# Guidelines for Whom? Rethinking AI Ethics in Resource-Constrained Migration Services

Anonymous ACL submission

## Abstract

Responsible AI principles have had limited influence on practice in humanitarian settings. A growing body of published guidelines now governs AI and data use in these contexts, responding to documented risks including surveillance, data misuse, and discriminatory outcomes affecting refugee populations. For high-risk applications such as biometric identification and asylum adjudication, such guidelines address genuine and serious concerns. Many differentiate risk tiers in principle, yet the compliance expectations they establish—staff capacity, technical infrastructure, formal evaluation—reflect the organizational contexts in which such guidelines are most often developed. Across the humanitarian sector, however, deploying organizations vary considerably in resources, and many of the nonprofits providing frontline services to refugees operate with limited administrative capacity. When compliance requirements exceed what these organizations can reasonably meet, formal AI adoption stalls while informal adoption proceeds without oversight or recourse. Current guidelines also tend to treat non-adoption as a neutral default, without accounting for the service gaps that follow when AI-assisted language access is unavailable. Drawing on collaboration with refugee-serving practitioners, we show that this gap between governance design and organizational reality has real consequences for the people these guidelines are meant to protect. We argue that evaluating AI guidelines requires the same realist logic that evaluation research has long applied to social programs: not "does this guideline exist?" but "for which deployers, under what conditions, and does it produce its intended protective outcomes?"

## 1 Introduction

Migration service organizations need AI-assisted language access but cannot obtain it through formal channels. The issue is not reluctance or ignorance. Jobin et al. (2019) identified 84 AI ethics documents worldwide; Corrêa et al. (2023) expanded the count to 200. The number has continued to grow since, and an organization that attempts to formally adopt AI translation for refugee services now encounters governance requirements from multiple domains simultaneously: AI-specific regulation such as the EU AI Act (European Parliament and Council of the European Union, 2024), sector-wide data responsibility frameworks such as the IASC Operational Guidance (Inter-Agency Standing Committee, 2023), and broader digital development principles such as the Principles for Digital Development (Principles for Digital Development, 2024). The risks these documents respond to are real: biometric surveillance of refugee populations (Kreutzer et al., 2025), data processing failures affecting crisis-affected populations (Kreutzer et al., 2025), and the concentration of AI procurement in the Global North (Png, 2022). For high-risk applications, including asylum adjudication, predictive profiling, and biometric identification, strict oversight is necessary. But most of these guidelines were written for government agencies and large technology companies. They assume compliance staff, legal counsel, bias auditing infrastructure, and dedicated evaluation budgets. Community-based nonprofits serving refugee populations were not the intended audience and, in most cases, were not part of the development process.

The gap between what guidelines require and what organizations can do has widened as resources have contracted. By mid-2025, 71% of refugees

were hosted by low- and middle-income countries (UNHCR, 2025a). The United States Agency for International Development (USAID) was officially shut down in July 2025; roughly 83% of its programs were cancelled, removing an estimated $2.3 billion in migration-related funding (Huang et al., 2025). The OECD recorded a 9% decline in official development assistance in 2024 alone (OECD, 2025). The Danish Refugee Council estimated that U.S. aid cuts could contribute to an additional 3.95 to 7.85 million displaced people in 2025 (Vallentine et al., 2025). For organizations that work with these populations, these cuts reduced already-insufficient budgets for professional interpretation, increasing pressure to find lower-cost alternatives for multilingual service delivery. UNHCR's AI strategy prioritizes multilingual access to services (UNHCR, 2025b). UN agencies are co-designing translation tools with refugee communities for low-resource languages. The need is acute: many refugee-serving organizations work with populations speaking dozens of languages for which professional interpreters are unavailable at any price point.

