



**Selected Papers of #AoIR2021:
The 22nd Annual Conference of the
Association of Internet Researchers**
Virtual Event / 13-16 Oct 2021

DISEMBEDEDNESS IN MACHINE LEARNING DATA WORK

Julian Posada
University of Toronto

Machine learning algorithms rely on mathematical models that learn from massive amounts of data, or “training data,” to make automated decisions. This type of artificial intelligence has become widely used in the last decade due to an abundant data produced by increased Internet connectivity (Alpaydin, 2020) and pervasive data collection methods (Zuboff, 2019). Firms and research organizations require humans to annotate this training data to make it compatible with machine learning algorithms or calculation processes (Gray & Suri, 2019; Posada, 2020). Data workers worldwide train, verify, and even impersonate algorithms through digital labor platforms (Tubaro et al., 2020). These platforms are “(re-)programmable digital infrastructures that facilitate and shape personalized interactions among end-users and complementors” (Poell et al., 2019) and serve as marketplaces where labor is exchanged as a commodity. Like in other gig economy platforms, firms that operate these marketplaces for machine learning development consider workers as “independent contractors,” paying them few cents per task, and denying them any recognition, rights, and social protections (Prassl, 2018), while placing systems of surveillance and control (Casilli, 2019).

This paper draws from decolonial theory (Mohamed et al., 2020), theories on social and economic embeddedness (Tubaro, 2021; Wood et al., 2019), and the political economy of platforms (Casilli & Posada, 2019; van Dijck et al., 2019). It studies how Latin American data workers are “embedded and enmeshed in institutions, economic and non-economic” (Polanyi, 2001) and how this situation affects social reproduction from the perspective of social structures and institutions (Bourdieu & Passeron, 1970) as well as forms of gendered and embodied labor (Bhattacharya, 2017; Huws, 2020). The paper uses data collected through the study of four digital labor platforms specialized in machine learning development that employs Latin American data workers. Since there is little known about the composition of this invisibilized population, the platforms were studied by analyzing web traffic data, the companies’ documentation, and the instructions for data annotation and algorithmic verification. The data workers were reached through an online survey sent through the platforms and a series of in-depth

Suggested Citation (APA): Posada, J. (2021, October). *Disembeddedness in Machine Learning Data Work*. Paper presented at AoIR 2021: The 22nd Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

semi-structured interviews with platform workers and members of their social circles to identify their background, their experience with online work, the composition of their personal networks, and how their work interacts, intersects, and depends on their social connections.

The data analysis suggests the continuation of long historical patterns of domination in how the platform labor market is configured from two levels. At the platform level, a geographical analysis of the web traffic from the platforms shows the continuation of a north-south divide in the distribution of work present in other forms of online work such as freelancing, where the demand for labor comes mainly from advanced economies and the supply from countries in the global south (Graham et al., 2017). While historically most of the workers in the global south come from African and Asian countries, in the case of data work for machine learning, there is more traffic coming from Latin America and especially Venezuela, a country currently experiencing a severe political and economic crisis, where most of the interviewed and surveyed workers come from.

At the worker level, the analysis of the instructions and the workers' interviews suggests that platforms' algorithmic management constraints their judgment and their labor process. These intermediaries compel them to reproduce the categorization of datasets according to the ideological preferences of requesters, even if they do not always align with the worldviews of workers. The analysis of the data from the interviews also suggests that, despite the individualized and alienated nature of platform labor, workers are also embedded in networks of trust within households, online, and in local communities, that provide social and economic support the "disembedded" markets of data work, which is currently unconstrained by government regulations (Tubaro, 2021; Wood et al., 2019). Thus, instead of enduring the problems of labor commodification alone, the support of workers' social networks plays a vital role in their social and working experience.

These findings show a continuation of exploitative supply chains in the current artificial intelligence market (Posada et al., 2021), where wealthy companies and research institutions in advanced economies profit from the economic and political situation of developing countries to access disembedded labor. From the perspective of institutional and structural reproduction, the design of crowdsourcing platforms and their configuration of the labor process provides evidence of a continuation of indigenous knowledge suppression by those in power positions and the imposition of their worldviews to individuals from exploited communities (Maldonado-Torres, 2007). At the same time, from the perspective of embodied social reproduction, while experiencing high degrees of disembeddedness as "independent contractors," data workers, mostly identifying as male, rely on the domestic, economic, and emotional support of their

families, friends, and communities to compensate for the social and economic risks of their primary source of income.

The paper concludes that contemporary data work continues centuries-long exploitative relations that are detrimental for both the development of nations and communities in the global south and the plural and ethical development of machine learning systems. The data workers necessary for the training and verification of algorithms are not given opportunities to grow professionally and contribute to their local economies by placing them at the margins of the data supply chains for artificial intelligence. Their voices and an improvement in their working conditions are necessary for AI to be truly able to serve the public good.

References

- Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). MIT Press.
- Bhattacharya, T. (Ed.). (2017). *Social Reproduction Theory. Remapping Class, Recentering Oppression*. Pluto Press.
- Bourdieu, P., & Passeron, J.-C. (1970). *La Reproduction. Éléments pour une théorie du système d'enseignement*. Les éditions de minuit.
- Casilli, A. A. (2019). *En attendant les robots*. Éditions du Seuil.
- Casilli, A. A., & Posada, J. (2019). The Platformisation of Labor and Society. In M. Graham & W. H. Dutton (Eds.), *Society and the Internet* (Vol. 2). Oxford University Press.
- Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*, 23(X), 1–28. <https://doi.org/10.1177/1024258916687250>
- Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.
- Huws, U. (2020). Social Reproduction in Twenty-First Century Capitalism. *Socialist Register*.
- Maldonado-Torres, N. (2007). On the coloniality of being: Contributions to the development of a concept. *Cultural Studies*, 21(2–3), 240–270. <https://doi.org/10.1080/09502380601162548>
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Poell, T., Nieborg, D. B., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1425>
- Polanyi, K. (2001). *The Great Transformation: The Political and Economic Origins of Our Time* (2nd Ed.). Beacon Press.
- Posada, J. (2020). The Future of Work Is Here: Toward a Comprehensive Approach to

Artificial Intelligence and Labour. *Ethics in Context* ,56.
<https://arxiv.org/abs/2105.10990>

- Posada, J., Weller, N., & Wong, W. H. (2021). We Haven't Gone Paperless Yet: Why the Printing Press Can Help Us Understand Data and AI. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*.
- Prassl, J. (2018). *Humans as a Service: The Promise and Perils of Work in the Gig Economy*. Oxford University Press.
- Tubaro, P. (2021). Disembedded or Deeply Embedded? A Multi-Level Network Analysis of Online Labour Platforms. *Sociology*. <https://doi.org/10.1177/0038038520986082>
- Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720919776>
- van Dijck, J., Nieborg, D. B., & Poell, T. (2019). Reframing platform power. *Internet Policy Review*, 8(2), 1–18. <https://doi.org/10.14763/2019.2.1414>
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Networked but Commodified: The (Dis)Embeddedness of Digital Labour in the Gig Economy. *Sociology*. <https://doi.org/10.1177/0038038519828906>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.