

## Diversity and Neocolonialism in Big Data Research

*Avoiding extractivism while struggling with paternalism*

Helm, Paula; de Götzen, Amalia; Cernuzzi, Luca ; Hume, Alethia; Diwakar, Shyam; Ruiz-Correa, Salvador; Gatica-Perez, Daniel

*Published in:*  
Big Data & Society

*DOI (link to publication from Publisher):*  
[10.1177/20539517231206802](https://doi.org/10.1177/20539517231206802)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

### *Citation for published version (APA):*

Helm, P., de Götzen, A., Cernuzzi, L., Hume, A., Diwakar, S., Ruiz-Correa, S., & Gatica-Perez, D. (2023). Diversity and Neocolonialism in Big Data Research: Avoiding extractivism while struggling with paternalism. *Big Data & Society*, 10(2), 1-14. <https://doi.org/10.1177/20539517231206802>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Diversity and neocolonialism in Big Data research: Avoiding extractivism while struggling with paternalism

Big Data & Society  
July–December: 1–14  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20539517231206802  
journals.sagepub.com/home/bds



Paula Helm<sup>1</sup> , Amalia de Götzen<sup>2</sup> , Luca Cernuzzi<sup>3</sup> ,  
Alethia Hume<sup>3</sup>, Shyam Diwakar<sup>4</sup> , Salvador Ruiz Correa<sup>5</sup>  
and Daniel Gatica-Perez<sup>6</sup>

## Abstract

The extractive logic of Big Data-driven technology and knowledge production has raised serious concerns. While most criticism initially focused on the impacts on Western societies, attention is now increasingly turning to the consequences for communities in the Global South. To date, debates have focused on private-sector activities. In this article, we start from the conviction that publicly funded knowledge and technology production must also be scrutinized for their potential neocolonial entanglements. To this end, we analyze the dynamics of collaboration in an European Union-funded research project that collects data for developing a social platform focused on diversity. The project includes pilot sites in China, Denmark, the United Kingdom, India, Italy, Mexico, Mongolia, and Paraguay. We present the experience at four field sites and reflect on the project's initial conception, our collaboration, challenges, progress, and results. We then analyze the different experiences in comparison. We conclude that while we have succeeded in finding viable strategies to avoid contributing to the dynamics of unilateral data extraction as one side of the neocolonial circle, it has been infinitely more difficult to break through the much more subtle but no less powerful mechanisms of paternalism that we find to be prevalent in data-driven North–South relations. These mechanisms, however, can be identified as the other side of the neocolonial circle.

## Keywords

Data colonialism, paternalism, diversity, Big Data, European Union, research policy

## Introduction

In recent years, there have been growing concerns about the extractive logic of data-driven technology and knowledge production. While the initial focus was on the impacts on Western societies, scholars are increasingly studying the consequences for countries and communities in the Global South.<sup>1</sup> With data being traded as “the new oil,” some commentators have gone so far as to call the prevailing dynamics neocolonial and imperialistic (Ricaurte, 2019; Varon and Peña, 2021; Zembylas, 2021). While imperialist and neocolonial dynamics are most evident with view on the activities of powerful private companies, we argue that publicly funded research projects that bring together partners from the Global South and North also need to be scrutinized through the lens of this critique. By this, we do not mean to say that these projects intentionally pursue extractivist interests (Taylor and Broeders, 2015). Nevertheless, they tend to contribute to or are part of culturally and structurally embedded neocolonialism. For

example, it can be observed that while the main goal in European Union (EU)-funded research needs to be benevolent towards a greater public good, the gains made often end up flowing in mostly one direction (Ricaurte, 2019;

<sup>1</sup>Department of Media Studies, University of Amsterdam, Amsterdam, the Netherlands

<sup>2</sup>Faculty of Engineering and Science, Department of Architecture Design and Media Technology, Aalborg University, Copenhagen, Denmark

<sup>3</sup>Universidad Católica Nuestra Señora de la Asunción in Paraguay, Asunción, Paraguay

<sup>4</sup>Amrita Mind Brain Center, Amrita Vishwa Vidyapeetham, Kollam, TN, India

<sup>5</sup>Instituto Potosino de Investigación Científica y Tecnológica, Potosi, Mexico

<sup>6</sup>Idiap Research Institute and Ecole Polytechnique Fédérale de Lausanne (EPFL), Martigny, Valais, Switzerland

## Corresponding author:

Paula Helm, Department of Media Studies, University of Amsterdam, Turfdragerpad 9, 1012 XT Amsterdam, the Netherlands.  
Email: p.m.helm@uva.nl



United Nations, 2021). This can lead to an impoverishment of research in partner countries outside the EU, as experience and data are siphoned off, and hardly any contribution is made to solving local problems and strengthening local research capacities (Armenteras, 2021). This practice is attributed to what is described as “helicopter,” “parachute,” or downright “colonial” science (Belhabib, 2021; Dalton et al., 2016; Vos, 2020).

The project that inspired this article, in our analysis, operates at precisely this intersection of benevolent but unequal collaboration, struggling with not replicating the above-described practices. It is situated in the context of data-driven technology and knowledge production. In this article, we analyze the dynamics of North–South collaboration in the project, thereby presenting a conceptual analysis that deviates from the official research agenda. This analysis was inspired by a project partner who joined the team at a later stage when most landmark design decisions had already been taken. That partner worked in the capacity of an embedded ethicist. Upon joining the project, the new team member stimulated a (self-)critical conversation about persistent inequalities in collaboration. This conversation involved half of the project partners, including the design team located in Denmark, the data analysis team in Switzerland, and some of the associated partners conducting the pilot studies, including in Paraguay, Mexico, and India. As part of our collective reflection and corresponding background research, we arrived at the conviction that many of our challenges seem symptomatic of some broader trends and issues in current data-driven research. With this in mind, we systematized our experiences towards this joint research paper.

