LEVERAGING OR EXPLOITING? BREAKING THROUGH DATA COLONIALISM AND ETHONCENTRISM WHEN BUILDING A DIVERSITY-AWARE SOCIAL PLATFORM

Paula Helm
University of Amsterdam

Amalia De Goetzen
Aalborg Univeristy Copenhagen

Luca Cernuzzi
Universidad Católica "Nuestra Señora de la Asunción"

Shyam Diwakar
Amrita University

Alethia Hume
Universidad Católica "Nuestra Señora de la Asunción"

Salvador Ruiz-Correa
Instituto Potosino de Investigación Científica y Tecnológica (IPICYT)

Niels Jørgen Gommesen
Aalborg University Copenhagen

Daniel Gatica-Perez
IPIAP Research Institute, Ecole Polytechnique de Lausanne

In recent years, there has been growing concern about the negative impacts of the extractive logic of data-driven technology and knowledge production. While the initial focus was on the impact on Western societies, scholars are increasingly concerned with the consequences for countries and communities in the Global South. Some commentators have gone so far as to describe the prevailing dynamics as neocolonial

and imperialist (Arora, 2019; Milan & Treré, 2019; Philip et al., 2012; Taylor & Broeders, 2015). Some go as far as describing the dynamics at play as neocolonial and imperialistic (Varon & Peña, 2021; Zembylas, 2021). While neocolonial dynamics are most evident regarding powerful private sector companies, they do extend to the realm of publicly funded research projects between partners from Souths and Norths countries. Here, it can be observed that while the initial idea may be benevolent, the benefits generated often end up flowing in only one direction. This can impoverish research in the disadvantaged country as experience and data are siphoned off and no contribution is made to local capacity and infrastructure (Armenteras, 2021). This practice has been named "helicopter," "parachute," or downright "colonial" science (Belhabib, 2021).

The present paper (self-)critically examines these dynamics by means of drawing on the experiences collected during one EU-funded research and innovation project that all authors have been involved with. The project that inspired this paper can be diagnosed at operating at precisely the intersection of benevolent but unequal and potentially neocolonial transnational scientific collaboration. It is set in the context of data-driven technology and knowledge production. In this paper, we present a spin-off study departing from the official research plan. In this spin-off study we reflect on and analyze the challenges, (missed) opportunities, and hurdles that we encountered in our collaboration. The authors are diverse. They include cultural studies and ethics scholars, data scientists, designers, and computer scientists, among them are full partners from EU countries (Denmark, Germany, Switzerland) and associated international partners (India, Mexico, Paraguay).

The basic idea of the research project bringing us together was to create a digital platform that connects users who want to solve complex tasks or answer questions by leveraging the diversity of their communities. The matching is based on profile data and self-learning algorithms.

By focusing on leveraging diversity, the WeNet-project responds to recent criticism of so- called "filter bubble" and "echo chamber" effects (Helberger et al., 2018; Pariser, 2011). In contrast to similarity-based matching algorithms, the goal of the WeNet-project is to develop a decidedly diversity-aware platform that enables machine-mediated social interactions between individuals who differ in their traits and competencies and thus complement each other productively (Helm et al., 2021). This is where our conceptual approach comes in, going beyond traditional understandings of diversity and enriching flat demographic categories with sociological and psychological ones. Data-based profiling is key to this approach. User profiles are therefore populated with data collected through self-reporting, mobile sensors, and geolocation. Terabits of clean and reusable data have already been obtained, and a platform is in place. Several pilot studies are underway. The longer-term goal consists of algorithmically analyzing user queries based on continuously updated profiles in order to identify suitable respondents in an automated way.

The users in the pilots are students from around the world. Pilot sites include China, Denmark, Great Britain, India, Italy, Mexico, Mongolia, and Paraguay. While the leaders of the European pilot sites receive full funding, this is not the case for the Souths

partners. This circumstance makes the situation substantially imbalanced. The imbalance does not only concern the cultural diversity at play, which makes the appropriateness of a one- size-fits-all design developed from a European perspective doubtful (this refers to what in the paper we discuss under "ethnocentric design"). Moreover, power asymmetries between European and international partners are also not trivial given the economic and epistemological value of the generated datasets operated and stored by the European partners but derived from international pilot sites (this refers to what in the paper we discuss under "data colonialism").

In this constellation, WeNet finds itself in an ambivalent position that holds both risks and potentials. On the one hand, the project partners are themselves caught up in the extractive logic of data capitalism (Sadowski, 2019), thus reproducing certain practices that have become standard in computer science: a scalability mindset as the ordering logic of innovation, prioritizing one-size-fits all solutions over plural design, the unequal distribution of data power, as well as time pressure and competition. On the other hand, by focusing on diversity, choosing an interdisciplinary approach, including ethicists on the core team, and working with partners from the Souths, WeNet simultaneously seeks to develop an inclusive and value-based alternative to existing private sector solutions. Given this tension, in this paper, we address the following questions: *Within the existing EU-funding framework and facing the prevailing pressures in innovation-driven research, what do we need to change to enable more equitable practices? What are the main problems? What activities have proven useful?*

To answer these questions, we proceed as follows: First, we provide an overview of recent studies and conceptual works that address the challenges of transnational research efforts and innovation projects under the conditions of postcolonial technoscience. This serves to clarify the structures in which the project discussed here is situated and needs to be assessed against. Based on the conceptual clarifications, we develop a model, visualizing the interplay of extraction and eurocentrism at the intersection of transnational research and datafication. We then zoom in on four exemplary pilot sites (Denmark, Paraguay, Mexico, India), examining how the project design played out on the ground, reflecting on successes, hurdles, challenges, missed opportunities and analyzing these against the background of the more general situation outlined in the previous section. After presenting the pilots, the various observations are analyzed comparatively. Based on this analysis, we draw some conclusions, that we hope will be instructive for future projects facing similar challenges.

References

Armenteras, D. (2021). Guidelines for healthy global scientific collaborations. *Nature Ecology & Evolution*, *5*(9), 1193–1194.

Arora, P. (2019). *The Next Billion Users: Digital Life Beyond the West*. Harvard University Press.

Belhabib, D. (2021). Ocean science and advocacy work better when decolonized. *Nature Ecology & Evolution*, *5*(6), 709–710.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207.

Helm, P., Michael, L., Schelenz, L. (2022) Diversity by Design? Balancing the Protection and Inclusion of Users in Online Social Networks, in: Conference Proceedings of the ACM/AAI Conference on Artificial Intelligence, Ethics, Society, https://doi.org/10.1145/3514094.3534149.

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television & New Media*, *20*(4), 319–335.

Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin UK.

Philip, K., Irani, L., & Dourish, P. (2012). Postcolonial Computing: A Tactical Survey. *Science, Technology, & Human Values*, *37*(1), 3–29.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549.

Taylor, L., & Broeders, D. (2015). In the name of Development: Power, profit and the datafication of the global South. *Geoforum*, *64*, 229–237.

Varon, J., & Peña, P. (2021). Artificial intelligence and consent: A feminist anti-colonial critique. *Internet Policy Review*, *10*(4).

Zembylas, M. (2021). A decolonial approach to AI in higher education teaching and learning: Strategies for undoing the ethics of digital neocolonialism. *Learning, Media and Technology*, *0*(0), 1–13.