RUPTURING "AI FOR GOOD": A FEMINIST DECOLONIAL THEORETICAL FRAMEWORK FOR ANALYZING AI INTERVENTIONS IN GENDER-BASED VIOLENCE

Lucía Fernanda Mesa Vélez (Upper & lower case, 12 pt Arial Regular)
Justus-Liebig University Giessen

This paper develops a critical theoretical framework for analyzing "AI for Good" interventions addressing gender-based violence (GBV) from feminist and decolonial perspectives. Current research reveals a tension between empirical evidence suggesting artificial intelligence (AI)'s potential benefits for GBV interventions and critical scholarship exposing how these technologies often reproduce harmful power structures. The proposed theoretical framework bridges this divide by enabling rigorous analysis of AI interventions that neither uncritically accepts technological solutionism nor dismisses potential benefits.

The analytical framework is developed through three interconnected methodological steps. First, I synthesize theoretical concepts from decolonial approaches to technology, such as the colonial matrix of power (Quijano, 1992), data colonialism (Couldry & Mejias, 2018), technocolonialism (Madianou, 2025), resistance and re-existense (Ricaurte, 2023) and the pluriverse (Escobar, 2018); as well as from black, community and decolonial feminism, such as the matrix of domination (Collins, 2002), the modern/colonial gender system (Lugones, 2007), situated knowledges (Haraway, 1988), the ethics of care (Gray & Witt, 2021), and *cuerpo-territorio* (Cabnal, 2010), to construct analytical lenses appropriate for examining these technologies. Based on those theoretical approaches, I conduct a critical discourse analysis (CDA), using Fairclough's (2013) proposal strengthened by feminist (Lazar, 2007) and decolonial (Maniglio & Silva, 2021) approaches, of three corpora: (1) "AI for Good" literature, (2) existing AI ethics and governance frameworks, and (3) meta-analyses of GBV interventions and empirical studies. Finally, I operationalize the theoretical concepts based on the gaps found through the CDA into a framework practical to analyze GBV and other "AI for Good" interventions.

Recent empirical studies demonstrate several promising applications of AI to address GBV. Natural language processing techniques effectively classify GBV content in news articles (Manzoor et al., 2022), AI-powered chatbots show potential for improving

access to domestic violence support services (Izaguirre Choperena et al., 2024), and AI approaches have demonstrated capacity to reduce judicial bias over medium to long timeframes, though with mixed short-term effects (Javed and Li, 2025; Feldfeber et al., 2022). However, these technological interventions exist within broader sociopolitical contexts marked by historical power asymmetries and gender norms. Those conditions permeate the funding, design, training and deployment of AI systems and have several implications, from the consent of the data used for training (Zuboff, 2019) to the labor conditions of the workers (Gray & Suri, 2019), the environmental impact of data centers (Strubell et al., 2019), and their discriminating outputs (Benjamin, 2019; Noble, 2018). The rapid proliferation of "AI for Good" interventions has occurred without adequate critical examination of those implications. Given that their users and material enablers are among the most vulnerable people, communities, and territories in the world, it is especially imperative to assess the impacts of these interventions and put guardrails in place to ensure their safety.

In this context, feminist and decolonial approaches to technological developments offer analytical tools to examine critically the implications of such AI interventions. Decolonial analyses expose how AI interventions often impose Eurocentric values over local knowledge; what Madianou (2025) calls "technocolonialism," Mohamed et al. (2020) identify as "algorithmic coloniality", and Ricaurte (2022) characterizes as a "necro-technological model" perpetuating oppressive global hierarchies. Meanwhile, feminist perspectives point to design choices that embed gender-based assumptions and hierarchies, as demonstrated by Cookson et al., (2023), D'Ignazio and Klein's (2020) and Costanza-Chock (2018) work on data feminism and design justice. These critical lenses are particularly valuable for interrogating the 'AI for good' discourse, which often function as the epistemological foundation through which the push for AI in every aspect of our public and private lives is justified; and do so through the examination of specific interventions.

The CDA showed that current hegemonic discourse and frameworks show significant limitations in addressing these tensions. Regarding AI for good, what "good" means depends on Global North actors and positions the Majority World as perpetually requiring Northern technological salvation. AI ethics and governance frameworks focus on algorithm performance or abstract notions of fairness, rather than on the relevance of AI interventions, their relevance to the contexts or assigning specific accountability. Those AI interventions aiming at addressing GBV dispossess this violence of its structural character and transform it into data to be collected. Overall, technical knowledge is privileged over the lived experiences of those affected.

