and institutional campuses

Ethics, Governance & Safeguards

Smart Kojin follows a privacy-by-design approach. No personally identifiable or sensitive health data is collected. All AI interactions are anonymized, and usage data is aggregated solely for operational insights. User consent is embedded at the application level. Bias risks are addressed through expert-validated content and inclusive language design. Women users contributed feedback during pilot phases, informing usability and content relevance. SmartEnviro Systems Pvt. Ltd. maintains system oversight with clear accountability for data security, system performance and grievance redressal.

Lessons & Replicability

What Worked

The integration of AI-driven education with physical disposal infrastructure proved effective. Privacy-first design, intuitive interfaces and real-time IoT monitoring emerged as critical success factors. Institutional partnerships supported smooth deployment and user acceptance.

Challenges Encountered

Initial resistance was linked to menstrual stigma, uneven digital literacy and connectivity constraints in some locations. Ongoing user orientation and simplified interfaces helped overcome adoption barriers.

Key Lessons for Replication

Successful replication requires reliable connectivity, institutional commitment and culturally sensitive AI content. Solutions must be adapted to local languages, norms and infrastructure contexts rather than transferred unchanged. The Smart Kojin ecosystem is modular and scalable, enabling deployment across schools, hostels, residential societies, workplaces, industrial campuses and public sanitation infrastructure

The solution is designed for replication across India and the Global South through partnerships with governments, NGOs and institutional stakeholders.

Conclusion & Future Directions

Smart Kojin offers a scalable and sustainable model for inclusive menstrual hygiene management by integrating education, access and data-driven governance.

Smart Kojin demonstrates how responsible AI and IoT can address gendered sanitation challenges at scale. Designed for expansion across public institutions and urban infrastructure, Smart Kojin contributes to inclusive and ethical AI innovation while advancing menstrual dignity across the Global South.

References

Central Pollution Control Board. (2016). *Solid waste management rules, 2016. Ministry of Environment, Forest and Climate Change, Government of India*. https://cpcb.nic.in/uploads/MSW/SWM_2016.pdf

Department of Science & Technology. (n.d.). *Green Dispo – Eco-friendly sanitary napkin incinerators reaching all corners of the country*. Government of India. https://dst.gov.in/green-dispo-eco-friendly-sanitary-napkin-incinerators-reaching-all-corners-country

Centre for Science and Environment. (2022). *Sanitary waste management in India: Challenges and agenda*. https://sbmurban.org/storage/app/media/pdf/sbm_knowledge_center/Sanitary_Waste_Report_FInal_02052022.pdf

Case Study 23

FINANCE

Gender-Bias Audited Credit Scoring for Retail Lending

Author: Maria Angel Margareth Pardede, Indonesia

Overview

Background & Problem Statement

Women across the Global South face persistent structural barriers in accessing formal finance due to gender gaps in income stability, asset ownership and financial documentation. Women-owned enterprises account for approximately one third of the global SME financing gap, exceeding USD 1.5 trillion. Women's account ownership also remains lower, at around 65 per cent, alongside higher rejection rates in both formal and trade finance.

As retail lending increasingly relies on automated credit scoring systems, these disadvantages risk being codified and scaled through algorithmic decision-making. Traditional credit models rely heavily on formal income and collateral, systematically excluding women with informal livelihoods. Prior non-AI interventions, including manual overrides or blanket inclusion policies, lacked scalability, consistency and the capacity to detect and correct bias embedded within decision systems.

Solution Summary

This solution presents a gender-bias-audited AI credit scoring framework piloted in institutional retail lending. The system integrates fairness auditing, explainability and governance controls using predictive analytics and gender-disaggregated data. This approach improves approval parity for women borrowers with informal income profiles and maintains stable portfolio risk performance.

The AI Solution

Solution Description

The AI-enabled credit scoring framework supports retail lending decisions while actively identifying and mitigating gender bias. The system uses predictive analytics, based on machine learning models, to analyse historical lending patterns and estimate credit risk. To assess whether women and men with similar risk profiles are treated differently, gender-disaggregated lending data

are evaluated through bias-auditing models. Fairness tests are applied to identify systematic disparities, while explainable AI tools generate clear, non-technical reason codes for each recommendation. These outputs enable credit officers to understand, review and challenge model recommendations, ensuring that the system functions as decision support rather than as a fully automated approval mechanism.

Deployment & Implementation

Stage of Deployment

The framework was piloted in Indonesia within an institutional retail lending context. Development and deployment were led by an internal analytics team working in collaboration with business units, risk governance and compliance functions. The system was integrated into existing credit assessment workflows and relied on internal application data, transaction histories and alternative income indicators. No additional customer-facing infrastructure was required. National institutions contributed through regulatory guidance, data protection standards and alignment with responsible AI and financial inclusion objectives.

