

Lost in Translation: How Universal Ethical Values Fail to Translate Across Global Contexts

Ozioma C. Oguine¹, Munachimso B. Oguine², Cesar Cervera¹, Jenny Yang³, Pooja Voladoddi³, Mario Rodriguez³, Saif Eddin Bani Malhem³, Karla Badillo-Urquiola¹, Daricia Wilkinson²

¹University of Notre Dame

²Arizona State University

³Independent Researcher

ooguine@nd.edu, mooguine@asu.edu, ccervera@nd.edu, jennyhyang03@gmail.com,
pooja.voladoddi@gmail.com, mariorodtcun@gmail.com, saif.banimalhem@insead.edu,
kbadill3@nd.edu, daricia.wilkinson@asu.edu

Abstract

AI ethics frameworks treat values such as fairness, transparency, and accountability as universal and uniformly operationalizable across contexts. We examined how 14 experts across 10 countries made sense of AI in practice, reinterpreted core values, and envisioned governance alternatives. We found that AI deployment is characterized by structurally unequal conditions, marked by infrastructural constraints, extractive practices, and a “mystification” of technology, which fundamentally shape perceptions of risks and opportunities. Our findings reveal that experts reinterpret values to fit local moral logics: privacy as collective and relational rather than individual; transparency as trust-building accountability rather than technical disclosure; and fairness as equity in access and representation rather than parity in outcomes. We identify these as translation gaps between encoded global frameworks and situated local practices. Finally, we propose pathways toward plural governance that redistributes epistemic authority and treats ethical negotiation as an ongoing, context-sensitive process rather than a settled technical standard.

Introduction

As artificial intelligence (AI) systems are increasingly deployed across diverse global contexts (Statista 2025), questions of ethics and governance have become central to both research and policy (Stanford Institute 2025; OECD 2026; Oguine et al. 2026). In response, a growing number of frameworks spanning institutions such as the OECD, UNESCO, and major technology companies have articulated core values including fairness, accountability, transparency, privacy, and safety (OECD 2026; UNESCO 2021; High-Level Expert Group 2019). These frameworks are often presented as broadly applicable guidelines for responsible AI development and deployment. However, despite their widespread adoption, they are largely grounded in normative assumptions that may not translate seamlessly across cultural, social, and geopolitical contexts (Sambasivan and Arora 2021; Varshney 2024). Emerging scholarship has begun to critique the universalist framing of AI ethics, ar-

guing that such approaches risk overlooking how values are interpreted, prioritized, and operationalized differently across communities (Mohamed, Png, and Isaac 2020; Mhlambi and Tiribelli 2023). In particular, perspectives from non-dominant regions, including the Global South and Indigenous communities, highlight how AI systems are entangled with historical inequities, cultural norms, and asymmetries in global power (Nemorin and Bonami 2026). In these contexts, AI is not only a technical system but also a sociotechnical artifact that can reinforce or challenge existing structures of authority, representation, and autonomy (Mohamed, Png, and Isaac 2020).

Despite these critiques, there remains limited empirical understanding of how experts¹ working across diverse cultural contexts, engage with dominant AI ethics frameworks in practice. Much of the existing literature has focused on defining ethical values or evaluating technical implementations (Jobin, Ienca, and Vayena 2019; Morley et al. 2021), with comparatively less attention to how these values are interpreted and adapted by those responsible for designing, governing, or critiquing AI systems in different regions. As a result, key questions remain about how AI ethics is interpreted across contexts and how local actors negotiate ethical values to align with their communities’ needs. While existing critiques have focused on the abstraction of ethical principles or the challenges of implementing them in practice, considerably less attention has been paid to how these principles acquire meaning as they move across cultural, social, and institutional contexts. We argue that understanding AI ethics therefore requires attention not only to principles or implementation, but also to translation and interpretation in situated contexts. Specifically, we ask: **RQ1:** How do experts describe and make sense of AI deployment within their respective cultural contexts? **RQ2:** How do experts interpret and negotiate the ethical values embedded in AI systems? **RQ3:** How do they envision culturally grounded approaches to the design and governance of AI systems?

¹In this paper, We define “experts” in AI as individuals with both substantive engagement in AI (e.g., research, policy, industry, or advocacy) and deep familiarity with the cultural contexts in which AI is designed, deployed, or governed.

To answer these research questions, we conducted a qualitative interview study with 14 experts across 10 countries in Africa, Asia, Latin America, Oceania, Europe, North America, and the Caribbean. Our findings show that infrastructural constraints, competing narratives, and existing power dynamics shape how experts experience AI, perceived risks and opportunities. These experiences, in turn, inform how participants reinterpret core ethical values such as fairness, privacy, and transparency as relational, context-dependent, and embedded within broader social and structural conditions. Finally, participants identify key barriers in current governance approaches, alongside pathways that emphasize local knowledge, participatory processes, and more distributed forms of authority.

This work makes three key contributions. This work makes three key contributions. First, we provide empirical evidence showing that the central challenge in AI ethics lies not only in defining universal values, but in how those values are interpreted across contexts. Second, we introduce the AI Ethics Translation Model, a conceptual framework that explains how universal ethical principles acquire situated meanings through local interpretation, producing translation gaps that shape governance outcomes. Third, we contribute to ongoing discussions in AI governance by identifying pathways toward more plural and power-aware approaches to AI governance, emphasizing the role of local knowledge, participatory processes, and more distributed forms of authority in shaping context-sensitive AI systems.

Related Work

The past decade has witnessed a proliferation of high-level AI ethics frameworks issued by international organizations and governmental bodies. The OECD (2019)'s recommendation on Artificial Intelligence, endorsed by 47 countries, established foundational commitments to human-centered values, including transparency, fairness, and accountability. UNESCO (2021) followed with its recommendation on the Ethics of AI, emphasizing human rights, dignity, and environmental sustainability as universal imperatives. The European Union's Ethics Guidelines for Trustworthy AI operationalized these principles through seven key requirements, including human agency, technical robustness, and diversity (High-Level Expert Group 2019). These frameworks share the underlying assumption that ethical values for AI can be articulated at a sufficient level of abstraction to apply across diverse sociopolitical and cultural contexts. However, scholars have increasingly critiqued this universalist orientation. Greene, Hoffmann, and Stark (2019) argue that such frameworks engage in ethical abstraction, distilling complex moral deliberation into decontextualized principles that fail to account for local power dynamics and value pluralism. Similarly, Jobin, Ienca, and Vayena (2019)'s comprehensive survey of AI ethics guidelines reveals how fairness metrics predominantly encode Western liberal assumptions about individual autonomy and procedural justice, rendering them inadequate for contexts prioritizing communal harmony or relational ethics. The abstraction inherent in these frameworks, while enabling broad international consensus, risks producing surface-level compliance that masks substantive

value misalignment, what Schultz, Conti, and Seele (2025) calls "ethics washing". These critiques suggest that the current architecture of global AI ethics may perpetuate epistemic injustice by privileging certain ways of knowing and valuing while marginalizing others.