At the heart of the project bringing us together was the vision to create a digital platform that connects users who want to solve tasks or questions by leveraging the diversity of their communities. By focusing on diversity, the WeNet-project responds to recent criticism of the so-called “filter bubble” and “echo chamber” effects (Bruns, 2019; Pariser, 2012). In contrast to similarity-based matching algorithms, the goal of the project was to develop a platform that enables machine-mediated social interactions between individuals who differ in their traits and competencies and may thus complement each other productively. This is where the project’s conceptual approach comes in. This approach goes beyond cosmetic understandings of diversity, enriching flat demographic categories with more profound sociological and psychological ones. Reflecting this understanding, in WeNet, diversity is defined as social practices. These are routines of human behavior that many members of a community enact while still maintaining individual differences. Through extensive representation in data sets, the longer-term goal is to develop self-learning algorithms capable of matching user queries automatically based on constantly updated profiles.

The envisioned approach requires that the diversity of users is reflected in the data. Data-based profiling is hence vital to the project idea. Accordingly, the bulk of the work revolves around populating user profiles with data collected through self-reporting (surveys), mobile sensors, and geolocation (I-Log), as well as by engaging them as test users of a chat application. This application has been implemented in three consecutive pilot studies. As a result, extraordinarily rich and valuable data sets on people’s skills, psychological make up, and behavior are being produced.

The test users are university students. To reflect diversity also in terms of geolocation, pilot sites are spread around the globe, including China, Denmark, Great Britain, India, Italy, Mexico, Mongolia, and Paraguay. This is where a neocolonial critique becomes relevant. While the leaders of the European pilot sites receive full funding for designing and conducting the pilots, this is not the case for the associated partners. Their budget is either non-existent or very small and limited to carrying out the pilots the EU partners have designed. This circumstance makes the situation substantially asymmetric. 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 (Helm et al., 2021). Moreover, power asymmetries between European and Global South partners are also not trivial given the economic and epistemological value of the generated data sets operated and stored by the European partners but derived from international pilot sites.

In this constellation, both the associated- and the EU partners find themselves in a complicated situation, as they are both caught in the logic of data capitalism (Sadowski, 2019). This logic pressures them to reproduce certain practices that have become standard in data-driven research and innovation: scalability thinking, preference for one-size-fits-all solutions over customized ones, time pressure and competition. But there are also ways to counter these trends. By focusing on diversity, taking an interdisciplinary approach, including ethicists in the core team, and collaborating as closely as possible with partners from the Global South, not just in the gathering but also analyzing data, the present project goes to great lengths to leverage existing margins and opportunities to practice a value-centered alternative. Please note that in highlighting our efforts, this article does not go as far as claiming to be part of a broader and more substantive decolonial movement. Instead, it discusses strategies for engaging in data-driven, transnational research constellations in the context of hegemonic structures but in ways that avoid reproducing neocolonial dynamics, which we identify to be consisting first and foremost of a combination of extractivist and paternalistic dynamics.

To do so, we proceed as follows: First, we provide an overview of recent studies and conceptual works addressing

the challenges of transcontinental data-driven research under postcolonial technoscience (Anderson, 2002). This serves to clarify the structures in which the discussed project is situated and needs to be assessed against. Based on the conceptual clarifications, we develop a model, visualizing the interplay of extractivism and paternalism at the intersection of transcontinental research and datafication. We then zoom in on four pilot sites. After presenting the pilots in depth, we offer a comparative analysis of the different observations. Based on this analysis, we reflect on viable strategies found to avoid the hegemonic dynamics of the unilateral data mining as one side of the neocolonial circle. In contrast, we show why it has been infinitely more difficult to break through the much more subtle but no less powerful mechanisms of paternalism, to which we refer as the other side of the neocolonial circle.

## Related works and conceptual background

The following section serves to situate the present project within the larger Big Data and design discourse to add a more nuanced classification to the blunt juxtaposition of the Global North and South. In doing so, we contribute to existing analyses of the logic by which power-saturated data practices influence public financing instruments, corresponding design decisions, private sector strategies, global inequalities, and local dynamics. To this end, we first differentiate between an approach that uses the term data colonialism to describe a new variant of capitalism and one that conceives of data colonialism as a continuation of historical colonialism - by different means. This helps to understand better the meaning ascribed to data in today's world and the standards by which data collection activities are evaluated. Second, we go into more detail about how our funding instrument influenced the design of the project, gave us room to maneuver, but also constrained us, and thus had a significant impact on our collaboration and outcomes. Finally, we address some design concepts that have influenced our approach. These serve as a normative compass, against which we now critically examine our work, than implemented practice.

### *Data capitalism and data colonialism*

The importance of data has changed dramatically with digitization. More specifically, the last three decades have seen an intensification of a long-standing tendency toward the capitalist valorization of information (Castells, 2009). In the private sector, data has become a monetizable resource and, in some cases, even a currency. In science, data have always been indispensable. They form the coagulated empirical foundation upon which truth claims are built. Here, we ask what implications the changing role of data in the private sector has in the way scientific projects are conceived and, more specifically, what implications this

has for North–South partnerships within such projects. If scientific projects are now more than ever measured by not only the quality but also the size of the data sets they produce or gain access to (Staunton et al., 2021): how does this affect the way we work together?