Whether these guidelines produce their intended outcomes in practice is a separate question from whether they exist. Mittelstadt (2019) argued that principles alone cannot ensure ethical AI practice. McNamara et al. (2018) tested this: presenting ACM's code of ethics to software engineers did not change their decisions. The gap between stated principles and on-the-ground behavior is not specific to AI. But in humanitarian settings, it takes a particular form. Pizzi et al. (2020) found that most humanitarian AI codes of ethics provide no answer about who bears the cost when principles are violated. Bhatnagar et al. (2025) found that practitioners identified a persistent disconnect between high-level governance policies and the tools needed for design, monitoring, and evaluation. Munn (2023) made a broader version of the same point: AI ethics, as currently structured, has limited influence on applied development. Hagendorff (2020) compared 22 AI ethics guidelines and found that they lack enforcement mechanisms, with their values easily overwritten by economic incentives. Heymans and Heyman (2024) argued that guidelines tend to encode the priorities of the powerful institutions that produce them rather than reflecting the needs of the broader populations they claim to protect.

We do not argue against these guidelines. We argue that the current approach, which applies the same requirements to all AI uses and all deployers regardless of risk or capacity, produces unintended consequences. When formal adoption pathways are inaccessible, what follows is not non-adoption, but informal adoption without institutional oversight. Drawing on collaboration with refugee-serving practitioners, we show how this plays out in practice and suggest that evaluating AI guidelines requires the question realist evaluation (Pawson and Tilley, 1997) has long asked of social programs: whether formal guidelines produce their intended protective outcomes across the range of organizations expected to follow them.

## 2 What Current Guidelines Assume

Three assumptions run through existing guidelines. Each is reasonable in the context where the guidelines were written. Each becomes a problem when applied to resource-constrained settings.

**Assumption 1: Guidelines address all deployers equally.** Jobin et al. (2019) found that the majority of AI ethics documents originate from institutions in North America and Europe, with African and South American countries not independently represented in the corpus. The deployers these documents assume are government agencies and large technology companies. The compliance requirements they specify, including dedicated ethics review, bias auditing infrastructure, and ongoing monitoring, reflect what those kinds of organizations can reasonably do.

A community-based nonprofit that adopts AI translation for refugee services encounters these requirements not through a single document but through the accumulation of multiple governance layers. AI-specific regulation, such as the EU AI Act (European Parliament and Council of the European Union, 2024), classifies AI in migration and asylum as high-risk. Sector-wide frameworks such as the IASC Operational Guidance on Data Responsibility (Inter-Agency Standing Committee, 2023) require data impact assessments, information-sharing protocols, and designated staff roles for any data-processing activity that AI use triggers. Broader digital development principles, such as the Principles for Digital Development (Principles for Digital Development, 2024), endorsed by over 300 organizations, ask endorsers to integrate privacy, security, and open standards into organizational policies. Each of these originates from a different governance domain. None was

written for an organization with three to five staff members and no dedicated compliance or technology personnel. But all apply simultaneously to a nonprofit using AI to translate services for refugee populations. The cumulative effect is a set of requirements that no single small organization can meet, even when each individual requirement is defensible on its own terms.

**Assumption 2: Non-adoption is always the safer choice.** In many contexts, this is reasonable. A government agency that delays deploying a predictive model until it has been audited for bias has chosen a defensible path. In migration services, the calculus is different. Practitioners were already relying on machine translation tools like Google Translate before generative AI existed, and no one required them to conduct bias audits or data protection assessments to do so. Generative AI has raised the stakes on both sides: translation quality has improved enough to handle contextual communication and some low-resource languages that earlier tools could not, while humanitarian funding cuts have eliminated interpretation budgets that were already insufficient. The choice practitioners describe to us is not between AI and some adequate human alternative. It is between using generative AI translation and having no access to language for the populations they serve. When formal adoption pathways are inaccessible, what typically follows is not non-adoption but informal adoption without institutional oversight.