Beyond critiques of universalism, researchers within the AIES community and beyond have begun articulating alternative epistemologies for AI ethics grounded in specific cultural traditions (Varshney 2024; Chi, Lurie, and Mulligan 2021; Schiff et al. 2020; Mhlambi and Tiribelli 2023). For instance, research on African philosophies in computing has highlighted how Ubuntu ethics, which emphasize interconnectedness and communal flourishing, offer distinct frameworks for algorithmic fairness that diverge from individual rights-based approaches (Birhane 2021; Mhlambi 2020; Metz 2022). Eke and Ogoh (2022) recover forgotten African AI narratives that center communal values over individual autonomy, demonstrating the epistemic violence of excluding non-Western ethical traditions. Indigenous and First Nations scholars have also advanced particularly trenchant critiques of AI ethics, centering data sovereignty and epistemic self-determination. In their paper, Lewis (2020) articulated how Māori, Aboriginal, and Native American communities develop AI governance grounded in relational accountability to land, ancestors, and future generations. Couldry and Mejias (2019) illuminates how the extraction of data from diverse cultural contexts mirrors historical resource extraction, undermining local value systems.

The governance of AI systems remains concentrated among a relatively small set of institutions, shaping how ethical values are defined and applied across contexts. Prior work shows that systems and policies developed within specific settings often fail to account for diverse social norms, infrastructures, and knowledge systems (Hassan 2023; D'Ignazio and Klein 2020). Platform governance research further illustrates how these dynamics manifest in practice: content moderation policies encode particular assumptions about harm and expression (Gillespie 2018), while AI systems may overlook linguistic diversity, social hierarchies, and local usage practices (Sambasivan and Arora 2021). These mismatches reflect broader structural asymmetries in AI governance, raising questions about whose values are embedded in these systems and whose are excluded. Our study builds on these critiques by providing an empirical examination of how AI ethics is interpreted and operationalized across diverse cultural contexts.

Method

We employed a qualitative, interpretivist approach (Pervin and Mokhtar 2022) to examine how experts across different cultural contexts interpret, negotiate, and envision culturally grounded approaches to AI ethics. This section outlines our participant recruitment, data collection procedures, analytical strategy, and ethical safeguards.

Recruiting & Participants

We used purposive and convenience sampling to recruit experts with knowledge of both AI systems and their sociocul-

tural contexts. Our sampling strategy aimed to capture perspectives across diverse global contexts. **Inclusion criteria** required participants to have: (1) expertise in AI development, deployment, or governance; and (2) deep familiarity with a specific cultural context, through lived experience or sustained professional engagement.

We recruited participants across multiple global regions, including Europe, North America, South America, Africa, Oceania, Asia, and the Caribbean, resulting in 14 experts from 10 countries. Recruitment was conducted through email outreach and social media platforms. We initially identified 62 potential participants who met our criteria; 17 agreed to participate, and 14 interviews were completed. The final sample (N=14) enabled in-depth exploration of cross-cultural perspectives while capturing variation across contexts. Table 1 summarizes participant demographics and regional distribution.

Study Procedures & Data Collection

We conducted semi-structured interviews with 14 experts over a seven-month period (December-June). Each interview lasted between 45-60 minutes and was conducted remotely via Zoom. With participant consent, interviews were audio-recorded, transcribed, and imported into Dovetail² for qualitative analysis. The interview protocol was designed to elicit culturally situated interpretations of AI ethics. To examine how experts engage with dominant AI ethics frameworks, we incorporated a structured probe centered on five widely cited values: bias, explainability, fairness, transparency, and privacy. These values were selected due to their prominence in existing AI ethics guidelines and policy documents identified in a prior systematic review we conducted (anonymized for review). Participants were asked:

“Listed below are commonly cited AI ethics values. Please define each value based on your cultural context.”

Following this structured component, we used open-ended questions to explore additional culturally specific values and perspectives not captured by mainstream ethics frameworks. Probes focused on: (1) perceived risks and opportunities of AI development and deployment within participants’ cultural contexts; and (2) visions for culturally grounded approaches to designing and governing AI systems. The semi-structured format allowed flexibility to pursue emergent themes while ensuring consistent coverage of the research questions across interviews.

Analysis Approach

We employed a multi-stage qualitative analysis combining iterative coding with reflexive thematic analysis (Clarke and Braun 2017). **Stage 1: Initial Coding.** The first, second, third, and fourth authors independently reviewed all transcripts to achieve familiarity with the data. The fourth author then conducted open coding and annotation using Dovetail, generating an initial codebook that captured partici-

pants’ meanings and low-level concepts. **Stage 2: Code Refinement and Categorization.** The first and second authors iteratively organized and refined initial codes to develop higher-level categories and identify relationships between concepts. This process involved constant comparison within and across transcripts to enhance conceptual clarity and coherence. **Stage 3: Reflexive Thematic Analysis.** We then engaged in reflexive thematic analysis (Clarke and Braun 2017) to develop themes across the dataset. Here, we moved beyond descriptive coding to examine patterns of meaning, with particular attention to how cultural context shaped expert perspectives on AI ethics. **Cross-Context Analysis.** To examine variation across contexts, we conducted both within-case and cross-case analyses (Møller and Skaaning 2017). Within-case analysis focused on the specificities of individual cultural contexts, while cross-case comparison identified convergent and divergent patterns across participants. This approach enabled us to preserve contextual nuance while developing broader analytical insights. The analysis process was conducted through weekly meetings involving all authors over several months. These sessions supported reflexive discussion of coding decisions, theme development, and interpretive tensions. We maintained an audit trail documenting analytical decisions and revisions to the codebook to enhance transparency and rigor.

Ethical Considerations

This study was reviewed and approved by our institution’s ethics review board. Prior to any data collection, all participants signed a consent form that included agreement to audio record their session. At the start of each interview, we reconfirmed verbal consent and addressed any questions. We reminded participants that their engagement was entirely voluntary; they could pause, skip questions, or stop the session at any time and still receive the full thank-you gift. We protected all study data, including video recordings, audio files, notes, and transcripts, through multiple safeguards: encrypting all records at rest, restricting access to only the core research team and institutional administrators, and requiring two-factor authentication with a physical security key for data access. We retained only anonymized notes for use in the publication process. Finally, we asked each participant whether they would like to be recognized in acknowledgments or materials produced as part of the research. As a best practice, we attribute quotes only to participant IDs and specifically omit unique details, phrases, or words from quotes to mitigate identification of participants.

