



PROJECT PROPOSAL

Towards Decolonial Artificial Intelligence

PhilTech Research Center for the Philosophy of Technology
Department of Philosophy, University of Milan
<https://philtech.unimi.it>

Main Proponents:

Maria Sapignoli maria.sapignoli@unimi.it,
Giuseppe Primiero giuseppe.primiero@unimi.it
Project duration: 24 months

Project Description

Artificial Intelligence is at the forefront of technological innovations across all professional and social domains, from automation to healthcare, education to the environment, sparking heated debates about its impact on society. There is a growing recognition that AI systems often perpetuate and amplify existing biases and inequalities, reflecting the values and perspectives of dominant cultures and societies.¹ In response to this last issue, the concept of decolonial AI has emerged, seeking to challenge and transform these biases by centring marginalized voices and perspectives in the development and deployment of AI technologies.

¹ See **Eubanks, V.** (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press; **Benjamin, R.** (2019). *Race after technology*. Massachusetts: Polity Press; **Noble S. U.** (2018) *Algorithms of Oppression. How Search Engines Reinforce Racism*, New York: New York University Press

In particular, the emergence of Generative AI and Large Language Models has brought a renewed focus on the different principles and models guiding the responsible development and use of AI. Centered on fairness, accountability, and transparency, the debates have led to the development of various sets of principles for defining *Responsible AI*, *Explainable AI*, *Trustworthy AI*. Furthermore, recent efforts to tackle bias and discrimination within AI systems, particularly for marginalized populations, have focused on addressing issues of AI and social inequities, data equity, data justice, and algorithmic discrimination.² In this light, studies have begun to explore the relevance of decolonial theories to the development of AI and related systems.³ Literature about decolonizing AI has emerged over the past decade addressing a range of issues from decolonizing technology,⁴ data,⁵ digital content and IT design,⁶ decentering technology,⁷ and increasing diversity in AI.⁸

These approaches have a social justice orientation which recognises that systemic biases perpetuated by algorithms mainly impact marginalised populations, reinforcing current manifestations of historical racial, gendered, and economic injustices. They reiterate the socio-technical nature of artificial intelligence systems, highlighting the cultural and ethical dimensions, and advocating for a more inclusive and balanced approach through decolonising values, decentring technology,

² See **Eubanks, V.** (2018); **World Economic Forum.** (2023). *Data Equity: Foundational Concepts for Generative AI*. World Economic Forum; **Peña Gangadharan, S., Niklas, J.** (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7), 882-899; **Kong, Y.** (2022). Are “intersectionally fair” ai algorithms really fair to women of color? a philosophical analysis. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. pp. 485-494; **Myers West, S.** (2020). Redistribution and Recognition: A Feminist Critique of Algorithmic Fairness, *Catalyst: Feminism, Theory, Technoscience Issue 6* (2):1-24; **Rodrigues, R.** (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005.

³ See **Mohamed, S., Png, M. T., Isaac, W.** (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659-684; **Birhane, A., Talat, Z.** (2023). It’s incomprehensible: on machine learning and decoloniality. In *Handbook of Critical Studies of Artificial Intelligence*, Lindgren, S. (ed). Massachusetts: Edward Elgar Publishing, pp. 1-19; **Couldry, N., Mejias, U. A.** (2023). The decolonial turn in data and technology research: what is at stake and where is it heading? *Information, Communication & Society*, 26(4), 786–802; **Chi, N., Lurie, E., Mulligan, D. K.** (2021). Reconfiguring diversity and inclusion for AI ethics. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 447-457.

⁴ See **Ali, S. M.** (2016). A brief introduction to decolonial computing. *XRDS: Crossroads, The ACM Magazine for Students*, 22(4), 16–21; **Bird, S.** (2020). Decolonising Speech and Language Technology. In D. Scott, N. Bel, C. Zong (Eds.), *COLING 2020 - 28th International Conference on Computational Linguistics*, Proceedings of the Conference, 1 ed., pp. 3504-3519.

⁵ See **Walter, M., Suina, M.** (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22(3), 233-243; **Magallanes-Blanco, C., Couldry, N., Mhlambi, S., Nothias, T., Nyabola, N. and Siminyu, K.** (2023). *Decolonizing Data, One Language at a Time*. Digital Civil Society Lab, Stanford University.