At least to some extent, academic research is implicated in the extractive logic of data capitalism (Sadowski, 2019). This logic has become so aggressive that some argue it can best be understood through the logic of colonialism. To coin this observation, (Thatcher et al., 2016) introduced the term data colonialism, which was taken up by (Couldry and Mejias, 2019). Data colonialism is described here as the normalization of the exploitation of people through data, similar (but not the same!) as historical colonialism appropriated territory and resources and dominated subjects for profit. In this understanding, the “colonizers” are not so much Western imperialist countries invading the South but powerful tech companies exploiting all kinds of users.

Recognizing the need for a nuanced perspective regarding the different socio-institutional contexts of data collection, let us further include those voices that have argued against the instrumentalization of the concept of colonialism in this context. First, it has been pointed out that it is inappropriate to compare the domination of private tech companies to colonial relations. While colonial relations were based on a necropolitics of physical violence and murder, political and economic domination through technology, while undesirable, fortunately, does not involve the same genocidal practices (Benjamin, 2019). Second, the generalized focus has been criticized for downplaying the struggles of marginalized populations and making it seem as if we live in a global data community where there is no difference between the impact on communities in the Global North and South (Arora, 2016). However, a closer look at our project context vividly illustrates that significant differences must be considered, both in the productive sense of diversity and in the destructive sense of power imbalances. To what extent are these differences accounted for in research and innovation collaboration? What measures are useful?

Following the critical view of the North–South axis of inequality, so-called Data for Development projects, which are situated at the intersection of research and development, have also been scrutinized (Daniels et al., 2019; Schelenz and Pawelec, 2022). The project discussed here resembles this category in many ways, especially with regard to the paternalistic attitudes common in this field (Hilbert, 2016). In paternalistic relationships, the authority figure or governing body assumes a position of superiority over those targeted as the object of development. Drawing on political philosophy, in which paternalism refers to a form of social behavior characterized by an authority figure or governing body making decisions or taking actions on behalf of others to promote their welfare or

protect them from harm (Mill, 1859: chapter IV), the concept has been mobilized for a critical analysis of the development sector, which today is increasingly data-driven (Anwaar et al., 2016). Paternalism is problematic here because it undermines the agency of individuals and/or autonomy of communities (Dworkin, 1972) and perpetuates or even reinforces existing power asymmetries between countries in the Global South and North. By the latter generating knowledge about the former and then using that knowledge to tell the former what is best for them, historically grown relationships of dependency are likely to be maintained (Madianou, 2019; Taylor and Broeders, 2015).

Having outlined these interwoven perspectives on the political-economic significance of data with a focus on capitalism on the one hand and neocolonial relations on the other, we proceed from an understanding that we believe combines the analytical advantages of both. While we acknowledge the conceptual advantages of a perspective critical of capitalism, breaking down the North–South dualism and focusing instead on the extractive and invasive logic of a particular kind of data economy that arguably does not stop at scientific enterprise and extends its demands to the public sector, we also believe that we need to pay particular attention to the North–South axis of inequality. Against this backdrop, we now examine the EU-funding instruments, which have played an important role in shaping our collaboration and which we believe cannot be evaluated in isolation from their postcolonial baggage.

### *Intercontinental collaborations in the EU-context*

EC-funded research undoubtedly has numerous advantages. Yet, it also comes with constraints, especially regarding North–South collaboration. In the present project, we took it as our challenge to identify margins and opportunities that can be leveraged to make collaboration as equal as possible.

Regarding advantages, the commitment to European values is most notable (European Commission, 1999). This commitment entails a range of practical consequences that imply strong protections for research participants and the affected public. This is especially pertinent since the General Data Protection Regulation came into force (European Commission, 2019). In addition, there is the public benefit orientation that must be demonstrated. Finally, numerous quality assurance mechanisms are in place to prevent fraud, manipulation of results, as well as nepotism among scientists and reviewers. All of these are genuine benefits, but they also operate on a safe harbor model, which implies that they are geared toward the European population but do not always necessarily benefit the rest of the world (Arora, 2019; Wetzling et al., 2021). Furthermore, it must be noted that the EU's 'pro-active equality action' requirements relate solely to

analyzing the impacts of a given project on women but not on, as clearly relevant in the given example, associated partners from the Global South and their communities.

A major obstacle to equitable collaboration is that most EC-funded research projects are carried out with a short-term time horizon. This is not only a problem for interdisciplinary research projects (Felt, 2022), but also for intercontinental ones, as meeting the needs of different locations requires conducting multiple in-depth studies over longer periods involving local experts. This requires time to coordinate and share data, results, and ideas. In addition, intercontinental projects need to accommodate different time zones, vacation schedules, and greater distances. Not to mention the sensitivity it takes to deal with cultural diversity. In this context, it is not only the EC-funding context that exerts pressure but also the incentive mechanisms for researchers at EC universities that expect multiple parallel projects to be solicited and that judge their staff on individual performance rather than collective achievements (Müller, 2012).

The biggest obstacle to collaboration at eye level is the asymmetrical funding situation. Although collaboration with associated partners from non-EU countries is explicitly desired (e.g. to promote data diversity and increase global impact), they can usually only receive limited funding, if any European Commission (2022). This, in turn, depends on bilateral agreements between the EU and non-EU countries. In our context, this has led not only to inequalities between EU partners and associated partners but also to inequalities among non-EU partners, with some receiving support and others not. These inequalities are clearly reflected in the quality of the results achieved. Of those that did receive funding, the money was generally allocated exclusively to operational activities such as data collection and travel to project meetings. At the same time, the actual research and analysis was not supported. Another problem is the uneven distribution of authority. While the associated research partners can actively participate in project consortium discussions, thereby contributing to decision-making processes, they can do so only unofficially since, ultimately, "the consortium is responsible for the proper execution of the tasks carried out by the associated partners (proper quality, timely delivery, etc.)" (European Parliament, 2021: 99). While this negation of authority and responsibility on the part of the associated partners is plausible from an insurance perspective, it is also highly problematic when viewed in a broader historical context. Indeed, as we will argue below, this constellation makes it almost impossible not to reproduce power asymmetries and paternalistic relations that still exist between Europe and its former colonies.