Positionality Statement

Our research team comprises authors from diverse cultural backgrounds across Africa, Asia, the Caribbean, the Middle East, South Asia, and Latin America. This diversity informed our interpretation of the data, particularly in how cultural context, power dynamics, and historical inequities shape perspectives on AI ethics. Several authors brought lived and professional experiences navigating marginalization in technology spaces, which sensitized our analysis to issues of representation, epistemic justice, and the uneven distribution of power in AI systems. We approached the

²Dovetail is an online tool for qualitative data analysis [https://dovetail.com/]

ID	Country	Region	Sector	Area of Expertise
P1	Mexico	North/Central America	Industry	AI
P2	Kenya	Africa	Tech Founder (Non-profit)	AI / Content Moderation
P3	New Zealand	Oceania	Industry	AI
P4	USA	North America	Researcher & Founder	AI
P5	New Zealand	Oceania	Research Fellow	AI
P6	Zambia	Africa	Academia	Journalism / Emerging Media
P7	Bangladesh	South Asia	Academia	HCAI
P8	Ghana	Africa	Industry / Co-founder	AI
P9	Bangladesh	South Asia	Academia	HCI / AI
P10	Nigeria	Africa	Researcher	HCI
P11	Nigeria	Africa	Researcher	HCAI
P12	France	Europe	Academia	AI Ethics
P13	Jamaica	Caribbean	Academia	Humanities
P14	Bangladesh	South Asia	Research Fellow	HCAI

Table 1: Participant demographics across regions, sectors, and areas of expertise.

analysis with an explicitly reflexive stance, recognizing that our positionalities both informed and shaped our interpretations. Throughout the coding and theme development process, we engaged in ongoing discussions to critically examine our assumptions, question interpretations, and account for how our cultural lenses influenced analytic decisions.

Findings

We organize our findings to reflect how participants made sense of AI in their own contexts. We begin with how they described their experiences with AI in practice, particularly in terms of risks and opportunities (RQ1). These experiences then shaped how they interpreted key AI ethical values (RQ2). Finally, we examine how these insights informed their perspectives on culturally grounded approaches to AI governance (RQ3).

Experts described AI Experience based on Risks and Opportunities (RQ1):

Our analysis revealed that participants did not describe AI as a uniformly adopted or experienced technology. Instead, AI entered different contexts under uneven conditions shaped by infrastructural limitations, limited exposure, and uncertainty about how these systems function. In many cases, access to AI was contingent on broader technological ecosystems, where communities lacking stable electricity, connectivity, or devices were effectively excluded from meaningful engagement. As one participant noted:

“They [tech corporations] would prefer to work with communities that have access to smartphones, electricity, [and] high-speed internet[...] but not every community is able to access some of this product.”
- P8

This uneven entry into AI ecosystems is not simply a matter of availability, but reflects how development and deployment decisions are optimized for already-resourced environments. As a result, AI adoption is structured in ways that reproduce existing inequalities in access, shaping who benefits from these systems and who remains peripheral to them. Beyond

infrastructure, participants emphasized that AI is often encountered through competing narratives of hype, fear, and uncertainty, which shape engagement even before adoption occurs. For example, one participant described how misinformation and sensationalized narratives led to fear-based interpretations of AI, noting that *“there’s a lot of fear[...] teenagers were scared there’d be killer robots”* (P3). At the same time, others pointed to the opposite dynamic, where AI is framed in overly optimistic terms, making it *“hard to think about the negative impacts”* (P4). These contrasting narratives shaped how participants made sense of AI in their contexts, with experiences of both uncertainty and optimism leading them to evaluate AI primarily in terms of its risks and opportunities. Across our analysis, four interrelated risk patterns emerged: (1) extractive data practices, (2) loss of local power and agency, (3) labor and economic exploitation, and (4) representation gaps and epistemic harm. Rather than functioning independently, these risks collectively reflect a broader dynamic in which AI systems extract value while limiting local control and representation.

At the same time, participants identified a set of opportunities that highlight AI’s potential when systems align with local needs and contexts. Four opportunity themes emerged: (1) expanding access to services and information, (2) supporting learning and knowledge acquisition, (3) improving efficiency in everyday tasks, and (4) enabling community empowerment and local innovation. We discuss these risks and opportunities in detail below.

a) Risks in AI Deployment: Participants consistently described risks as emerging not only from the technical properties of AI systems, but also from how these systems interact with existing social, economic, and political structures. First, participants characterized **AI development as fundamentally extractive**, particularly in how data, cultural knowledge, and intellectual property are collected and repurposed without meaningful consent or governance. This extraction extends beyond raw data to include the contextual meanings embedded within that data, which are often stripped away during model training. As one participant explained, *“AI is trained on data collected without people’s consent[...] they*

just pull data online[...] there is no transparency from development to deployment to training” (P2). This highlights how AI systems rely on large-scale data aggregation practices that prioritize efficiency over accountability, raising concerns about ownership, consent, and the erasure of context.

Closely tied to this was a perceived **loss of local power and agency**. Participants described how AI systems are often designed externally and introduced into communities without meaningful opportunities for input or oversight, positioning local users as passive consumers rather than active contributors. As one participant mentioned:

“Users don’t get to determine or be part of the designing process[...] we are more like[...] just use it.” - P8

Here, the concern is not only exclusion from design, but the broader reconfiguration of authority, where decision-making power is concentrated elsewhere. This reinforces asymmetries in who gets to define how AI systems operate and whose needs they prioritize.

Participants also emphasized **labor and economic exploitation** as a critical, yet often invisible, component of AI systems. Tasks such as data annotation and content moderation were described as underpaid, precarious, and psychologically harmful, particularly in contexts with limited labor protections. As one participant noted, *“laborers are[...] exposed to violent and disturbing content[...] there’s absolutely no work protection[...] so that is also unfair”* (P9). These accounts reveal that AI systems are sustained by forms of labor that are both essential and systematically undervalued, raising questions about fairness not only in outputs, but in the conditions of production.

Finally, participants highlighted **representation gaps and epistemic harm** as a persistent risk, particularly in how AI systems fail to reflect local languages, accents, and cultural knowledge. These gaps were not framed solely as technical limitations, but as indicators of whose knowledge is prioritized in AI development. For instance, one participant explained that *“we don’t have a chatbot that responds in Swahili[...] the accent[...] is not an African accent[...] so we feel we are left behind”* (P2). This implies that AI systems can marginalize users by failing to account for linguistic and cultural diversity, resulting in both practical exclusion and symbolic erasure. Overall, in addition to the new forms of harm AI might cause, participants also highlight its potential to amplify and reconfigure existing structural inequalities in data ownership, labor, power, and representation.

b) Opportunities in AI Deployment: Despite these concerns, participants also described AI as offering meaningful opportunities, particularly when systems align with local needs, constraints, and priorities. Importantly, these opportunities were not framed as universal or inherent to AI, but as context-dependent outcomes that emerge under specific conditions of alignment between technology and use context.