⁶ See **Nothias, T.** (2023). How to Fight Digital Colonialism. In *Imaging Global Futures*, Getachew, A. (ed) Boston Review; **Tunstall, E.** (2023) *Decolonizing Design. A Cultural Justice Guidebook*. The MIT Press; **Ali, S. M.** (2016); **Irani, L., Vertesi, J., Dourish, P., Philip, K., and Grinter, R.** (2010). Postcolonial computing: A Lens on Design and Development. *Proceedings of the SIGCHI Conference on Human Factors in Computing* pp. 1311–1320.

⁷ See **Peña Gangadharan, S., Niklas, J.** (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7), 882-899.

⁸ See **Baradaran, A.** (2024). Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. *AI & SOCIETY*, 39(1), 7-19; **Cachat-Rosset, G., Klarsfeld A.,** (2023) Diversity, Equity, and Inclusion in Artificial Intelligence: An Evaluation of Guidelines, *Applied Artificial Intelligence*, 37:1, 2176618,

decentralising applications and infrastructures, and increasing diversity.⁹

Some of the key elements to consider in developing a decolonial approach to AI include:

- **Data:** the basis of training reflects the biases of its collection and labelling context.
- **Algorithms:** the logic and language of AI systems are the result of historical-cultural processes that have informed the nature of the systems.
- **Cultural Expertise:** the developer class is not very diversified and perpetuates a limited set of values and worldview.
- **Physical infrastructure:** the control of technology in its physical components is centralised and beyond the influence of vulnerable groups.

Fully in line with the aims and scope of the Departmental Project 'Techne. Philosophy, sciences, technologies' (funded within the "Department of Excellence 2023-2027" Scheme of the Ministry of Universities and Research), the proposed research focuses on new philosophical, ethical and socio-anthropological - as well as technological – consequences of the development and use of AI systems. It specifically investigates the conditions under which a given artificial intelligence system can be adopted while respecting ethical, social and cultural principles specific to the context of application. The project will employ an interdisciplinary approach, drawing on insights from social anthropology, critical race theory, feminist theory, postcolonial studies, indigenous epistemologies, and considers its contribution towards system modelling and formal design. The research activities will focus on exploring the principles, methodologies, and implications of AI systems through empirical analyses of case studies where AI has impacted on ethnicity, gender, class, and other axes of marginalization, while investigating the pathways for de-centering and decolonizing these socio-technical assemblages. The researcher will examine how the promotion of culturally diverse values across core AI systems, and the social world in which AI is produced, could lead to more responsible and equitable outcomes for marginalized and vulnerable groups. A critical element will be the development of accountability methods through the formulation of policies and technical tools to verify under which conditions a given AI system has been sufficiently decolonized.

Research Objectives:

1. To examine the theoretical foundations of decolonial AI, including its roots in critical theory,

⁹ Peña Gangadharan, S., Niklas, J. (2019); Baradaran, A. (2024); Birhane, A., Talat Z. (2024).

postcolonial studies, and indigenous knowledge systems.

2. To identify and analyse case studies where AI applications impact on race, gender, class, and other axes of marginalization and/or in which decolonized/decentralized AIs have been developed and the values they promote.
3. In light of the empirical analysis, reconsider the fundamental elements for a decolonized AI in relation to data, algorithms, cultural expertise, physical infrastructures.
4. To explore methodologies for decolonizing AI, such as the formulation of policies and technical tools to verify under which conditions a given AI system can be sufficiently decolonized.

Expected Outcomes:

1. A deeper understanding of the theoretical foundations, technical requirements and practical implications of developing decolonial AI.
2. Recommendations for integrating decolonial principles into AI research, development, and policy.
3. Contributions to the scientific debate about ethics, equity, and justice in AI and technology more broadly through participation to International Conferences and publication in international venues.
4. Contributing to the activities of the PhilTech Research Center and the Associated Labs.