On a more general level, requirements to demonstrate the economic relevance of research and innovation projects are often in tension with considering the different sensitivities, worldviews, vulnerabilities, and constraints of the

communities involved and/or addressed (both inside and outside the EU). This is primarily because economic relevance is often understood as synonymous with smooth market integration based on rapid up-scaling potential (Pfotenhauer et al., 2021). Scaling data sets to promote diversity in automated matching may well be beneficial, such as in creating richer data sets that enable better/fairer performance of collaborative filtering (Deldjoo et al., 2023). However, rapid demands for upscaling are detrimental for a project that seeks to work on par with diverse communities if it is used to replicate the same pilot (ethics, rules, processes, forms of engagement, etc.) to be more efficient and create comparable data sets and thus use cases that can be abstracted from local contexts. To clarify the difference, Anna Tsing differentiates between scalable and meaningful diversity, observing a tendency for the former to often be privileged in the face of prevailing industry and market demands. In contrast, she argues that the value of “non-scalable diversity,” that is, “diversity that changes things,” should be weighted more heavily than scalable diversity, where only that which can easily fit into existing designs and standards is accepted (Tsing, 2012). Her argument is part of a more general debate that emphasizes the situatedness of knowledge and problematizes universalist claims and their broader implications (Haraway, 1988; Harding, 1995; Jasanoff, 2006; Tsing, 2004). Although these anti-universalist arguments are not limited to North–South research collaborations, they are particularly relevant in this context.

Rather than emphasizing the importance of working toward locally significant but small-scale solutions that are co-designed by culturally and politically diverse communities and thus reflective of their needs, recent EC research programs have been flagged in the name of “grand societal challenges.” This way of formulating research agendas implies a preference for scalable top-down solutions, whether intentionally or not. Projects like the one discussed here fall under the category of technology-assisted “social innovation,” which traditionally focuses on bottom-up dynamics and local alternatives rather than huge technocratic interventions (Musa and Rodin, 2016). As such, they should be evaluated and not made dependent on market-driven scaling imperatives, as is not exclusively but partly the case in current EC evaluation practices.

### *Ethnocentric versus pluriversal design*

Seeking to overcome the extractivist and paternalistic logics of data-driven research and innovation, currently dominating private sector activities, in our joint project work, we self-critically ask ourselves: What alternative modes of project, technology, and service design are available, in how far are we taking them into account and if not, what might be the reasons for that?

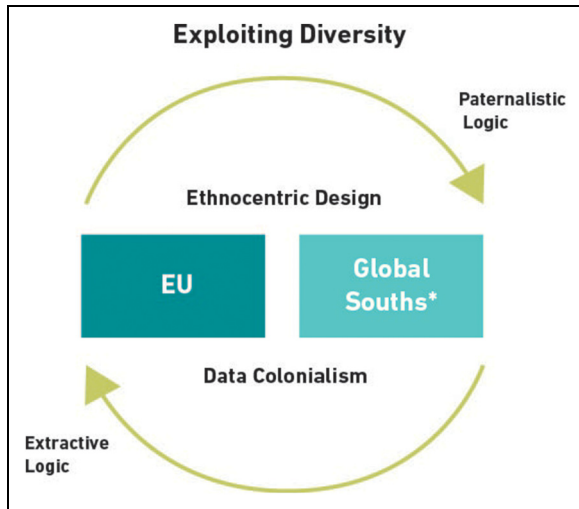
Costanza-Chock’s (2020) call for design justice is path-breaking here, as is Fry’s concept of “defuturing.” Both expose how patronizing, top-down design processes take the diversity of futures away (Fry, 2020: 238). Fry suggests an alternative route, grounded in a relational way of designing for, by and with the Global South (Fry, 2017: 26). He describes methods of designing that break with epistemological imperialism by selectively redirecting its value: designing that integrates activism and local needs and brings human–nonhuman materialities together. According to Fry, the problems and prospects of the Global South urgently require departing from the scalability mindset described above. Instead, pluriversality should be the benchmark. By this, we mean striving not towards a “one-world-world” (Escobar, 2017, 2018), but rather, to say it in the words of the Zapatista: *a world where many worlds fit*. Being conceptualized as such, pluriversality goes beyond the idea of tolerance of different approaches, which starts from one specific position, for example the European one. Instead, pluriversality is decentral (Mignolo, 2012).

Drawing on the struggles of autonomy movements in Latin America, Escobar’s (2017) autonomous-led design framework proposes a practice for alternative world-making in line with pluriversality. Such an approach is relevant for imagining new collective futures both in the Global South and North. At the centre is a renewed notion of communality, relationality and territory (Escobar, 2017, 2018). According to Escobar, the primary goal is to realize the communal in ways that create new essentials for the community’s persistent self-creation and relationship with local territory (Escobar, 2017). From this understanding, autonomy-led design is a counter-response to top-down ethnocentric design and data extraction and an invitation to a design practice attuned to the relational dimensions of life and more ethical ways of collaborating (Milan and Trere, 2019). From this perspective, the pluriversal design project is incomplete and an ongoing process of “unlearning and re-learning to see the world” (Schultz et al., 2018: 97).