**Assumption 3: "AI in humanitarian contexts" is a single risk category.** Current guidelines do not differentiate among AI used for border biometric surveillance, AI used to triage asylum claims, and AI used to translate an informational pamphlet about local services. Memon et al. (2024) documented the growing use of AI tools across European asylum systems, from language detection to case matching, each with distinct risks depending on the application. The EU AI Act classifies AI systems used in migration, asylum, and border management as high-risk under Annex III, Section 7, covering AI used as polygraphs, systems assessing migration-related risks, and systems processing asylum and visa applications (European Parliament and Council of the European Union, 2024). For these applications, the classification is appropriate. But the Act also builds in proportionality for smaller deployers: Article 62 specifies simplified technical documentation, simplified quality management systems, and priority access to regulatory sandboxes for SMEs and startups. The European Commission has proposed extending these provisions to small mid-cap companies (European Commission, 2025). This kind of differentiation, based on both use-case risk and deployer capacity, has not been adopted by most AI guidelines applicable to humanitarian contexts. The EU AI Act's proportionality provisions demonstrate that graduated requirements are technically and legally possible. Their absence in humanitarian AI governance is a design choice, not a necessity.

Realist evaluation (Pawson and Tilley, 1997) offers a way to see why these assumptions matter. The central question in realist evaluation is not "does this program work?" but "what works, for whom, in what contexts, and how?" Applied to AI guidelines: the same guideline produces formal adoption and protective oversight when deployed in an institution with compliance capacity, but produces avoidance or informal workarounds when deployed in a resource-constrained nonprofit without that capacity. Current guideline evaluation does not ask this question. It asks whether guidelines exist and whether they are inclusive in scope. It does not ask whether the same guideline produces different outcomes depending on the capacity of the organization expected to implement it.

## 3 What Happens When Uniform Standards Meet Uneven Capacity

The observations in this section are informed by the authors' ongoing participatory research with refugee-serving organizations in South Korea, including field notes. South Korea has recognized over 700 refugees and granted humanitarian status or permits to several thousand more in recent years, but the infrastructure to serve these populations has not kept pace. Community-based organizations that provide legal aid, case management, and social services to refugees operate with minimal staff, high turnover rates, and limited funding. Professional interpretation for the range of languages spoken by refugee populations is expensive when available and often unavailable entirely. The following case illustrates what AI-assisted language access makes possible when it is available.

> At a community forum serving refugee populations, organizers had previously relied on sequential human interpretation, limiting sessions to one or two languages and requiring participants in other language groups to

wait or go without. When AI-powered synchronous translation was deployed across multiple languages, practitioners reported that participants who had previously been excluded from multilingual sessions were able to follow along in real time. The experience made staff more willing to consider AI translation for other service areas, including intake interviews and benefit navigation.

This is not unique to migration services. Two recent organizational surveys, neither peer-reviewed but both drawing on large samples, suggest the pattern is widespread. A survey of over 1,300 nonprofit professionals found that 76% of organizations in the U.S. had no AI strategy, 80% had no acceptable use policy, and 43% relied on a single staff member for all IT and AI decisions (TechSoup and Tapp Network, 2025). A larger survey of 2,539 humanitarian professionals across 144 countries found that 70% used AI tools daily or weekly, but fewer than one in four organizations had formal AI policies, and only 8% reported organization-wide AI integration (Johnson et al., 2025). The concern practitioners raise is not that ethical standards are too high but that uniform requirements, without differentiation by deployer capacity, leave organizations unable to adopt formally while the need for AI-assisted language access remains. The same organization's experience illustrates what happens when practitioners attempt to adopt these tools through formal channels.

When practitioners investigated formal adoption of the AI translation tools, the requirements they encountered included data protection impact assessments, informed consent protocols for AI-mediated communication, bias auditing procedures, and community consultation. Each requirement is defensible. But the organization in question had recently lost a primary funding source, reducing an already minimal staff and accelerating turnover that had left institutional knowledge thin. There was no compliance officer, no legal counsel, and no dedicated technology staff. The person responsible for evaluating the adoption requirements was typically the same person managing intake, coordinating volunteers, and answering client calls, often someone who had been in the role for only a few months. After reviewing the burden, the organization gave up on formal adoption. The tools continued to be used, but at the discretion of individual staff members, without organizational policy, without documentation, and without recourse if something went wrong.

