
Structural Amplifiers of AI-Induced Harm: A Five-Dimension Sector Vulnerability Framework for South and Southeast Asia

Sakshi Chaubey

Aaraam Research & Learning Studio

With

Apart Research

Abstract

Artificial intelligence is being deployed at scale across South and Southeast Asia in sectors that directly govern the livelihoods, health, credit access, welfare entitlements, and physical safety of hundreds of millions of informal, low-income, and historically marginalised people. Yet existing AI risk frameworks assess harm at either the model level or the country level, neither of which tells policymakers or civil society organisations which sectors require the most urgent intervention. This paper proposes a five-dimension sector vulnerability framework that assesses the structural conditions under which AI deployment poses the greatest risk to users and workers across sectors in South and Southeast Asia. Applying the framework to five sectors across six countries, i.e India, Vietnam, Indonesia, Philippines, Thailand, and Bangladesh, we find that two sectors score at critical risk (rural healthcare and public welfare, 15/15) and three at high risk (gig platforms, financial services, and surveillance, 14/15 each). Workforce vulnerability scores maximum across all five sectors, and harm visibility represents the most consistently unaddressed structural gap. The framework is designed to be replicable and updatable by NGOs, policymakers, and researchers without technical expertise.

1. Introduction

AI deployment in South and Southeast Asia is moving faster than the governance infrastructure meant to keep up with it. Across India, Vietnam, Indonesia, the Philippines, Thailand, and Bangladesh, AI systems now have a hand in who receives welfare entitlements, who gets approved for credit, who gets flagged by law enforcement, who can earn enough to eat that day, and whose medical complaint is believed. The people most affected by these systems — informal gig workers, rural healthcare patients, smallholder borrowers, welfare-dependent households, communities under algorithmic policing — are also, almost without exception, the people least present in the rooms where AI risk frameworks get written.

Existing frameworks do not fill this gap. Country-level governance indices can tell you how regulatory-ready a nation is overall, but they say nothing about which sector inside that country needs attention first. Expert elicitation studies are good at ranking which categories of AI risk worry experts the most, but they don't disaggregate by geography nor explain why some populations end up absorbing significantly more harm than others. The sector-specific harm studies that do exist tend to live in silos with no shared structure that lets you compare across them. Mohamed, Png and Isaac (2020) describe this kind of exported framework as a form of algorithmic coloniality: AI governance built around Global North risk tolerances and institutional assumptions, then dropped onto populations whose actual conditions look nothing like that.

This paper attempts to address this gap through a narrower question: given a sector, what structural conditions make the people inside it especially exposed when something does go wrong, independent of what that something turns out to be? This is done through a five-dimension sector vulnerability framework assessing what structural conditions make the people in a given sector especially exposed to harm when AI is deployed there, regardless of which specific risk materialises. The five dimensions I use are deployment intensity, workforce and user vulnerability, regulatory gap, data exploitation risk, and harm visibility and accountability, which work less like risk categories and more like amplifiers.

This paper makes three contributions:

- A five-dimension sector vulnerability rubric, grounded in critical AI scholarship filling the gap in the existing frameworks and research done in this area
- A comparative vulnerability scoring of five priority sectors across six South and Southeast Asian countries, which produces a scored sector-level risk map for the region
- A set of sector-specific findings relevant for NGOs, policymakers, and researchers, grounded in the structural conditions of each sector

2. Related Work

Four bodies of prior work inform this framework and together define the gap it fills.

Theoretical foundations. Mohamed, Png and Isaac (2020) introduce algorithmic coloniality: the replication of colonial patterns of extraction and dispossession through AI systems, identifying ethics dumping (exporting harmful AI practices to countries with weaker regulatory protection) and algorithmic exploitation (extracting value from marginalised workers through opaque AI systems) as the theoretical basis for this framework's harm visibility and workforce vulnerability dimensions respectively. Chan, Okolo, Turner and Wang (2021) demonstrate empirically that Southeast Asia is entirely absent from the top ten publication indices of the world's largest AI research conferences, establishing that the populations most at risk from AI deployment in the target countries have the least representation in spaces where AI risks are defined. Their distinction between inclusion and redistribution grounds this paper's recommendation logic: better ethics guidelines applied to unchanged systems will not alter the structural conditions that produce harm.