While these ideas and concepts sound promising and, if implemented, can certainly contribute to true social innovation, we encounter limitations when putting them into practice. Institutional and political discourses on diversity and inclusion are often disconnected from design practice, quickly leading to inconsistencies when implementing intercontinental projects. Although participatory and co-design approaches are lauded, we find that practice often falls short, not because the approaches themselves are flawed, but because the conditions under which they are carried out are more than sub-optimal.

### *Graphical model: Circle of extraction and paternalism*

The following model (Figure 1) summarizes the previous three subsections. It visualizes specifically those circular



**Figure 1.** The circle of extraction and paternalism.

dynamics between EU-funded research and partners from the Global South that we aim to avoid contributing to. It forms the background against which we evaluate our experiences.

### How are these considerations reflected in concrete practice?

Within the WeNet-project, the Danish partner coordinated the pilots and designed the application that ran on top of the WeNet-platform. The pilots were planned in three iterations over four years and across different pilot sites: Aalborg University in Denmark (AAU), the London School of Economics in the United Kingdom (LSE), the University of Trento in Italy, the National University of Mongolia, and the Universidad Católica in Paraguay (UC), Aamrita University in India, Jilin University in China, and the Instituto Potosino in Mexico.

Each pilot iteration informed the subsequent one while ideally providing enough data to the technical partners to develop the needed algorithms to learn the different social practices in the students' communities. The pilots consisted of three data collection phases through different means: a survey, a sensing application Zeni et al. (2014) and a chat-application developed by the project. Table 1 summarizes the number of participants versus the total number of invitations sent in each pilot's site's different data collection phases.

Because the project was funded under a research and innovation line, it was clear from the outset that the success of the pilots and, ultimately, the project would depend on the amount of data collected, that is, the number of students participating and the duration of their engagement, and that the project would be evaluated on whether it could be "scaled up." However, this both

**Table 1.** Participants across pilots' locations.

Country	Survey	I-Log	Application
Paraguay	1546/13,403	28/193	22/193
Mexico	84	84	
India	5061/21,653	151/350	
Denmark	491/2843	24/46	34/46

economic and technical requirements conflicted with the ethico-political goal of accounting for and collaborating with a diversity of cultural contexts and applications in each country. This created the risk that the desired large data sets would not materialize, and thus robust and generalizable algorithms could not be trained because the individual data sets would be too diverse.

Apart from these ethico-political tensions, the request of scaling up also appeared as a mismatched challenge for a research project in which the developed application served as a technical probe and was not actually meant to take the shape of a product ready for marketing and sale. This, however, turned out to appear as one of the key criteria of project evaluation.

Two other constraints limited the design space: first, the tight timeframe of such a complex project (which was also influenced by the pandemic affecting the different pilot sites differently), and second, the fact that the budget for the development activities was strongly focused on one European partner, which meant that there were not enough resources to develop different applications for the different contexts and needs.

Given these constraints, the Danish partner, at the prompting of the ethics partner and in agreement with the rest of the consortium, decided to develop an application that could be used in the different pilots to address a pressing need in a global pandemic: a chat application through which students could connect in moments of struggle and isolation. With this application, students could help and support each other while exploring the diversity of their community and benefiting from the different skills, competencies, and values that the algorithms could leverage.

All pilot sites found this scenario and the associated app relevant. Despite limited or no funding, all local pilot partners contributed to the co-design of the app, the profiling of the students, and the development of the data collection tools that would be used to evaluate the students' experiences subsequently. Nevertheless, the development process was friction-ridden and complex, largely because of the need to balance the needs of technical partners, who had to meet their predetermined research agenda under pressure to scale, and those who did the grassroots work at the pilot sites and were primarily concerned with the needs of their local communities.

After analyzing the lessons learned from the first two iterations of the pilot, all stakeholders recognized the need to design the third iteration in a way that allowed the

non-European pilots to work more autonomously on their scenario and application with the support of the consortium. This was possible thanks to the commitment of some pilots who actively contributed their opinions to the consortium meetings and then independently selected a relevant scenario to work on in pilot three and adapted the application accordingly.

As exemplary cases, in the following, the supervisors of the pilots from Denmark, Paraguay, Mexico, and India provide their perspective on the data collection process, positioning themselves in relation to power dynamics and reflecting on challenges, progress, achievements, and (missed) opportunities.

### *The Denmark pilot*

As explained in the previous paragraph, the AAU team, compared to the Associated Partners, has a privileged position in terms of decision-making and implementation powers in the Consortium. Moreover, as a European partner, the AAU has a full budget for hiring staff and covering all expenses related to the research activities.

Despite its privileged position, the AAU faced challenges mainly due to the technically dominated nature of the collaboration and the funding instruments.

The student induction process was designed to replicate the consolidated process at another European pilot site. At the outset, participants were asked to complete an extensive survey and then use the I-Log application to collect data from various sensors, with limited customization depending on the pilot context. In Denmark, however, the number of collaborators was lower than expected. A cultural difference may have played a role: Danish students are not used to working on long surveys. They also feel free not to collaborate on research activities proposed by their professors. Therefore, local adaptation seems necessary not only on the content level but also in the design of the recruitment process.

In general, the Danish team questioned the planned top-down procedure both in general and specifically due to the diversity orientation of the project. Given the top-down approach, missed opportunities were also pointed out by the Danish students in the field, who are very sensitive to the problems of data capitalist models. They would have appreciated the data being shared within the community and working with it themselves. Although the value of the data was immediately apparent to all, due to the now widespread skepticism, more work should have been done to be transparent about how the data gathered by the I-Log sensors were transformed into meaningful insights. An interesting research question for this pilot site is how to build, organize, and sustain a community of users interested in maintaining their data sets as a commons and using them themselves. However, we also recognize that a single project that must fulfil their promises to the EU Commission is unlikely to realize the potential of

more participatory approaches given the limitations of the funding instrument.