Participants highlighted the role of **AI in supporting learning and education**, particularly as a tool for augmenting rather than replacing existing practices. AI was

seen as enabling new forms of engagement when integrated thoughtfully. As one participant noted, AI *“is to enhance learning[...] not to duplicate it[...] but to support what’s already there”* (P13). This reflects a broader orientation toward AI as a complementary resource, rather than a disruptive force.

In addition, AI was described as **expanding access to essential services**, particularly in contexts where traditional infrastructure is limited. Participants pointed to examples such as mobile banking and e-governance systems as ways in which AI-enabled technologies can extend reach. One participant explained:

“The phone itself is a bank[...] you don’t need to have a smartphone[...] or even a bank.” - P6

This illustrates how AI can be embedded within existing infrastructures to lower barriers to access, particularly in resource-constrained environments.

Participants also emphasized AI’s role in improving efficiency in everyday tasks, where the value of AI is often tied to convenience rather than technical understanding. As one participant noted, *“they don’t really care how the system is working[...] it’s convenient for them, and that’s everything they care about”* (P7). This highlights how adoption is often driven by practical utility, even in the absence of deep technical knowledge.

Finally, participants described AI as a potential tool for community empowerment, particularly when communities are able to shape how systems are developed and used. In such cases, AI can serve as a mechanism to amplify local voices and address existing biases. As one participant suggested, *“if we can work with the data[...] we can empower our culture and our people”* (P3). This points to the possibility of AI not only reinforcing inequalities, but also supporting more equitable forms of participation and representation when aligned with local control. Overall, these opportunities demonstrate that AI’s benefits are not guaranteed, but emerge when systems are designed and deployed in ways that reflect the specific needs and contexts of the communities they serve.

Ethical Values Interpreted Differently Across Cultural Contexts (RQ2)

Building on participants’ experiences with AI in practice, we found that interpretations of core AI ethical values varied significantly across contexts. Participants reflected on widely cited values such as bias, privacy, fairness, transparency, and explicability not as fixed or universal values, but as concepts whose meanings are shaped by cultural norms, lived realities, and structural conditions. Across interviews, participants challenged dominant assumptions embedded in existing AI ethics frameworks, particularly those grounded in Western, individualistic, and decontextualized understandings of these values. Instead, they articulated interpretations that emphasized relational, contextual, and collective dimensions of ethical practice. In doing so, participants also introduced locally grounded values such as community, respect, and collective well-being, which they

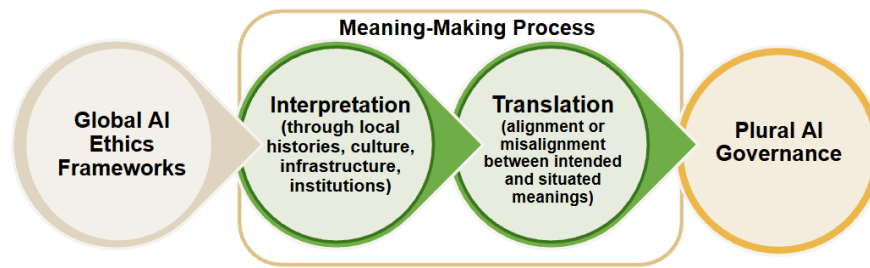


Figure 1: **Conceptual model of AI ethics translation.** Universal AI ethics frameworks acquire context-specific meanings through local interpretation. This process may produce alignment or translation gaps, which subsequently shape culturally grounded AI governance.

viewed as essential yet often overlooked in mainstream AI ethics discourse.

Our analysis identified five major themes in how participants interpreted AI ethical values: (1) reinforcement of marginalized and colonial ideals, (2) privacy as collective and relational practice, (3) transparency as a foundation for trust and accountability, (4) fairness as equity and structural justice, and (5) explicability as meaningful and contextual interpretation.

a) Reinforcement of Marginalized and Colonial Ideals:

Most participants interpreted bias not simply as a technical issue of data imbalance, but as a form of “*epistemic imperialism*” that reflects the broader power dynamics embedded within AI systems. Rather than viewing bias as an isolated flaw, participants emphasized how AI systems reproduce and reinforce existing social hierarchies by prioritizing certain knowledge systems, values, and identities over others. In this sense, bias was understood as structural, emerging from historical patterns of exclusion that are encoded into data and algorithms. For instance, P9 noted, “*when you try to apply your understanding of ethics in a culture, which you have no idea about. You are basically doing like practicing colonization, then a colonial practice, you are telling them what is right and what is wrong*”.

Participants emphasized that many AI systems are trained on datasets that fail to adequately represent local contexts, leading to outputs that are misaligned with users’ realities. As one participant noted, models are often “*trained on data that does not represent the community*” (P8), resulting in systems that fail to produce relevant or accurate outcomes. This lack of representation was not seen as incidental, but as indicative of deeper imbalances in whose data is collected and whose knowledge is valued. Beyond representation, participants described how AI systems embed and enforce dominant cultural norms, particularly those associated with Western perspectives. One participant explained:

“If it aligns with Western values, then it’s good. If it does not align[...] then it’s bad[...] we have to tell you what is human rights.” - P6

Here, bias is not only about misrepresentation, but about the imposition of particular moral frameworks as universal standards. Participants framed this as a continuation of colonial dynamics, where certain epistemologies are privileged while

others are marginalized. Together, these accounts demonstrate that bias in AI systems is not merely technical but deeply tied to questions of power, representation, and epistemic authority.

b) Privacy as Collective and Relational Practice: Participants challenged dominant interpretations of privacy as an individual right, instead framing it as a collective and relational practice shaped by social norms, family structures, and community expectations. In many contexts, privacy was not understood as control over personal data in isolation, but as something negotiated within relationships and shared social spaces. As one participant explained, assumptions that marginalized communities do not value privacy are misleading, noting that:

“So in the West, we would say the individual is everything, right? It’s about my privacy. Whereas in eastern and indigenous cultures, it’s much more of a group dynamic. And so you have extended family unit in Maori culture. Um, we talk about us, rather than I.” - P9

This highlights a disconnect between how privacy is operationalized in AI systems, typically as individual data ownership, and how it is experienced in practice. Participants emphasized that privacy decisions are often made collectively, taking into account the well-being and reputation of families or communities. Additionally, privacy was closely tied to trust and social accountability. Participants noted that sharing data is often contingent on relationships and familiarity, rather than formal consent mechanisms.