When formal adoption is impractical, informal use continues without institutional support. Among the nonprofits surveyed by TechSoup (TechSoup and Tapp Network, 2025), only 7% had adopted AI through formal organizational processes; 42% reported that staff had independently learned AI on their own. The global humanitarian survey found the same pattern: individual workers adopting ChatGPT and Google Translate at their own discretion, outside institutional channels (Johnson et al., 2025). In humanitarian settings, where the data involved may include asylum claims, medical histories, or personal identification, the risks of such informal adoption are high, and the oversight is thin. The risks of informal adoption are not hypothetical. Documented cases from neural machine translation tools, which share the same underlying architecture as current generative AI systems, show what happens when AI-mediated translation enters legal settings without institutional quality review.

Respond Crisis Translation documented a case in which an automated translation tool swapped first-person pronouns in a Pashto-speaking refugee's asylum statement, changing "I" to "we." The resulting inconsistency led a judge to deny the claim (Rogin and Corkery, 2023). In a separate case reported through the same investigation, a domestic violence survivor described her abuser as "mi jefe," a common colloquialism for one's father. The translation tool rendered it literally as "my boss," and the asylum application was initially denied (Bhuiyan, 2023).

Pronoun substitution, literal rendering of idiomatic expressions, and lack of cultural and dialectal awareness are well-documented failure modes in neural language models. The practitioners we work with were already using Google Translate routinely before generative AI entered the conversation, and generative AI has not replaced these tools so much as merged with them: Google integrated its Gemini model directly into Google Translate in December 2025, redesigning the service to handle idioms, slang, and contextual expressions through a large language model rather than a conventional neural translation pipeline (Google, 2025). ChatGPT and other generative AI tools have entered these workflows in parallel, used by individual staff for drafting intake documents, translating longer case narratives, and communicating in low-resource languages that earlier tools handled poorly (Deck, 2023; Johnson et al., 2025). The boundary between "machine translation" and "generative AI" is no longer a meaningful distinction in practice. What has not changed is the condition under which these tools are used: informally, without organizational policy, quality review, or documentation of errors. Guidelines designed to prevent exactly this kind of harm may be producing the conditions for

it by making formal adoption impractical for the organizations that most need it.

The relevant comparison is not between AI-assisted translation and some adequate human alternative, nor between AI-assisted translation and no translation at all. It is between formal adoption with appropriate safeguards and informal adoption with none. Dalal et al. (2024) argued that participatory AI structures do not account for the barriers that prevent marginalized communities from realizing the benefits of AI. We extend this to the structures of ethical AI evaluation: when guidelines assign evaluation tasks uniformly to all deployers, well-resourced institutions evaluate and comply. Under-resourced organizations either skip the evaluation or skip the tool. Neither outcome is what the guidelines intended.

## 4 Toward Differentiated Standards

We argue that three things are missing from the current guidelines. The first is use-case risk differentiation. AI-assisted translation for an informational pamphlet is not the same as AI used to process asylum claims or to detect dialects to infer country of origin. For applications involving asylum adjudication, biometric identification, or border surveillance, strict oversight should remain in place. But within the broad category of "AI in humanitarian contexts," applications vary widely in what can go wrong and for whom. The EU AI Act provides one reference point: it classifies AI in migration as high-risk while also allowing proportionate requirements for smaller deployers.

The second is attention to deployer capacity. The EU AI Act offers simplified documentation requirements, reduced fees, and sandbox access for SMEs on the principle that compliance obligations should reflect what an organization can reasonably do. Rather than requiring every deployer to independently audit, document, and evaluate its AI tools, guidelines could specify which responsibilities can reasonably be carried by organizations at different capacity levels. Some recent initiatives point in this direction: the Decoded Futures program in New York City has supported over 1,100 organizations through AI capacity-building (Tech:NYC, 2025), and the IRC's Signpost AI initiative plans to extend responsible AI infrastructure to smaller organizations (International Rescue Committee, 2024). But most of these are U.S.-based and English-language focused, and none have been incorporated into the formal guidelines that govern AI use in humanitarian settings.