Skills and Expertise:

- MA in artificial intelligence, computer science, sociology, social anthropology, cultural studies, philosophy or related fields.
- A completed PhD (or proof of submission) in artificial intelligence, computer science, sociology, social anthropology, cultural studies, philosophy or related fields, is highly desirable.
- A track record of publications in top-tier conferences and journals is desirable.
- Interest and familiarity with both AI research and frameworks relevant to postcolonial studies
- Desirable proficiency in qualitative and/or quantitative research methodologies, with experience conducting empirical studies and analysing data.

- Strong communication skills, with the ability to collaborate effectively with interdisciplinary teams and engage with diverse stakeholders.

Motivation:

Applicants should be motivated by a commitment to advancing socially responsible AI that addresses the needs and perspectives of diverse communities. They should be excited about the opportunity to contribute to cutting-edge research at the intersection of technology, ethics, and social justice. Applicants should demonstrate a willingness to engage with complex theoretical concepts and apply them to real-world problems, with the goal of making meaningful contributions to both the field of AI and the broader quest for social justice and equity.

We encourage applications from individuals with diverse backgrounds and experiences, including members of historically marginalized groups. We are committed to fostering an inclusive and equitable research environment.

References

- Ali, S. M. (2016). A brief introduction to decolonial computing. *XRDS: Crossroads, The ACM Magazine for Students*, 22(4), 16–21.
- Baradaran, A. (2024). Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. *AI & SOCIETY*, 39(1), 7-19.
- Benjamin, R. (2019). *Race after technology*. Cambridge: Polity Press.
- Bird, S. (2020). Decolonising Speech and Language Technology. In D. Scott, N. Bel, & C. Zong (Eds.), *COLING 2020 - 28th International Conference on Computational Linguistics*, Proceedings of the Conference. 1 ed., pp. 3504-3519.
- Birhane, A., Talat, Z. (2023). It's incomprehensible: on machine learning and decoloniality. In *Handbook of Critical Studies of Artificial Intelligence*, Lindgren, S. (ed). Massachusetts: Edward Elgar Publishing, pp. 1-19.
- Cachat-Rosset G., Klarsfeld A. (2023) Diversity, Equity, and Inclusion in Artificial Intelligence: An Evaluation of Guidelines, *Applied Artificial Intelligence*, 37:(1).
- 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*, pp. 447-457.

- Couldry, N., Mejias, U. A. (2023). The decolonial turn in data and technology research: what is at stake and where is it heading? *Information, Communication & Society*, 26(4), 786–802.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.
- Irani, L., Vertesi, J., Dourish, P., Philip, K., and Grinter, R. (2010). Postcolonial computing: A Lens on Design and Development. Proceedings of the SIGCHI *Conference on Human Factors in Computing*, pp. 1311–1320.
- Kong, Y. (2022). Are “intersectionally fair” ai algorithms really fair to women of color? a philosophical analysis. Proceedings of the 2022 ACM. *Conference on Fairness, Accountability, and Transparency*, pp. 485-494.
- Magallanes-Blanco, C., Couldry, N., Mhlambi, S., Nothias, T., Nyabola, N. and Siminyu, K. (2023). *Decolonizing Data, One Language at a Time*. Digital Civil Society Lab, Stanford University.
- Mohamed, S., Png, M. T., and Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659-684.
- Myers West, S. (2020). Redistribution and Rekognition: A Feminist Critique of Algorithmic Fairness, *Catalyst: Feminism, Theory, Technoscience Issue*, 6 (2):1-24.
- Noble S. U. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism*, New York: New York University Press.
- Nothias, T. (2023). How to Fight Digital Colonialism. In *Imaging Global Futures*, Getachew, A. (ed) Boston Review.
- Peña Gangadharan, S., Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7), 882-899
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges, and vulnerabilities. *Journal of Responsible Technology*, 4, 100005.
- Tunstall, E. (2023). *Decolonizing Design. A Cultural Justice Guidebook*. The MIT Press.
- Walter, M., Suina, M. (2019). Indigenous data, indigenous methodologies, and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22(3), 233-243.
- World Economic Forum. (2023). *Data Equity: Foundational Concepts for Generative AI*. World Economic Forum. <https://www.weforum.org/publications/data-equity-foundational-concepts-for-generative-ai/>.