c) Transparency as a Foundation for Trust and Accountability: Participants interpreted transparency not simply as access to information about how AI systems operate, but as a foundation for building trust and ensuring accountability. Rather than focusing on technical explanations alone, participants emphasized the importance of being able to understand, question, and trace decisions made by AI systems. A recurring concern was that many AI systems are perceived as opaque and difficult to engage with, which undermines trust. As one participant noted, “*you can’t see it, you can’t touch it[...] I’m not trusting this thing*” (P13). This highlights how abstract and intangible systems challenge expectations of accountability, particularly in contexts where trust is built through direct interaction and visibility. Participants also emphasized that explanations must be meaningful

within local contexts. Technical descriptions were often insufficient, particularly when they did not align with users' knowledge or experiences. One participant illustrated this through the example of agricultural AI tools:

"When farmers ask how [AI] knows whether it's going to rain[...] no one can explain it[...] so they become uncomfortable sharing their data." - P9

Participants viewed transparency as a mechanism for enabling trust, where users can make informed decisions about whether to engage with AI systems.

d) Fairness as Equity and Structural Justice: Participants interpreted fairness not as equality of outcomes, but as equity in inclusion, emphasizing the need to account for structural inequalities in how AI systems are designed and deployed. Rather than treating fairness as a neutral or technical property, participants framed it as inherently political, shaped by access to resources, representation in data, and participation in decision-making. Participants highlighted that AI systems designed without considering infrastructural disparities may be technically fair but practically exclusionary. For example, access to AI services often depends on connectivity, literacy, and economic resources, raising questions about who benefits from these systems. As one participant noted, *"you can't call it fair if people cannot even access it,"* emphasizing that fairness must be evaluated in relation to real-world conditions.

Furthermore, fairness was linked to representation and inclusion in system design. Participants stressed the importance of ensuring that diverse voices are involved in shaping AI systems, rather than relying solely on external perspectives. This reflects a broader understanding of fairness as a process, rather than an outcome, one that requires ongoing attention to equity, participation, and structural conditions.

e) Explicability as Meaningful and Contextual Interpretation: Participants emphasized that explicability extends beyond technical transparency to include the ability to communicate AI processes in ways that are culturally meaningful and accessible. While many systems provide outputs with high confidence, participants noted that users often lack the ability to assess their validity or understand how they were generated. As one participant explained:

"The system gives very confident answers[...], but you have no idea what went into that[...] the only way to know is if you already knew the answer." - P5

This highlights a gap between system outputs and user understanding, where explanations fail to support meaningful engagement. Participants emphasized that effective explicability requires grounding explanations in local languages, cultural references, and everyday experiences. Rather than expecting users to adapt to technical systems, participants argued that AI should adapt to users' ways of knowing. This includes designing explanations that are intuitive, contextually relevant, and aligned with users' interpretive frameworks.

In addition to reinterpreting mainstream AI ethics values, participants also **articulated locally grounded values** that they viewed as essential to ethical AI design and deploy-

ment. These values, such as respect, community, and collective well-being, extend beyond dominant frameworks that prioritize individual rights and technical performance. Instead, they emphasize relational accountability, social cohesion, and alignment with cultural norms. Participants described these values as shaping how technologies should function within their communities, particularly in ensuring that systems are not only technically effective but also socially appropriate and respectful of local practices. As one participant noted:

"There are also cultural practices that AI systems fail to capture or respect, such as women covering their faces for religious or cultural reasons" - P9

This implies that ethical AI cannot be fully captured through universal values, but must also account for locally embedded value systems and meaning-making that define what is considered acceptable, respectful, and beneficial in specific contexts. Participants consistently described how global AI ethics frameworks were interpreted through local histories, cultural norms, infrastructures, and institutional realities. We define *translation gaps* as the divergence between the intended meanings of universal AI ethical principles and their situated interpretations across contexts. Figure 1 synthesizes this conceptual process, illustrating how universal AI ethics frameworks are interpreted through local cultural, historical, infrastructural, and institutional contexts, and how these situated interpretations inform more plural and culturally grounded approaches to AI governance.

Barriers and Pathways to Culturally and Plurally Grounded AI Governance (RQ3)

Drawing on participants' experiences with AI in practice (RQ1) and their interpretations of core ethical values (RQ2) as depicted in Figure 1, participants framed AI governance as a sociotechnical challenge shaped by misalignment between existing frameworks and local realities. Rather than viewing governance as a matter of implementing predefined values, participants emphasized the need to reconsider how ethical values are defined, operationalized, and enforced across contexts. Across our analysis, participants identified key barriers embedded within current approaches to AI ethics as well as pathways for rethinking governance in ways that are more culturally grounded and context-sensitive.

a) Barriers to Culturally Grounded AI Governance: Participants described three primary barriers that constrain the development of culturally grounded AI governance: (1) misalignment between ethical frameworks and contextual value systems, (2) institutional structures that limit adaptation, and (3) structural inequalities that shape participation.

Many participants emphasized that widely adopted AI ethics frameworks often **fail to translate across contexts** because they are built on specific cultural assumptions about values such as harm, fairness, and appropriateness. These frameworks tend to operationalize ethics through fixed categories, while participants' interpretations of these values are relational, contextual, and dynamic (RQ2). As a result, governance systems enforce standards that may conflict with local norms and practices. For example, one participant de-

scribed how content moderation policies reflect external interpretations of acceptable behavior:

“How we perceive nudity is different. . . even our laws are different[...] they bring their own rules.” - P8

This highlights how governance frameworks can impose singular interpretations of ethical values, overriding local legal and cultural systems. Similarly, linguistic expressions are often interpreted out of context, where socially embedded forms of communication are misclassified as harmful. These tensions reveal that governance challenges emerge not from a lack of values, but from how those values are translated into system rules and policies.

Next, participants further described how **governance is shaped by institutional systems that codify narrow interpretations** of ethical values, limiting flexibility in how these values can be applied. Legal and regulatory frameworks often reflect dominant epistemologies, making it difficult to incorporate alternative understandings. One participant illustrated this through privacy regulations:

“Privacy laws[...] only recognize the individual[...] but from our perspective, privacy includes the collective.” - P9

Additionally, participants highlighted how **governance decisions are influenced by centralized sources of authority**, particularly industry and external policy actors. As one participant explained, decision-makers are often exposed primarily to optimistic narratives about AI, which shape regulatory priorities and limit critical engagement. Together, these dynamics reinforce governance models that are difficult to adapt to diverse contexts.