The third is deployment-level guidance for practitioners. Even when risk levels are differentiated and organizational capacity is acknowledged, individual practitioners still face real-time judgments about whether and how to use an AI tool in a specific service interaction—judgments that no policy document fully anticipates. Recent scholarship has identified this as a structural limitation of principles-based AI ethics: high-level guidelines do not translate into the micro-level decisions that arise during deployment (Hagendorff, 2020; Mittelstadt, 2019; Munn, 2023). A caseworker deciding whether to use an AI translation tool with a client who speaks a low-resource dialect, or a frontline worker unsure whether a chatbot response is accurate enough to act on, is not well served by guidelines written for compliance officers. What is missing is deployment-level guidance: practical tools, heuristics, or escalation protocols that help practitioners make safe choices under uncertainty, without requiring organizational capacity that most refugee-serving nonprofits do not have.

Realist evaluation offers a way to structure what is missing. Pawson and Tilley (1997) argued that evaluation should produce context-mechanism-outcome configurations: statements about how a program activates specific mechanisms among specific actors in specific conditions. Applied here, this means asking three questions that current guidelines do not. First, does the guideline differentiate requirements by the risk level of the specific AI application? Second, does the guideline specify the organizational capacity needed to implement it? Third, does the guideline provide deployment-level guidance for practitioners who encounter novel situations during service delivery, and does it account for what follows when organizations that need AI-assisted services cannot meet the conditions for formal adoption?

## 5 Conclusion

Ethical evaluation of AI in migration services sits between two sets of concerns. Refugee populations face documented risks from AI systems: surveillance, data misuse, and discriminatory outcomes. They also face service gaps when AI-assisted language access is unavailable: inability to communicate with providers, delays in legal processes, and exclusion from safety-relevant information. Cur-

rent evaluation norms are designed to address the first set of issues. They say little about the second. Applying the same requirements to all use cases and all deployers produces a specific outcome: organizations with resources comply formally, organizations without resources adopt informally, and the people those guidelines were written to protect end up less protected than they would be under a differentiated approach.

## References


Tigmanshu Bhatnagar, Maarya Omar, and Davor Orlic. 2025. Bridging AI and humanitarianism: An HCI-informed framework for responsible AI adoption. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. ACM. Accessed: 2026-3-4.

Johana Bhuiyan. 2023. Lost in AI translation: growing reliance on language apps jeopardizes some asylum applications. https://www.theguardian.com/us-news/2023/sep/07/asylum-seekers-ai-translation-apps. Accessed: 2026-3-3.

Nicholas Kluge Corrêa, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, Rodrigo Mambrini, Luiza Galvão, Edmund Terem, and Nythamar de Oliveira. 2023. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns (N. Y.)*, 4(10):100857. Accessed: 2026-3-4.

Samantha Dalal, Siobhan Mackenzie Hall, and Nari Johnson. 2024. Provocation: Who benefits from "inclusion" in generative AI? *arXiv [cs.CY]*. Accessed: 2026-3-4.

Andrew Deck. 2023. AI translation is jeopardizing afghan asylum claims. https://restofworld.org/2023/ai-translation-errors-afghan-refugees-asylum/. Accessed: 2026-2-25.

European Commission. 2025. Proposal for a regulation of the european parliament and of the council amending regulation (EU) 2024/1689 on the digital omnibus on AI regulation. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Google. 2025. Bringing state-of-the-art Gemini translation capabilities to Google Translate. https://blog.google/products/search/gemini-capabilities-translation-upgrades/. Accessed: 2026-3-5.

Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.*, 30(1):99–120.

Frederic Heymans and Rob Heyman. 2024. Identifying stakeholder motivations in normative AI governance: a systematic literature review for research guidance. *Data Policy*, 6:e58.