In terms of the cultural challenges posed by the Eurocentric top-down design of the pilot, we also note a high sensitivity to gender issues in the liberal Danish context: the binary distinction in the survey was perceived as outdated and offensive.

Also, the tone of the application, used as a research tool to collect diversity data, was perceived as overconfident and even insulting when trying to persuade students to cooperate. It should be mentioned that in the Danish pilot, it was decided to stick to English, as only one language could be used per pilot due to limited time and budget resources to develop the application. In addition, it was considered important to connect with the many international students at the university. Nevertheless, some Danish students felt discouraged from discussing personal issues or answering intimate questions in a foreign language. This problem was made particularly relevant by the COVID-19 pandemic and the isolation that accompanied it. In this context, it should be noted that the two supervisors involved in the Copenhagen pilot are Italian and Hungarian, used to speaking English in their work environment.

### *The Paraguay pilot*

The Paraguay pilot was carried out following the generalized procedure described above. Although in this paper we also include India and Mexico under the term “Global South,” Paraguay is the only pilot site geographically located in the Southern Hemisphere, where the academic calendar is opposite to that of the North. This came with a host of challenges in terms of timing, rhythm, etc.

Despite the organizational challenge, the consortium team was able to accommodate the timing and rhythm of the overall project. Furthermore, everyone involved in the project design agreed that the pilot would require special knowledge of the local culture and slang, especially because in Paraguay, most people speak a mixture of Guaraní and Spanish, which we call Jopara. The need for understanding this specific slang and the culture embedded within it makes the participation of local experts elementary. During that participatory design process, the applicability of Eurocentric approaches was fundamentally challenged.

Despite many efforts in terms of alignment, it was difficult to engage students in the pilots, especially in the I-Log and chat application activities. In Paraguay, there is no established culture of collaboration in these types of studies. Given this situation and the fact that most students work long hours apart from their studies, they would have needed a strong thematic or economic incentive to make the effort to participate. Despite these obstacles, those students who nevertheless participated were remarkably active and enthusiastic about the general project idea.



Another interesting feature is the use of technology. Instead of the original purpose (i.e. to ask for and offer help in solving specific tasks), UC students used the WeNet application primarily for casual and non-targeted forms of interaction. This could be due to the physical distancing requirement during the COVID-19 pandemic.

Another particularly relevant point in terms of the limits of Eurocentrism is related to the preparation of the questionnaire, which was initially supposed to be unified to generate comparable data sets but had to be readopted many times to speak to diverse living realities. Given the different cultural contexts and the variances from European higher education contexts, for a transcontinental project like WeNet, it turns out to be hardly feasible to use a uniform survey and collect the same data indicators, even if this would feed scaling aspirations and desires for extracting easily reusable data sets.

For the European partners, for example, one of the relevant aspects to be covered is student mobility, housing, and living conditions. In Paraguay, over the past 25 years, many universities have been established. Currently, 58 universities are accessible to Paraguay's 7 million inhabitants. Many have multiple campuses in different cities and cover the most populous regions of the country. Despite significant differences in the quality of study, it is the norm for students not to move to another city to study but to stay with their families and look for opportunities nearby. This reduces opportunities for shared housing and other types of social interaction among students that seem common in European higher education contexts. In addition, full-time students are a small minority. The vast majority are part-time students whose primary occupation is paid work. Under these conditions, university life is essentially limited to attending lectures and, when necessary, laboratories. Students are used to studying at home, and there are almost no rooms for studying at the university. Social relationships, friendships, and the feeling of belonging to a community with fellow students are also lower due to the double burden of work and care responsibilities. This is a striking difference from the pilot projects in Europe as, for example, at LSE in London, participants reported a strong sense of belonging to a community, which also reduced their concerns about privacy.

Housing and living conditions are also different. Unlike in the European cities where the first two pilot iterations were designed, citizens in Paraguay try to avoid public transportation at all costs. This is mainly due to safety issues and the overall poor reliability of public services in Paraguay. As a result, students are accustomed to seeking safe parking on university campuses or in the neighborhood. In response, we worked with our European partners, who were sensitive to this problem, to develop a variant of the uniform questionnaire tailored to the Paraguayan context. The third pilot phase also tested several possible scenarios for Paraguay. For example, a scenario aimed at bringing users together for community-based transportation

solutions such as carpooling would have been very valuable to the local community but could not be implemented in the short term. Instead, a much easier adoptable scenario was implemented that focused on using the app for students to tutor each other in their studies.

Another issue of relevance is the state of data collection, with large differences in privacy protection between Europe and Paraguay. In Paraguay, there is currently no data protection law. In this way, we were able to comply with the binding regulation for EU-funded projects. In the meantime, Paraguay is working on its own data protection law, and once it is passed, it could pose new challenges to the applicability of the GDPR standard.

Finally, it is worth noting that despite obstacles, a way was found to allocate some financial resources from the European project to conduct not only the pilot study but, to a limited extent, also research. This way, buying into the extractive dynamic attributed to data colonialism that has been found to be prevalent in private sector tech innovation endeavors was indeed avoided. This was thanks to the proactive engagement of the European partners and their willingness to share their university funds in order for the Paraguayan pilot to go beyond simply carrying out western ideas. As a result, comparative data analysis has been conducted, leveraging the collected data for tackling region-specific and broader research questions, thereby overcoming blunt extractivism.