Lastly, participants emphasized that participation in AI governance was shaped by broader structural conditions, including access to resources, knowledge, and infrastructure. Engagement with governance processes requires a baseline level of awareness and capacity, which is unevenly distributed across contexts. As one participant explained, *“you can’t really talk about trust unless people understand what AI is”* (P13), highlighting how limited familiarity with AI constrains meaningful participation. In addition, competing priorities, such as meeting basic needs, reduce the feasibility of engaging with governance discussions. These conditions shape whose voices are represented in governance processes, reinforcing existing inequalities in decision-making.

b) Pathways Toward Culturally and Plurally Grounded AI Governance. In response to these barriers, participants articulated pathways that move beyond adapting existing frameworks toward reconfiguring how AI governance is structured. Four key themes emerged: (1) Centering Local Knowledge in AI Design and Evaluation, (2) Enabling Co-Constructed Governance Through Participatory Approaches, (3) Community Sovereignty over Data, Knowledge, and Representation, and (4) Rebalancing AI Knowledge and Governance Across Contexts.

Participants expressed the need and urgency of **grounding AI systems in local epistemologies**, rather than adapting externally developed models to local contexts. This involves recognizing local knowledge as a foundation for design, rather than as supplementary input. For instance,

one participant described how existing cultural frameworks could guide AI development:

“We should be creating frameworks for AI development[...] much like our own research methods.” - P14

This reflects a shift toward designing systems that emerge from local ways of knowing, enabling more contextually aligned interpretations of ethical values.

Several participants also highlighted the **need for participatory governance models** that involve communities in shaping both system design and policy. Participation was framed as co-creation rather than consultation, where stakeholders actively contribute to defining ethical priorities. As one participant explained, *“you need to work with communities[...] ethics is not universal”* (P3), emphasizing that ethical values must be negotiated rather than imposed. This approach enables governance systems to reflect diverse perspectives and adapt to local contexts.

Participants further emphasized the importance of **community control over data and knowledge**, framing it as essential for addressing extractive practices identified in RQ1. Data sovereignty was described as a mechanism for ensuring that communities retain authority over how their data is used. One participant noted that *“ensuring consent and control over data[...] would be a good start”* (P2), highlighting how governance must extend beyond access to include ownership and accountability. This reframes data governance as a question of stewardship and responsibility.

Finally, participants described the need to **rebalance global AI governance by redistributing authority** and recognizing multiple knowledge systems. This involves moving away from centralized models toward more distributed and context-sensitive approaches. As one participant explained:

“We need to stop extractive practices[...] build locally[...] and learn from experts within those contexts.” - P6

This reflects a broader shift toward plural governance, where ethical values are not standardized but negotiated across contexts. Overall, participants framed AI governance as an ongoing process of negotiating across diverse value systems, rather than implementing universal principles. Addressing the identified barriers and advancing these pathways requires rethinking how ethical values are translated, how authority is distributed, and how participation is structured in AI systems. Synthesizing our findings across RQ1-RQ3, we present a conceptual overview of how participants’ experiences with AI shape the interpretation of ethical values, where misalignments emerge, and how these tensions inform governance implications. This synthesis highlights key patterns that connect experiences, interpretations, and governance across contexts.

Discussion

In this section, we situate our findings within broader conversations in AI ethics and governance, focusing on how participants’ experiences, interpretations of ethical values, and proposed pathways reveal key limitations in current approaches. We highlight two key contributions that both align with and extend prior work.

Rethinking deficit framing in AI ethics positions local knowledge as Governance Infrastructure

In our findings, participants did not position themselves or their communities as lacking ethical understanding of AI. Instead, they articulated rich, contextually grounded interpretations of values, alongside locally rooted values such as community, respect, and collective well-being. These perspectives were not framed as alternatives to mainstream AI ethics, but as already-existing systems of meaning that shape how technology is evaluated and used in practice. Dominant trends in AI ethics has focused on identifying and aggregating shared ethical principles such as fairness, transparency, and accountability across institutions and national contexts (Hagendorff 2022; Jobin, Ienca, and Vayena 2019). These efforts have been instrumental in establishing a common vocabulary for AI governance, and in this sense, our findings align with these prior works in recognizing the continued relevance of these values.

However, many of the prior papers also rely on an implicit assumption that once defined, these values can be uniformly interpreted and operationalized across contexts (Hagendorff 2022). Our findings complicate this assumption. Participants consistently demonstrated that the central challenge is not defining ethical values, but translating them into practice in ways that align with local contexts. This resonates with longstanding scholarly critiques of abstraction, particularly (Tidjon and Khomh 2022; Goffi and Momcilovic 2022), which argue that systems often fail when they impose simplified representations onto complex social realities. In our study, this abstraction manifests as what we conceptualize as “*translation gaps*”: which are divergences between the meanings intended by universal AI ethics frameworks and the situated interpretations these values acquire across cultural, historical, and institutional contexts.

Crucially, these gaps are not only interpretive but material. As seen in RQ1 and RQ3, they shape how systems are adopted, trusted, and resisted. Misalignment leads to misclassification of culturally situated behaviors, breakdowns in trust, and governance systems that feel externally imposed rather than locally relevant. This extends prior critiques of “ethical abstraction” (Greene, Hoffmann, and Stark 2019; Munn 2023) by showing how these issues are experienced in practice across diverse contexts. At the same time, our findings challenge the prevailing “inclusion” paradigm in HCI and AI ethics (Dine 2025; Oguine et al. 2025). Much prior work (Mayeesha, Islam, and Ahmed 2025; Chi, Lurie, and Mulligan 2021) has framed marginalized communities as needing access, representation, or inclusion within existing systems (e.g., digital divide literature, participatory design). While these efforts are important, our findings suggest that inclusion alone is insufficient. In many cases, inclusion operates within systems whose underlying assumptions remain unchanged, resulting in what participants described as forms of epistemic extraction where local data, knowledge, and perspectives are incorporated without shifting who defines the system (Munn 2023). Our findings suggest that when communities are included only as data sources or consulted without veto power, AI ethics risks becoming a form

of ‘participatory extraction’ in which local nuances are harvested to make global systems more resilient, without ceding any structural authority to local actors. Hence, participants’ interpretations of values and their articulation of locally grounded values demonstrate that governance can be built from these knowledge systems, rather than layered onto them. This shifts the focus from “how do we include diverse users?” to “how do we design governance systems that emerge from diverse epistemologies?”

Moving from Universalism to Plural, Power-Aware AI Governance

Our findings also speak directly to ongoing debates about universalism in AI ethics (Floridi and Cowls 2019; Munn 2023). Many global frameworks, like the OECD and UNESCO, operate on the premise that ethical alignment can be achieved through convergence on shared values (Munn 2023; Birhane et al. 2022). This model of “value universalism” assumes that concepts such as fairness or transparency possess stable meanings that can be applied across contexts (Floridi and Cowls 2019; Mohamed, Png, and Isaac 2020). Participants in our study did not reject these values. Instead, they revealed how these values take on multiple, co-existing meanings depending on context. As illustrated in Figure 1, governance emerges only after ethical values have been interpreted within local contexts. This suggests that supporting plural AI governance requires designing mechanisms that acknowledge and negotiate these situated interpretations, rather than assuming that shared principles will be uniformly understood. These interpretations align with growing critiques in HCI and AI ethics that question the universality of ethical values (Greene, Hoffmann, and Stark 2019; Khan et al. 2021), while providing empirical grounding for how these tensions manifest in practice.