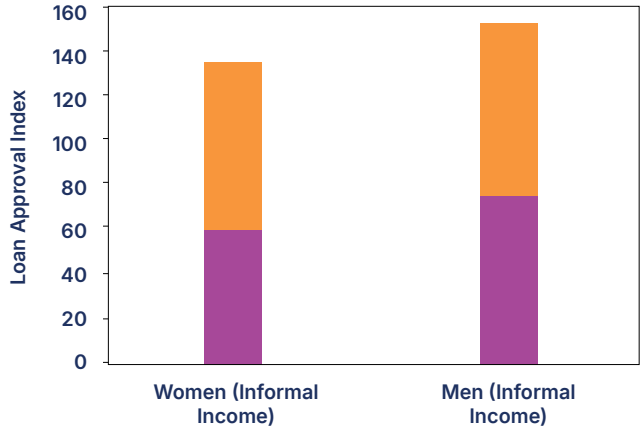
Users & Scale

Primary users were credit analysts and risk officers, while indirect beneficiaries were retail loan applicants, particularly women borrowers with informal income profiles. During the pilot, the system supported loan assessments within a single institutional environment. Women accessed its benefits through standard loan application channels, with all decisions subject to mandatory human review.

Impact & Outcomes

Measured Gender Impact

During the pilot, the framework was applied to several hundred retail loan assessments within an institutional lending context in which approximately half of applicants were women. Compared with pre-intervention outcomes, approval parity improved for women borrowers with informal income profiles, while default rates and overall portfolio risk remained stable. The intervention supported women's economic participation and advanced SDG 5 (Gender Equality) and SDG 8 (Decent Work and Economic Growth) by addressing exclusion linked to informal livelihoods and limited collateral. Figure 1 illustrates indexed changes in loan approval outcomes for women and men with informal income profiles before and after the introduction of gender-bias auditing.



Improvement in Loan Approval Parity Following Gender-Bias Auditing

Ethics, Governance & Safeguards

Ethical safeguards and governance mechanisms were embedded by design. The system used anonymized and access-controlled data in accordance with local privacy regulations and operated within pre-existing institutional infrastructure. Bias mitigation was implemented through routine gender-disaggregated monitoring and fairness audits during both model validation and ongoing use. Explainable outputs enabled credit officers to review, interpret and challenge model recommendations. Internal risk governance and compliance functions reviewed and monitored system performance through clearly defined escalation processes. Potential unintended effects, including over-reliance on automated scoring, were mitigated through mandatory human review and continuous governance oversight.

Lessons & Replicability

What Worked

Embedding gender-bias auditing within existing credit workflows proved more effective than introducing standalone systems. The combination of predictive analytics, explainability and human review supported institutional trust, regulatory alignment and operational adoption. Strong internal governance and access to relevant data were critical enablers for timely testing and refinement.

Challenges Encountered

Balancing fairness objectives with established risk thresholds required iterative calibration and sustained internal dialogue. Initial scepticism towards algorithmic adjustments underscored the importance of change management and transparency. Data gaps related to informal income required careful feature selection to avoid introducing new forms of exclusion.

Key Lessons for Replication

Successful replication requires adapting fairness metrics to local labour markets, data environments and regulatory contexts rather than replicating models directly. Early engagement with regulators, robust governance structures and continuous bias monitoring are essential. Institutions should prioritize explainability and human oversight from the outset to sustain trust and accountability.

Conclusion & Future Directions

This case demonstrates that gender bias in AI-enabled credit decisions is neither inevitable nor acceptable. Integrating fairness, explainability and governance into lending systems provides a scalable pathway for institutions across the Global South, enabling geographic expansion, diversified product offerings and more inclusive AI governance.

References

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. https://fairmlbook.org

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). *Explainable AI in credit risk management. Computational Economics, 57*(1), 203–216. https://doi.org/10.1007/s10614-020-10042-0

International Finance Corporation. (2023). *MSME finance gap: Assessment of the shortfalls and opportunities in financing micro, small and medium enterprises in emerging markets*. World Bank Group. https://www.ifc.org/insights-reports/msme-finance-gap