The proposed theoretical framework addresses these findings through the lens of the decolonial and feminist perspectives to provide a comprehensive approach for analyzing AI interventions addressing GBV. The framework comprises four interconnected analytical dimensions: (1) epistemic foundations; (2) material infrastructure; (3) power distribution; and (4) transformative potential.

First, the epistemological dimension examines whose knowledge shapes these interventions. Drawing on decolonial scholars' call for epistemic diversity (Ricaurte, 2022; Escobar, 2018) and Haraway's (1988) situated knowledges, this dimension

interrogates how AI systems encode particular understandings of gender, violence, social problems and appropriate solutions. It questions whether victim/survivor epistemologies inform technologies without retraumatization or further exploitation.

Second, the material infrastructure dimension builds on Crawford's (2021) work and Madianou's (2025) concept of "infrastructuring" to examine the extractive processes associated with AI interventions. This dimension analyzes who bears the environmental and social costs and who benefits from them, with a special focus on the victims/survivors' bodily and territorial autonomy, the ecological impacts, and the labor relations that enable the intervention.

Third, the power distribution dimension builds on my previous work on feminist decolonial approaches to AI governance (Mesa-Velez, 2025) to analyze how AI interventions operate within and affect intersecting systems of oppression. This dimension examines the decision-making processes, funding structures, and allocation of accountability.

Finally, the transformative potential dimension draws on decolonial feminists' call for resistance and re-existence (Ricaurte, 2023) to examine whether the interventions opens or closes possibilities for transformation. It looks closely into how "change" is understood by those proposing the intervention, if the structural causes of the problem are considered and if it creates conditions for alternative futures beyond the current colonial, extractive, oppressive one.

This is certainly not the first attempt to address the interconnected systems of oppression that are reproduced by artificial intelligence. Critical scholars, activists and communities have proposed approaches such as the aforementioned "Data Feminism" (D'Ignazio and Klein, 2020) and "Design justice" (Costanza-Chock, 2018), as well as the "Decolonial AI Manyfesto" (Kirshnan et al., 2022), "Not my AI" (Varon & Peña, 2022), and "CARE Principles for Indigenous Data Governance" (Global Indigenous Data Alliance, 2018). This is my attempt to address their collective concerns and proposals with a single pragmatic tool that was developed through a rigorous methodology and rich theoretical and political structure.

By developing this analytical framework, my hope is to generate new knowledge about the sociotechnical assemblages through which AI interventions that intend to address GBV, in the broader "AI for Good" context, operate. This is also the first stage and theoretical groundwork of a broader research project that enquires about the extent to which AI can address social justice issues such as GBV. The framework will be applied to various cases across the Majority World in the next stage. The larger project aims to reimagine technological engagement with gender-based violence that centers survivor knowledge, community sovereignty, and transformative justice rather than technological solutionism.

References

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. John Wiley & Sons.

Cabnal, L. (2010). Feminismos Diversos: El feminismo comunitario. https://porunavidavivible.wordpress.com/wp-content/uploads/2012/09/feminismos-comunitario-lorena-cabnal.pdf

Collins, P. H. (2002). Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment (2nd ed.). Routledge. https://doi.org/10.4324/9780203900055

Cookson, T. P., Fuentes, L., Zulver, J. M., & Nelson, A. (2023). Fit for purpose? Assessing the accessibility, theory of action, and accountability of digital technology interventions for sexual and gender-based violence prevention and response. Gender, Technology and Development, 27(2), 184–206. https://doi.org/10.1080/09718524.2023.2182035

Costanza-Chock, S. (2018). Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice (SSRN Scholarly Paper No. 3189696). Social Science Research Network. https://papers.ssrn.com/abstract=3189696

Couldry, N., & Mejías, U. A. (2018). Data Colonialism- Rethinking Big Data's Relation to the Contemporary Subject. Television & New Media, 20(4), 336–349.

Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press. https://doi.org/10.2307/j.ctv1ghv45t

D'Ignazio, C., & Klein, L. F. (2020). Data Feminism. MIT Press.

Escobar, A. (2018). Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds. Duke University Press.