Importantly, our findings suggest that the challenge of AI governance is not simply cultural, but deeply tied to power and epistemic authority. Decisions about how ethical values are defined, operationalized, and enforced remain concentrated within specific institutions, regions, and industries. This reflects broader patterns of epistemic inequality, where certain forms of knowledge are privileged while others are marginalized (Birhane et al. 2022; D’Ignazio and Klein 2020; Munn 2023). As participants described, local actors are often positioned as adopters of externally developed systems rather than contributors to their design and governance. This insight extends existing calls for participatory and inclusive AI governance by emphasizing that participation alone does not address underlying power asymmetries. Instead, what is needed is a shift toward plural and power-aware governance, where multiple value systems are not only represented but actively shape decision-making processes. In this model, governance is not about enforcing consensus on a single interpretation of ethical values, but about enabling coordination across diverse perspectives. The pathways identified in our findings, such as participatory governance, local knowledge integration, and data sovereignty, point toward how this shift might be realized in practice. These approaches redistribute authority, recognize diverse forms of expertise, and create space for ongoing

ing negotiation of values. Rather than treating ethical values as fixed standards to be implemented, plural AI governance requires institutional mechanisms that support the ongoing interpretation, negotiation, and revision of ethical values across contexts.

Implications for AI Design, Policy and Governance

We outline implications that follow from our findings, focusing on how AI systems, governance frameworks, and participation models must shift to better support the contextual interpretation and plural negotiation of ethical values.

Design Implications: Our findings suggest that AI system design must move beyond treating ethical values as fixed inputs and instead support their contextual interpretation in use and systemic contestability (Alfrink et al. 2023). Prior work within AIES and allied fields (Varshney 2024; Delgado et al. 2023; Whittlestone et al. 2019) has emphasized the importance of designing for situated action and meaning-making, yet many AI systems continue to rely on static representations of harm, fairness, and appropriateness. Designing for translation requires developing systems that can accommodate multiple interpretations of ethical values, potentially through pluralistic auditing, localized models, or mechanisms that allow users to contest and reinterpret system outputs through local community councils or groups. This shifts the role of AI from enforcing predefined norms to supporting ongoing negotiation of meaning, which is critical in culturally diverse and globally deployed systems.

Governance and Policy Implications: At the policy level, our findings challenge the dominance of one-size-fits-all governance frameworks. While global guidelines provide useful high-level principles, they often lack mechanisms for contextual adaptation. Subsidiarity in AI governance should therefore operate across multiple levels, allowing ethical values to be interpreted and operationalized locally while maintaining broader coordination (CAPP-USA 2026). This aligns with emerging discussions in AI policy that advocate for flexible and context-sensitive regulatory approaches (Whittlestone et al. 2019). Importantly, such models must also account for the structural conditions identified in RQ2, ensuring that governance frameworks are not only adaptable in principle but feasible in practice across different infrastructural and socio-political contexts.

Implications for Participation and Stakeholder Engagement: Our findings also reinforce the need to move from consultation-based approaches toward co-constructive models of participation. While participatory design has long been a focus in HCI, its application in AI governance remains limited and often superficial (Oguine et al. 2026; Sloane et al. 2022). Participants in this study emphasized that meaningful engagement requires not only inclusion but also the ability to influence decision-making processes. This requires addressing structural barriers such as access, literacy, and resource constraints that limit participation, as well as creating sustained mechanisms for engagement rather than one-off consultations. In doing so, governance becomes a continuous, collaborative process that evolves with community needs and interpretations.

Implications for Data Practices and Governance: The

findings further highlight the need to reconceptualize data as socially and culturally embedded, rather than as neutral inputs to AI systems (Loukissas 2019). Existing data practices often prioritize extraction and aggregation, overlooking the relationships and meanings associated with data in different contexts. Participants' emphasis on data sovereignty aligns with growing calls for more equitable and accountable data governance models (Mhlambi 2020; Kukutai and Taylor 2016). Supporting community control over data requires not only technical solutions but also institutional changes that enable ownership, consent, and stewardship at the community level. This shifts data governance from a model of access to one of accountability and reciprocity.

Limitations and Future Work

This work should be interpreted in light of several limitations that also point to future research. First, our purposive sample of 14 experts enabled the depth of qualitative analysis we sought. While the findings are not intended to be generalizable, they offer theoretically transferable insights into how ethical translation operates across diverse contexts. Accordingly, the perspectives presented here should be understood as situated interpretations shaped by participants' professional and lived experiences, rather than representative accounts of the cultures in which they are situated. Future work should examine how translation operates across a broader range of stakeholders, particularly marginalized and non-expert communities. Second, our findings rely on self-reported accounts, which may differ from how AI ethics is negotiated in practice. Ethnographic or longitudinal studies could examine these dynamics over time. Finally, while this paper highlights the need for plural AI governance, it remains primarily analytical. Future work should explore governance mechanisms that balance locally situated interpretations of ethical values with globally shared commitments to human rights, safety, and accountability.

Conclusion

This paper examined how experts across diverse contexts experience AI systems, interpret core ethical values, and envision pathways toward more culturally grounded governance. Our findings show that widely cited AI ethics principles, such as fairness, are not universally interpreted, but are shaped by contextual and relational factors. We argue that the central challenge in AI ethics is not defining shared values, but translating them across contexts. These translation gaps contribute to misalignment between systems and the communities they serve, reinforcing the need for more context-sensitive approaches. In response, participants highlighted pathways that emphasize local knowledge, participatory processes, and more distributed forms of governance. Hence, this work reframes AI governance as an ongoing process of negotiating ethical values across diverse contexts, pointing toward the need for more plural and power-aware approaches to designing and governing AI systems.

Acknowledgments

We sincerely thank all of our participants for sharing their time, experiences, and perspectives for this study. Their thoughtful reflections made this work possible.