Existing risk frameworks and their limits. Gandhi, Cihon, Larter and Anselmetti (2025) introduce the Societal Capacity Assessment Framework (SCAF), an indicators-based approach to measuring country-level vulnerability and resilience to AI-related risks, the closest existing methodological precedent to this work. The authors explicitly note SCAF "could be further adapted to assess vulnerability and resilience to sector-specific risks"; this paper undertakes that adaptation for the Global South context. The MIT AI Risk Repository's Delphi study (Noetel et al., 2025), aggregating judgments from 272 AI experts, takes risks as the primary unit and asks which sectors are exposed to each risk, the inverse of this framework's logic, which takes sectors as the primary unit and asks what structural conditions amplify whatever harm occurs. The AGILE Index (2025) evaluates 40 countries across 43 indicators but operates entirely at the country level. None of these frameworks asks what structural conditions make a sector inherently vulnerable to AI-induced harm in the Global South, which is the question at the centre of this paper.

Harm and accountability scholarship. Raji and Buolamwini (2019) demonstrate that algorithmic harm correction through auditing is highly contingent, requiring harm to be visible enough to attract an auditor, disclosure to generate reputational pressure, and companies to have sufficient market exposure to respond. Their finding that companies not publicly targeted showed no improvement even after comparable bias was documented directly grounds this framework's harm visibility dimension: the accountability chain that functions conditionally in high-income contexts is structurally unavailable in the sectors assessed here. Shelby et al. (2023) provide the cross-sector taxonomy of sociotechnical harms used to define what counts as harm in each sector profile.

Data and labour exploitation scholarship. Couldry and Mejias (2019) establish that data relations appropriate human activity through social arrangements that render consent fictional, such that

the more essential the service, the less real the choice, providing the foundational logic for the data exploitation dimension. Sambasivan et al. (2021), based on 36 interviews across India, demonstrate that conventional algorithmic fairness assumptions fail systematically in the Indian context, providing the foundational argument for why a framework built from Global South conditions outward is necessary. Ramesh et al. (2022) document platform power asymmetry in Indian lending platforms, showing empirically how financial dependency prevents users from contesting algorithmic decisions.

No existing framework uses a consistent set of structural dimensions to comparatively score sectors by their vulnerability to AI-induced harm in the Global South. This is the gap that this paper aims to fill.

3. Methods

Research design

This study employs a qualitative comparative analysis design, applying a consistent five-dimension rubric across five sectors in six countries to produce ordinal vulnerability scores. The methodology follows the codebook-guided scoring approach used in AI governance research (AGILE Index, 2025), yielding ordinal indicators scored 1–3 per dimension rather than quantitative measurements. This is appropriate given the heterogeneous nature of available evidence across countries and sectors. Every score is accompanied by at least one cited source; scores are comparative vulnerability indicators, not cardinal rankings.

Unit of analysis

The sector as experienced by users and workers within it across the six target countries collectively. Scores reflect structural conditions in aggregate, **not** country-specific scores. The sector analysis makes a note of it whenever there is strong country-specific evidence.

Country selection

The countries selected for this study are India, Vietnam, Indonesia, Philippines, Thailand, and Bangladesh. They were selected on three criteria: rapid AI deployment growth, high informality in target sectors, and regulatory environments in active but incomplete transition. Singapore was excluded as a high-income outlier whose conditions are not representative of the populations this framework serves.

Sector selection. Five sectors were selected based on their placement at the intersection of two criteria: designation as high-risk in EU AI Act Annex III, and documented AI deployment in at least two target countries in secondary literature. Sectors were excluded where primary harm falls on institutions or democratic processes rather than directly on workers and users.

The **five dimensions** are defined as follows, with the rationale for how each is scored, with a score of 1 implying low vulnerability and a score of 3 implying high vulnerability:

Table 1: Five-Dimension Scoring Rubric

Dimension	Score 1	Score 2	Score 3
AI deployment intensity	Emerging/pilot	Significant active use	Pervasive, opt-out impossible
Workforce & user vulnerability	Formal, protected, recourse exists	Mixed formal/informal, weak enforcement	Informal, marginalised, no practical recourse
Regulatory gap	Binding sector-specific law	General law, no AI-specific provision	No applicable binding rule anywhere
Data exploitation risk	Disclosed collection, meaningful consent	Broad terms, limited user awareness	Essential service dependency, consent structurally impossible
Harm visibility & accountability	Documented, functional redress mechanisms	Aggregate harm documented, redress inaccessible	Invisible at all levels, no mechanism exists

Data sources. Primary sources per dimension: AI deployment intensity > OECD AI Policy Observatory, Rest of World, national AI strategy documents; workforce vulnerability > ILO informal employment data, World Bank, sector-specific labour studies; regulatory gap > national AI laws (Vietnam Law 134/2025, India MeitY frameworks), ASEAN Guide 2024, EU AI Act Annex III as benchmark; data exploitation risk > Access Now, Internet Freedom Foundation, AIAAIC incident database; harm visibility > AIAAIC, civil society documentation, academic audit studies.