Fairclough, N. (2013). Critical Discourse Analysis: The Critical Study of Language (2nd ed.). Routledge. https://doi.org/10.4324/9781315834368

Feldfeber, I., Quiroga, Y. B., Guevara, C., & Felice, M. C. (2024). Feminisms in Artificial Intelligence: Automation Tools towards a Feminist Judiciary Reform in Argentina and Mexico | AymurAI. Feminist AI. https://feministai.pubpub.org/pub/z83eyq54/release/1

Global Indigenous Data Alliance. (2018). CARE Principles of Indigenous Data Governance. Global Indigenous Data Alliance. https://www.gida-global.org/care

Gray, J. E., & Witt, A. (2021). A feminist data ethics of care framework for machine learning: The what, why, who and how. First Monday, 26(12). https://firstmonday.org/ojs/index.php/fm/article/view/11833

Gray, M. L., & Suri, S. (2019). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Houghton Mifflin Harcourt.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies, 14(3), 575. https://doi.org/10.2307/3178066

Izaguirre Choperena, A., López Belloso, M., & Sanz Urquijo, B. (2024). Empowering Change: Unveiling the Synergy of Feminist Perspectives and AI Tools in addressing Domestic Violence. Communication Papers. Media Literacy and Gender Studies, 13(27), Article 27. https://doi.org/10.33115/udg_bib/cp.v13i27.23087

Javed, K., & Li, J. (2025). Bias in adjudication: Investigating the impact of artificial intelligence, media, financial and legal institutions in pursuit of social justice. PLOS ONE, 20(1), e0315270. https://doi.org/10.1371/journal.pone.0315270

Krishnan, A., Abdilla, A., Moonx, A. J., Çetin, B., Souza, C. A., Adamson, C., Lach, E. M., Ghazal, F., Fjeld, J., Taylor, J., Havens, J. C., Jayaramx, M., Morrow, M., Rizk, N., Ricaurte, P., Chatila, R., Dotan, R., Mhlambi, S., Jordan, S., & Rosenstock, S.

(2022). AI decolonial Manyfesto. Feminist AI.
https://feministai.pubpub.org/pub/q1go9pgy/release/1

Lazar, M. M. (2007). Feminist Critical Discourse Analysis: Articulating a Feminist
Discourse Praxis. Critical Discourse Studies, 4(2), 141–164.
https://doi.org/10.1080/17405900701464816

Lugones, M. (2007). Heterosexualism and the Colonial / Modern Gender System.
Hypatia, 22(1), 186–209.

Madianou, M. (2025). Technocolonialism: When Technology for Good is Harmful. Polity
Press.
https://www.wiley.com/en-us/Technocolonialism%3A+When+Technology+for+Good
+is+Harmful-p-9781509559022

Maniglio, F., & Silva, R. B. da. (2021). El Análisis Crítico del Discurso y el giro
decolonial ¿Por qué y para qué? Critical Discourse Studies, 18(1), 156–184.

Manzoor, M. A., Hassan, S.-U., Muazzam, A., Tuarob, S., & Nawaz, R. (2023). Social
mining for sustainable cities: Thematic study of gender-based violence coverage in
news articles and domestic violence in relation to COVID-19. Journal of Ambient
Intelligence and Humanized Computing, 14(11), 14631–14642.
https://doi.org/10.1007/s12652-021-03401-8

Mesa-Vélez, L. (2025). Towards a Feminist Decolonial Governance of AI: Epistemic
Justice for the Majority World. In P. Hacker (Ed.), Oxford Intersections: AI in Society
(p. 0). Oxford University Press. https://doi.org/10.1093/9780198945215.003.0043

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as
Sociotechnical Foresight in Artificial Intelligence. Philosophy and Technology, 33(4),
659–684. https://doi.org/10.1007/s13347-020-00405-8

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism.
In Algorithms of Oppression. New York University Press.
https://www.degruyterbrill.com/document/doi/10.18574/nyu/9781479833641.001.00
01/html

Quijano, A. (1992). Colonialidad y Modernidad/Racionalidad. Perú Indígena, 13(29),
11–20.

Ricaurte, P. (2022). Artificial Intelligence and the Feminist Decolonial Imagination [Bot
Populi].
https://botpopuli.net/artificial-intelligence-and-the-feminist-decolonial-imagination/

Ricaurte, P. (2023). Resistencia como reexistencia: La defensa del cuerpo-territorio en
la sociedad algorítmica. Revista Pléyade, 32, 64–92.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for
Deep Learning in NLP (No. arXiv:1906.02243). arXiv.
https://doi.org/10.48550/arXiv.1906.02243

Varon, J., & Peña, P. (2022). Not My AI: Towards Critical Feminist Frameworks to Resist
Oppressive AI Systems (No. 10; Carr Center Discussion Paper). Carr Center for
Human Rights Policy. Harvard Kennedy School.
https://www.hks.harvard.edu/centers/carr/publications/not-my-ai-towards-critical-fem
inist-frameworks-resist-oppressive-ai

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at
the New Frontier of Power: Barack Obama's Books of 2019. Profile Books.