Lawrence Huang, Samuel Davidoff-Gore, and Susan Fratzke. 2025. Can innovation help blunt the impact of foreign aid cuts on migration management programs? https://www.migrationpolicy.org/news/foreign-aid-cuts-migration-management. Accessed: 2026-2-24.

Inter-Agency Standing Committee. 2023. IASC operational guidance on data responsibility in humanitarian action. https://interagencystandingcommittee.org/operational-response/iasc-operational-guidance-data-responsibility-humanitarian-action. Accessed: 2026-3-3.

International Rescue Committee. 2024. The impact of signpost: Bridging the information gap for people in crisis. https://www.rescue.org/article/impact-signpost-bridging-information-gap-people-crisis. Accessed: 2026-3-3.

Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.*, 1(9):389–399.

Madigan Johnson, Ka Man Parkinson, and Lucy Hall. 2025. Initial insights report: How are humanitarians using artificial intelligence in 2025? https://www.humanitarianleadershipacademy.org/resources/initial-insights-report-how-are-humanitarians-using-artificial-intelligence-in-2025/. Accessed: 2026-3-2.

Tino Kreutzer, James Orbinski, Lora Appel, Aijun An, Jerome Marston, Ella Boone, and Patrick Vinck. 2025. Ethical implications related to processing of personal data and artificial intelligence in humanitarian crises: a scoping review. *BMC Med. Ethics*, 26(1):49.

Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. Does ACM's code of ethics change ethical decision making in software development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, New York, NY, USA. ACM.

Amina Memon, Zoe Given-Wilson, Derya Ozkul, Karen Mcgregor Richmond, Julia Muraszkiewicz, Ella Weldon, and Cornelius Katona. 2024. Artificial intelligence (AI) in the asylum system. *Med. Sci. Law*, 64(2):87–90.

Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.*, 1:501–507. Accessed: 2026-3-4.

Luke Munn. 2023. The uselessness of AI ethics. *AI Ethics*, 3(3):869–877.

OECD. 2025. Cuts in official development assistance. https://www.oecd.org/en/publications/cuts-in-official-development-assistance_8c530629-en/full-report.html. Accessed: 2026-2-24.

Ray Pawson and Nick Tilley. 1997. *Realistic Evaluation*. SAGE Publications, Thousand Oaks, CA.

Michael Pizzi, Mila Romanoff, and Tim Engelhardt. 2020. AI for humanitarian action: Human rights and ethics. *Int. Rev. Red Cross*, 102(913):145–180.

Marie-Therese Png. 2022. At the tensions of south and north: Critical roles of global south stakeholders in AI governance. In *2022 ACM Conference on Fairness Accountability and Transparency*, pages 1434–1445, New York, NY, USA. ACM.

Principles for Digital Development. 2024. Principles for digital development. https://digitalprinciples.org/. Accessed: 2026-3-3.

Ali Rogin and Andrew Corkery. 2023. How language translation technology is jeopardizing afghan asylum-seekers. https://www.pbs.org/newshour/show/how-language-translation-technology-is-jeopardizing-afghan-asylum-seekers. Accessed: 2026-2-24.

Tech:NYC. 2025. 2025 decoded futures annual report. https://www.annualreportdecodedfutures.nyc/. Accessed: 2026-2-25.

TechSoup and Tapp Network. 2025. Benchmark report: The state of AI in nonprofits 2025. https://page.techsoup.org/ai-benchmark-report-2025. Accessed: 2026-2-25.

UNHCR. 2025a. Mid-year trends. https://www.unhcr.org/us/publications/mid-year-trends. Accessed: 2026-3-3.

UNHCR. 2025b. UNHCR AI approach. https://www.unhcr.org/digitalstrategy/ai-approach/. Accessed: 2026-2-24.

Jennifer Vallentine, Bram Frouws, and Roberto Forin. 2025. Mixed migration review 2025. https://mixedmigration.org/wp-content/uploads/2025/11/MMR_2025_Online.pdf. Accessed: 2026-2-24.