References

- Alfrink, K.; Keller, I.; Kortuem, G.; and Doorn, N. 2023. Contestable AI by Design: Towards a Framework. *Minds & Machines*, 33(4): 613–639.
- Birhane, A. 2021. Algorithmic Injustice: A Relational Ethics Approach. *Patterns*, 2(2): 100–103.
- Birhane, A.; Ruane, E.; Laurent, T.; S. Brown, M.; Flowers, J.; Ventresque, A.; and L. Dancy, C. 2022. The Forgotten Margins of AI Ethics. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 948–958. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.
- CAPP-USA. 2026. Subsidiarity in the Age of AI: A Catholic Framework for AI Governance. *CAPP-USA*.
- Chi, N.; Lurie, E.; and Mulligan, D. K. 2021. Reconfiguring Diversity and Inclusion for AI Ethics. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 447–457. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384735.
- Clarke, V.; and Braun, V. 2017. Thematic analysis. *The journal of positive psychology*, 12(3): 297–298.
- Couldry, N.; and Mejjias, U. A. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.
- Delgado, F.; Yang, S.; Madaio, M.; and Yang, Q. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703812.
- D'Ignazio, C.; and Klein, L. F. 2020. *Data Feminism*. MIT Press.
- Dine, J. 2025. The Digital Inclusion Outlook: What It Looks Like and Where It's Lacking. *Information Technology and Innovation Foundation | ITIF*.
- Eke, D.; and Ogoh, G. 2022. Forgotten African AI Narratives and the Future of AI in Africa. *International Review of Information Ethics*, 32: 1–8.
- Floridi, L.; and Cowls, J. 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Sci. Rev.*, 1(1).
- Gillespie, T. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Goffi, E. R.; and Momcilovic, A. 2022. Respecting cultural diversity in ethics applied to AI : a new approach for a multicultural governance. *Misión Jurídica*, 15(23): 111–122.
- Greene, D.; Hoffmann, A. L.; and Stark, L. 2019. Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning.
- Hagendorff, T. 2022. Blind spots in AI ethics. *AI Ethics*, 2(4): 851–867.
- Hassan, Y. 2023. Governing Algorithms from the South: A Case Study of AI Development in Africa. *AI & Society*, 38: 1429–1442.
- High-Level Expert Group. 2019. Ethics Guidelines for Trustworthy AI. Technical report, European Commission.
- Jobin, A.; Ienca, M.; and Vayena, E. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 1(9): 389–399.
- Khan, A. A.; Badshah, S.; Liang, P.; Khan, B.; Waseem, M.; Niazi, M.; and Akbar, M. A. 2021. Ethics of AI: A Systematic Literature Review of Principles and Challenges. *arXiv*.
- Kukutai, T.; and Taylor, J. 2016. *Indigenous data sovereignty: Toward an agenda*, volume 38. ANU press.
- Lewis, J. E. E. 2020. Indigenous Protocol and Artificial Intelligence Position Paper. Technical report, The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR), Honolulu, Hawai'i.
- Loukissas, Y. A. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA, USA: The MIT Press. ISBN 978-0-26235221-5.
- Mayeesha, T. T.; Islam, F.; and Ahmed, N. 2025. AI4Bangladesh: AI Ethics for Bangladesh - Challenges, Risks, Principles, and Suggestions. In *Proceedings of the 13th International Conference on Information & Communication Technologies and Development*, ICTD '24, 260–272. New York, NY, USA: Association for Computing Machinery. ISBN 9798400710414.
- Metz, T. 2022. *A Relational Moral Theory: African Ethics in and Beyond the Continent*. Oxford University Press.
- Mhlambi, S. 2020. From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance. Technical Report Discussion Paper 009, Carr Center for Human Rights Policy, Harvard Kennedy School.
- Mhlambi, S.; and Tiribelli, S. 2023. Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms. *Topoi: An Alternative Review of Philosophy*, 42: 867–880.
- Mohamed, S.; Png, M.-T.; and Isaac, W. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4): 659–684.
- Møller, J.; and Skaaning, S.-E. 2017. Explanatory typologies as a nested strategy of inquiry: Combining cross-case and within-case analyses. *Sociological Methods & Research*, 46(4): 1018–1048.
- Morley, J.; Floridi, L.; Kinsey, L.; and Elhalal, A. 2021. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. In *Ethics, governance, and policies in artificial intelligence*, 153–183. Springer.

- Munn, L. 2023. The uselessness of AI ethics. *AI Ethics*, 3(3): 869–877.
- Nemorin, S.; and Bonami, B. 2026. AI ethics through a decolonial lens: what does AI ethics look like if we take seriously the push to decolonise it? *AI & Soc.*, 1–13.
- OECD. 2019. Recommendation of the Council on Artificial Intelligence. Technical Report OECD/LEGAL/0449, Organisation for Economic Co-operation and Development. Adopted May 2019; amended November 2023.
- OECD. 2026. AI Use by Individuals Surges Across the OECD as Adoption by Firms Continues to Expand.
- Oguine, O. C.; Alvarado Garcia, A.; Muller, M.; and Badillo-Urquiola, K. 2026. Who Gets to Define Safety? A Systematic Review of How Generative AI Research Addresses Youth Online Safety. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. New York, NY, USA: Association for Computing Machinery. ISBN 9798400722783.
- Oguine, O. C.; Radford, D.; Hoang, S.; and Badillo-Urquiola, K. 2025. Inclusion by Design or by Default? Examining How HCI Studies Engage Underrepresented Youth in Makerspaces. In *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing*, CSCW Companion '25, 326–334. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714801.
- Pervin, N.; and Mokhtar, M. 2022. The interpretivist research paradigm: A subjective notion of a social context. *International Journal of Academic Research in Progressive Education and Development*, 11(2): 419–428.
- Sambasivan, N.; and Arora, S. 2021. Re-imagining AI for the Global South. Technical report, Google Research India.
- Schiff, D.; Biddle, J.; Borenstein, J.; and Laas, K. 2020. What's Next for AI Ethics, Policy, and Governance? A Global Overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, 153–158. New York, NY, USA: Association for Computing Machinery. ISBN 9781450371100.
- Schultz, M. D.; Conti, L. G.; and Seele, P. 2025. Digital ethicswashing: a systematic review and a process-perception-outcome framework. *AI Ethics*, 5(2): 805–818.
- Sloane, M.; Moss, E.; Awomolo, O.; and Forlano, L. 2022. Participation Is not a Design Fix for Machine Learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394772.
- Stanford Institute. 2025. The 2025 AI Index Report. Technical report, Stanford University.
- Statista. 2025. AI Market Size Worldwide 2020-2031.
- Tidjon, L. N.; and Khomh, F. 2022. The Different Faces of AI Ethics Across the World: A Principle-Implementation Gap Analysis. *arXiv*.
- UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence. Technical Report SHS/BIO/REC-AIETHICS/2021, United Nations Educational, Scientific and Cultural Organization, Paris.
- Varshney, K. R. 2024. Decolonial AI Alignment: Openness, Viśeṣa-Dharma, and Including Excluded Knowledges. In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 1467–1481.
- Whittlestone, J.; Nyrup, R.; Alexandrova, A.; Dihal, K.; Cave, S.; et al. 2019. Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *London: Nuffield Foundation*, 1–59.