### *The India pilot*

The India pilot was conducted at Amrita University with its six campuses. It involved a survey followed by I-log data collection. Regardless of students' different native languages, most university courses in India are taught in English. This made the survey easy to implement, at least language-wise. While there was no direct funding from the EU, as there is no bilateral agreement, part of our expenditure on pilot implementation mimicking the centralized design was made possible with direct funding from Trento University, where the consortium lead was based. The Eurocentric approach had advantages and disadvantages. Among the advantages was the opportunity to compare with European data to see how the data collection methodology can help transcend European boundaries and reflect diversity among participants and between countries, for example, in examining student behavior and mood when using the prototypes. However, because this was the only Indian pilot project with even less funding than Paraguay or Mexico and was not adapted in any way, comparative data analysis at the local level to address local research needs was not possible.

Given this situation, the Indian pilot project, like the Paraguayan one, encountered several practical challenges. In our case, the diversity survey was an initial component sent to students at a time when sensitivity to data reuse

was heightened due to the “remote-only” phase caused by lockdown and computer use outside the classroom. More than three-quarters of the students did not want to cooperate. Among the incomplete questionnaires were those in which students did not indicate which faculty or department they were studying in, which academic year, where they were living during the semester, and in what type of housing. This can be taken as a first indication of skepticism about privacy. Although the goals of the survey were communicated, some students emailed to ask if there was a pretext, such as control or surveillance as one of our goals. Since this was, of course, not the case, we received positive responses about their willingness to cooperate. This cautious, even suspicious, attitude can be taken as another indication of heightened privacy concerns.

When we asked students who had completed the questionnaire to cooperate with I-log data collection, we received several messages from students complaining about that data collection disrupting their daily routines because of the frequently interrupting questions on the app. Only a small percentage of respondents cooperated. The request was made at a time when most students had to sit in front of their computers for six or more hours straight, and this computer fatigue was one of the reasons for refusal. Some students told us that the questions about activities of daily living had become an “extra workload.” Other students wrote that they had problems with server synchronization of the collected data.

In terms of content, similar to Paraguay, key aspects of the descriptions in the standardized survey, including housing and living situations that reflect European standards, were experienced as confusing because most Indian students go through only one (living with parents or a guardian) or two options (living in a dormitory or in a private apartment in or near the university) in their lives.

Several conclusions can be drawn from this experience, which provides additional context for the case study given the comparatively low number of participants. We hypothesize that the reason why only a small proportion of participants responded fully is that, unlike the funded European partners, the Indian pilot was unable to offer incentives or provide resources for co-design, which would have resulted in a more customized and thus locally meaningful version of the pilot that would also have motivated students to participate. Overall, there was a lack of sensitivity for how to best manage the inclusion of ICT-use into the lives of local students, their living circumstances, and the existing digital infrastructure. Even if data collection is governed by GDPR-compliant methods, local sensitivities must be taken into account, and this should be reflected in the recruitment process.

### *The Mexico pilot*

The Mexican pilot took place in San Luis (population 825,000), the capital of the Mexican state of San Luis

Potosí, located in central Mexico. There are several universities in the city with a large and diverse student body. In this case, bilateral agreements with the EU were in place and the budget was also doubled by the same amount from local research institutions, who, on top of the operational money from the EU, gave additional money for design and research. On that basis, a multidisciplinary team of researchers from IPICYT and Idiap designed a tailored pilot project divided into three main phases. This made the case decidedly different from the other Global South pilots. First, the research team conducted a customized and structured campaign to recruit students through social media and presentation workshops. This was based at Universidad Autonoma de San Luis Potosi and Universidad Tangamanga, two leading universities in the city. The campaign’s focus was to address the obesity epidemic in Mexico, a health problem that strongly concerns and severely affects students throughout the country.

The pilot thus addressed an issue that many Mexican students intrinsically care about (Meegahapola et al., 2021). Cases that have little impact on local communities, in turn, are difficult to promote among Mexican students, even when there are economic incentives. The main motivation for the volunteers who participated was the opportunity to provide mobile data that would allow documentation of the influences that the environment, opportunities, or life circumstances have on obesity. The appropriate characterization of this data will inform the development of community-based health interventions working with local student communities. Because of the clear relevance of the study, the recruitment campaign was successfully carried out by supporting volunteers, including students, teachers, institutional authorities, and local media.

The second phase of the pilot project consisted of an experiment run from September to December 2019 to test and validate the methodology developed to collect real-world data. The mobile data collected during the pilot provided information about the context surrounding food consumption, captured by the participating students. However, some limitations of the overall approach also became apparent during the development of the pilot project. The application used for data collection was not compatible with many of the phones used. Most students in Mexico have outdated cell phones and cannot afford phones with the latest operating system, which was assumed as the norm by the EU-based technical partners. In addition, the brands and features of phones sold in Mexico are often different from those in Europe. Therefore, many students had to borrow a compatible cell phone or refrain from participating. Software differences led to difficulty for students to receive important notifications needed to provide data at the expected time and place. This situation led to some participants’ frustration and hindered the pilot’s development. This experience highlights the importance of involving the local research team and communities from the

beginning. Adapting the methodology and technology to the local context is essential to address the different circumstances of the target groups.

Finally, in the third phase, a pilot like the one implemented in the other sites was run in June 2021. The approach of the pilot changed significantly due to the COVID-19 pandemic. Schools in Mexico were closed in March 2020 and reopened in September 2021. For this reason, all interactions with students took place online, which had a detrimental effect on motivation.