NOTE: In sectors where harm documentation is sparse, harm visibility is scored 3 (hidden) rather than 1 (visible). Absence of documentation is treated as evidence of accountability infrastructure failure, not evidence of absence of harm, based on the work by Raji and Buolamwini (2019) and Chan et al. (2021).

4. Results

Table 2: Complete Sector Vulnerability Scoring Matrix

Sector	Deployment	Vulnerability	Reg. Gap	Data Exploitation	Harm Visibility	Total /15	Tier
Gig & delivery platforms	3	3	2	3	3	14	HIGH
Rural & community healthcare	3	3	3	3	3	15	CRITICAL
Financial services & credit	3	3	2	3	3	14	HIGH
Public welfare & benefits	3	3	3	3	3	15	CRITICAL
Surveillance & law enforcement	3	3	3	3	2	14	HIGH

Sector profiles

Gig and delivery platforms (Swiggy, Zomato, Grab, GoTo/Gojek: India, Vietnam, Indonesia, Philippines, Thailand): Algorithmic management controls task assignment, pay, ratings, and deactivation for an estimated 12 million gig workers in India alone, with comparable deployment across Southeast Asia. Workers have no visibility into how decisions are made and no formal mechanism to contest them. Regulatory gap scores 2 rather than 3 because Karnataka's 2025 state law and India's Social Security Code represent partial steps. Data is extracted as a condition of working; HRW documented rights abuses across nine countries including India in 2026 with no accountability response.

Rural and community healthcare (e-Sanjeevani, AI diagnostics: India, Vietnam, Indonesia, Philippines): India's e-Sanjeevani supported 282 million AI-assisted consultations as of March 2025. AI is deployed by ASHA workers in India and Barangay Health Workers in the Philippines, which are state-mandated community health workers without training in AI failure modes. ICMR Ethical Guidelines exist but are explicitly non-binding; patient data is collected as the only means of accessing care; misdiagnosis harms are systematically undercounted. ECRI named AI diagnostic dilemmas the number one global patient safety concern for 2026. This is the only sector to achieve a perfect vulnerability score i.e. every structural amplifier operates simultaneously at maximum.

Financial services and algorithmic credit scoring (digital lending, alternative credit scoring: India, Vietnam, Indonesia, Philippines): Approximately 91 million adults in Indonesia are unbanked or underbanked; AI-based credit scoring fills this gap using smartphone metadata, behavioural biometrics, and psychometric indicators collected without explicit consent for credit assessment purposes. Indonesia's POJK 29/2024 introduces baseline consumer protection for licensed alternative credit scoring providers but covers only third-party providers, not internal banking algorithms. A November 2025 review confirmed systematic disadvantages for minority groups and women persisting despite industry efficiency claims.

Public welfare and benefits targeting (Aadhaar, Samagra Vedika, G2P systems: India, Bangladesh): Over 140 crore Aadhaar numbers generated as of early 2026 with private companies being able to ask for Aadhar information; Bangladesh's NID was integrated into G2P delivery. A 2024 Al Jazeera and Amnesty International investigation revealed Samagra Vedika causing arbitrary welfare denial for thousands with the private company responsible declining to respond to Amnesty's enquiries. This is the most existentially consequential sector: exclusion errors affect survival. Documentation of harm generates zero institutional accountability response, producing a harm visibility score of 3 despite external documentation.

Surveillance and law enforcement (facial recognition, predictive policing: India, Vietnam, Indonesia): India's Ministry of Law and Justice confirmed AI integration in policing, surveillance, and criminal investigations as of 2025; predictive policing systems expanded across multiple Indian states. FRT is deployed without legal authorisation, necessity assessments, or accuracy disclosures in any target country. Vietnam's smart city surveillance camera expansion in Hanoi and Ho Chi Minh City is documented but largely unaudited by civil society which creates a noted evidence gap. IFF (2023) documented FRT accuracy failures and vendor accountability gaps in India, explaining why harm visibility scores 2 rather than 3: active Indian civil society has produced aggregate-level documentation, even without individual accountability mechanisms.