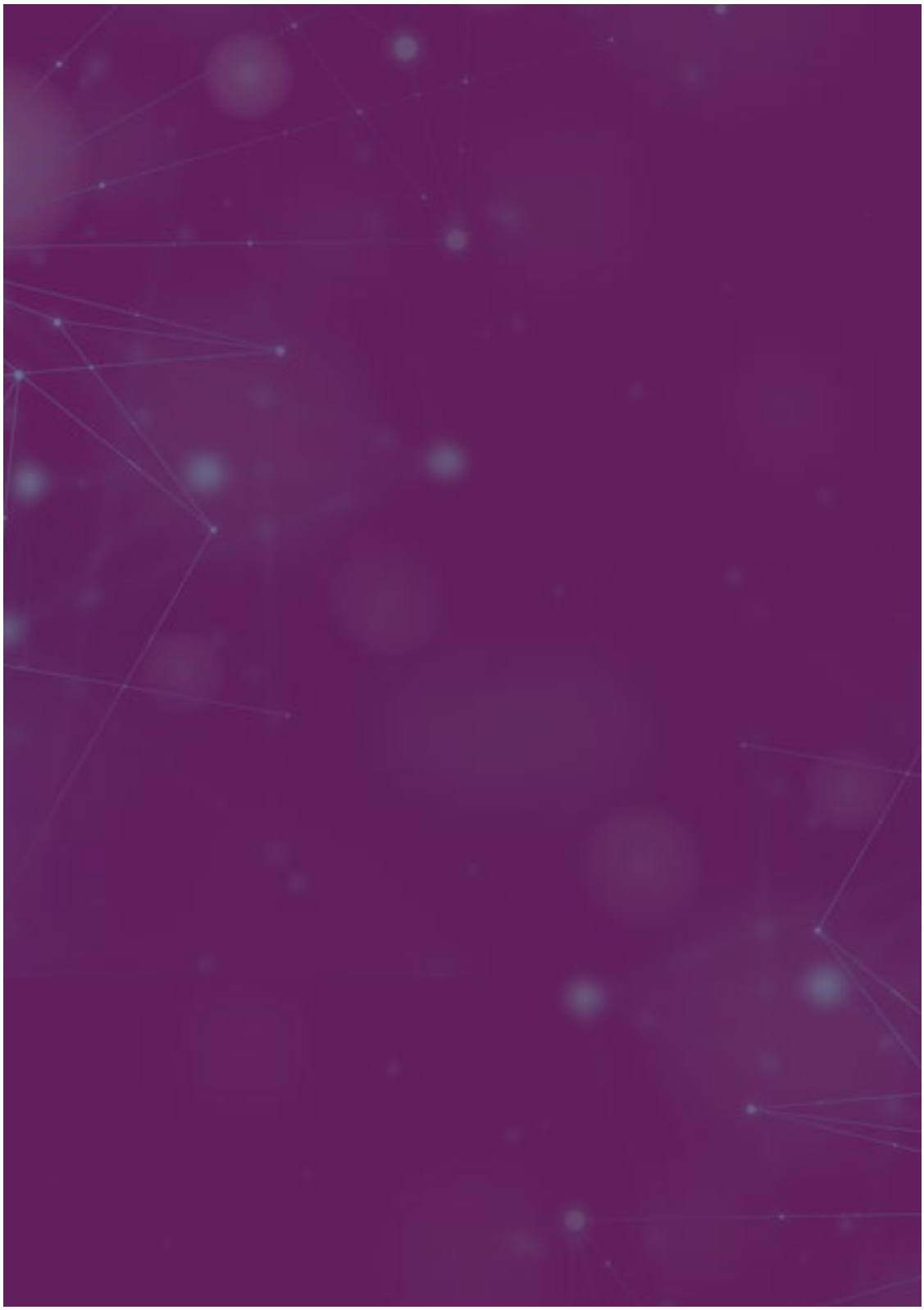
International Monetary Fund. (2022). *Financial access survey: Trends in gender gaps in financial inclusion*. IMF. https://www.imf.org/en/Data/Statistics/Financial-Access-Survey

Organisation for Economic Co-operation and Development. (2021). *Trustworthy artificial intelligence in finance*. OECD Publishing. https://www.oecd.org/finance/trustworthy-artificial-intelligence-in-finance.htm

UN Women. (2023). *Gender equality and artificial intelligence: Addressing bias and advancing inclusion*. https://www.unwomen.org/en/digital-library/publications/2023/gender-equality-and-ai

World Bank Group. (2021). *Advancing women's financial inclusion: Addressing structural barriers*. World Bank. https://www.worldbank.org/en/topic/financialinclusion/publication/advancing-womens-financial-inclusion

World Bank. (2022). *Global Findex database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19*. https://www.worldbank.org/en/publication/globalfindex



Disclaimer

The case studies included in this compendium have been evaluated based on information submitted by the respective authors and participating organizations. Responsibility for the accuracy of data, metrics, and representations rests solely with the submitting authors. The evaluation committee and partner institutions shall not be held liable for any discrepancies, omissions, or subsequent changes in the information provided.

Partners:

IndiaAI Mission, Ministry of Electronics and Information Technology (MeitY), and UN Women.

We thank the Ministry of Women and Child Development for its support.