Cross-sector findings

Finding 1: Workforce vulnerability is universal and total

Every sector scores maximum on this dimension. The populations bearing the consequences of AI deployment across the region are consistently those least able to contest, resist, or exit the systems governing them, including but not limited to informal gig workers, rural patients, unbanked borrowers, welfare-dependent households, and communities subject to discriminatory policing share the structural combination of informality, power asymmetry, economic disadvantage and institutional exclusion that Sambasivan et al. (2021) and Mohamed et al. (2020) identify as the defining feature of AI harm in the Global South. Any intervention that improves AI systems without improving the structural position of those they affect will be insufficient.

Finding 2: The regulatory gap is near-total, with hyper-localised exceptions

Six of ten possible regulatory gap scores are 3. The two partial exceptions being gig platforms

(India's Karnataka law) and financial services (Indonesia's POJK 29/2024) which represent single-country, sector-specific steps with no regional equivalents. No sector has comprehensive binding coverage across the six target countries. The most urgent gaps are in rural healthcare and public welfare, where AI decisions which can be existentially threatening beyond a point are made with little to no binding oversight anywhere in the region.

Finding 3: Data exploitation at score 3 is near-universal

Four out of five sectors score maximum on data exploitation. Every sector involves sensitive personal data collected from populations who cannot practically or meaningfully refuse because in majority of these cases, the service is a gatekeeper to something essential, confirming Couldry and Mejias's (2019) core argument at the sector level.

Finding 4: Harm visibility differentiates sectors by existing civil society capacity

The only sector scoring below 3 on harm visibility is surveillance. This is not because sufficient accountability mechanisms exist, but because India's digital rights civil society (IFF, SFLC, Amnesty) has produced documented evidence of aggregate harm. Rural healthcare and public welfare score 3 precisely because affected populations have the least civil society representation. As Chan et al. (2021) would describe it, it is structural invisibility produced by exclusion agenda-setting and decision-making.

5. Discussion

This framework surfaces a finding with direct implications for how AI safety is conceptualised in the Global South: the primary AI safety problem in South and Southeast Asia is the structural conditions under which everyday AI deployment causes compounding, invisible, unaddressed harm to populations who in actuality have no power to contest it. The sectors scoring CRITICAL in this analysis (rural healthcare and public welfare) are national programmes, active at scale, affecting hundreds of millions of people, with no limited oversight. These sectors experience a severe gap of better accountability infrastructure which should include auditing capacity, incident reporting mechanisms, legal redress pathways, and governance frameworks built with and for the populations they are meant to protect.

The framework also surfaces a structural insight about the relationship between harm visibility and harm magnitude: the sectors where harm is least visible are those where affected populations have the least power to define harm in the first place. This, along with the other findings, imply that harm visibility is a power problem requiring redistribution of the capacity to define, document, and demand accountability for AI-induced harm, something we see missing across sectors and countries.

Limitations

This paper is a part of the Apart Research Global South hackathon which spanned over a 3-day period, thus including time constraint issues in addition to other limitations. In this study, scoring is qualitative and involves interpretive judgement, mitigated by consistent rubric application and cited evidence for every score. Secondary source analysis means dependence on what has already been documented, which itself reflects the harm visibility gaps the framework identifies. Country coverage in available sources is substantially uneven: India is extensively documented by civil society and academic research, while Bangladesh and Thailand have materially thinner coverage, and Vietnam's surveillance sector is a noted evidence gap. The framework produces a snapshot of conditions as of mid-2026; scores in the regulatory gap dimension in particular may shift as Vietnam's PDPL 2026 and India's DPDP Act enforcement develop. No primary data collection, expert elicitation, or community consultation with affected populations was conducted during the hackathon.

Future Work

The immediate next step I would like to pursue with this study is a disaggregation by country producing a full 5×6 matrix, one score per sector per country, enabling identification of specific country-sector combinations requiring urgent intervention. Furthermore, three additional sectors, which are hiring and recruitment screening, garment and electronics manufacturing, and EdTech and student assessment, need to be included in the next iteration.

With more resources and time, the study would include structured interviews with workers and users in target sectors, a Delphi-style expert elicitation to validate scores, and dynamic longitudinal scoring updated annually as regulatory environments evolve. Most importantly, a participatory iteration would involve affected communities in co-designing the rubric which is consistent with the redistribution of epistemic power that Chan et al. (2021) identify as necessary for AI governance to genuinely serve Global South populations.

The framework's value ultimately depends on reaching the people who can act on it: future iterations should be developed in active partnership with NGOs, think tanks and civil society organisations so that scores translate into advocacy priorities and regulatory submissions.

6. Conclusion

This paper has proposed and applied a five-dimension sector vulnerability framework for assessing the structural conditions under which AI deployment poses the greatest risk to users and workers in South and Southeast Asia. By taking sectors as the primary unit of analysis and measuring structural amplifiers of harm rather than specific AI risk categories, it fills a gap between

country-level AI governance indices and sector-specific AI harm studies, neither of which tells policymakers or NGOs which sectors need intervention first and why.

This work produced three main findings: 1) workforce and user vulnerability scores maximum in every sector assessed: the people bearing the consequences of AI deployment across the region are universally those least able to contest or exit the systems governing them, 2) the regulatory gap is near-total, with hyper-localised exceptions in individual countries, and 3) harm invisibility is structurally produced by the same exclusion from power that characterises every other dimension i.e the sectors where harm is least visible are precisely those where the affected populations have the least capacity to surface it.

The framework is aimed to be replicable. Any NGO, policymaker, or researcher can apply the five-dimension rubric to new sectors or countries using the scoring criteria and data sources specified here. This is meant to be a preliminary project that needs much more exhaustive research in order to be holistic.

Code and Data

This project does not involve code or datasets. All analysis is based on publicly available secondary sources cited in the references.

Author Contributions (optional)

This paper was researched and written in full by a single author over the course of the Global South AI Safety Hackathon (June 19–21, 2026).

References

1. Mohamed, S., Png, M., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33, 659–684.
2. Chan, A., Okolo, C. T., Turner, Z., & Wang, A. (2021). The Limits of Global Inclusion in AI Development. AAAI Workshop on Reframing Diversity in AI. arXiv:2102.01265.
3. Sambasivan, N., et al. (2021). Re-imagining Algorithmic Fairness in India and Beyond. ACM FAccT 2021. arXiv:2101.09995.
4. Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, 20(4), 336–349.
5. Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. AAAI/ACM AIES 2019.

6. Cihon, P., Maas, M. M., & Kemp, L. (2020). Should Artificial Intelligence Governance be Centralised? AAAI/ACM AIES 2020. arXiv:2001.03573.
7. Gandhi, H., Cihon, P., Larter, R., & Anselmetti, M. (2025). Societal Capacity Assessment Framework (SCAF). ICML TAIG Workshop 2025. arXiv:2509.22742.
8. Noetel, M., et al. (2025). Prioritisation of Risks from AI: MIT AI Risk Repository Delphi Study. MIT AI Risk Initiative. airisk.mit.edu.
9. Shelby, R., et al. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. AAAI/ACM AIES 2023.
10. Ramesh, A., Kameswaran, V., et al. (2022). How Platform-User Power Relations Shape Algorithmic Accountability. ACM FAccT 2022.
11. AGILE Index (2025). AI Governance International Evaluation. arXiv:2507.11546.
12. Jeena Joseph (2025). Algorithmic Bias in Public Health AI: A Silent Threat to Equity in Low-Resource Settings. *Frontiers in Public Health*.
13. Internet Freedom Foundation (2023). Facial Recognition Technology in India. internetfreedom.in.
14. Amnesty International (2024). Samagra Vedika Investigation.
15. Human Rights Watch (2026). Platform Labour Rights: Nine-Country Report.
16. ECRI (2026). Top Patient Safety Concerns 2026.
17. ASEAN (2024). ASEAN Guide on AI Governance and Ethics.
18. European Parliament (2024). EU AI Act — Annex III: High-Risk AI Systems. artificialintelligenceact.eu/annex/3.
19. Karnataka Platform Based Gig Workers (Social Security and Welfare) Act (2025).
20. Vietnam Law on Personal Data Protection (2026). Effective January 1, 2026.
21. Indonesia OJK POJK 29/2024 on Alternative Credit Scoring.
22. ICMR (2023). Ethical Guidelines for Application of AI in Biomedical Research and Healthcare.
23. TechPolicy.Press (2026). India AI Governance Guidelines Analysis.
24. SFLC.in (2026). AI Surveillance and Law Enforcement in India. sflc.in.
25. World Bank (2025). Beyond the AI Divide. Policy Research Working Paper 11073.

LLM Usage Statement

Claude (Anthropic) was used to assist with literature identification and synthesis of secondary sources in addition to formatting, stylistic suggestions and grammar corrections of the final draft. The scoring, framework design, sector selection, dimension rationale, and findings are the author's own.