The Mexican pilot was the only one among the pilots in the Global South that used a design tailored to the needs of local people, which attracted many participants in the first round. However, this co-design process should not only consider a comprehensive understanding of the local context but also consider technological limitations and the specific interests of the target population, which can vary widely even within communities in the same city. In our experience, pilot studies that take the needs and interests of collaborating volunteer communities seriously may not achieve the largest data sets, but they can achieve richer and more meaningful ones. Addressing community challenges requires creating synergies among all participants to be effective. When co-designed with local stakeholders' help, pilot projects can foster such synergies. This is important because an experiment like the one discussed here, which focuses on diversity, should benefit all partners equally to live up to its objective of making a difference in private sector activities. Ultimately, this is the only safe way to avoid neocolonial dynamics, where one-size-fits-all data extraction is paramount to creating big and monetizable data sets, and the benefits are unequally distributed. From the beginning, the Mexican pilot took this into account in the expanse of voluntary additional engagement of local researchers. In the process, all dimensions must be considered, from attention and time to tailored local expertise, as well as addressing technical incompatibilities in the field. In conclusion, community-based, locally co-designed pilots are more likely to contribute to efforts that move away from neocolonial practices, but they require more and different funding opportunities.

## Comparative analysis

The following is a comparative analysis of the different pilot sites. This analysis is situated on two interrelated levels. One relates to the broader political-economic context ("Related works and conceptual background" section) and the other to the more specific project context ("How are these considerations reflected in concrete practice?" section). In the following, we focus on the interplay between these two.

A first striking feature that comes to light through the comparative perspective is the apparent need to start the co-design process already in the recruitment phase, as all

pilots report difficulties with the standardized process, such as a lack of sensitivity in dealing with privacy concerns. The exception here is the Mexican pilot, which began customization at this stage. Another striking feature in this regard is the adaptation to different technical standards and infrastructural conditions. Here, the Danish pilots had the least difficulties, working with a fixed norm according to EU standards, while all others reported technical difficulties with this standard. In these cases, the need and value of comprehensive co-design becomes very clear.

Things become more complicated when it comes to the content of the data collection. While Mexico makes a strong case for tailored design throughout, which is useful for solving local problems but makes comparability difficult, Denmark, though critical of it, expresses an understanding of the technical needs to work with standardized data sets that can be more easily merged to train robust algorithms, and to meet the scaling requirements that some of the reviewers demand. This is especially true when participant numbers are low, which was the case in Denmark, but not in Mexico. This, in turn, is a good argument for the Mexican strategy of throughout tailoring, because it is precisely this strategy that has led to high participant numbers. At another level, the value of comparability was highlighted by Paraguay and India, as this allowed them to conduct relevant research with the data despite limited resources, which was seen as a way to avoid repeating neocolonial practices of unidirectional data extraction. In all cases, however, it was also clear that without some level of co-design, it is impossible to conduct meaningful pilot projects. At a minimum, this means that a standardized survey must incorporate different living realities to not deter participants. To achieve meaningful results and avoid replicating private-sector, data-capitalist, or even colonialist practices, in some cases where time and funding allowed, a customized design was used in the third pilot iteration. This can be seen as the result of a learning process that took place within the project. Reflecting on this, the contribution of the pilot sites in the first two iterations can be described as "translators" of a Eurocentric design into diverse local contexts.

In all cases, with or without comprehensive co-design, research was conducted, and joint publication projects were initiated (including this one). This was despite the fact that there was no official funding for such activities. Notwithstanding, all project partners agreed that it was elementary to involve local experts not only in data collection, but also in research, in order to share the benefits and not fall into an extractive logic. However, in the absence of formal funding for such activities, the important role of Global South partners in co-design, translation, diversification, analysis, and co-authorship was largely carried out "pro bono" or thanks to voluntary grants from EU partners, with the latter receiving all funding for research and design.

This meant not only that Global South partners were financially and structurally dependent on EU partners, but also that they had to do most of the co-creation work in addition to their regular jobs to earn a living.

The results are, therefore, ambivalent: while working on projects like the one discussed here was perceived by all partners as a great opportunity and enrichment, the structurally embedded inequalities make it a delicate challenge to build working relationships on an equal footing and to counteract existing power asymmetries instead of reinforcing them. In summary, the inclusion of local expertise is essential to the success and quality of a project that seeks to look beyond a Eurocentric horizon. However, this is carried out on the shoulders of people who work overtime and receive little compensation for it. Under these conditions, our co-design and research efforts have done well to avoid neocolonial logics of unilateral data extraction, but were limited in terms of overcoming paternalistic relations of inequality.

### **Confronting the two sides of the neocolonial circle: Extractivism and paternalism**

Reflecting on our collective experience in the context of an EU-funded but transcontinental research and innovation project, we can draw some broader conclusions that we explicitly situate within historically rooted but still effective structures of inequality:

Avoiding data extractivism is a major challenge in a project constellation where data collection plays a central role and where most of the funding comes from the EU but is intended to be used only to carry out the research planned and targeted by the EU partners, without distributing design and research expertise equally among the partners. If such a policy is implemented unquestioningly, it puts partners in the Global South in the position, so to speak, of doing the manual labor that EU partners then use as a basis for their intellectual work. When this is coupled with a unidirectional flow of data, where data is the valuable commodity that is taken from the South and enriches the West, we are indeed dealing with neocolonial structures. These neocolonial structures ultimately harm all stakeholders by undermining the innovative potential of learning from diverse communities. To avoid this, two strategies can be employed. First, standardization must be challenged despite its scaling-oriented benefits. Although there are pro and con arguments, the ultimate deciding factor here is the question: *cui bono*? Despite the arguments favoring standardization, we have learned from our deliberations that local adaptation is necessary to make participation more attractive to people living in very different sociocultural contexts. To do this, local teams must be involved not only in the implementation phase of the

research, but also in the early stages of project planning and design. Second, it is elementary that those who collect the data be involved in the research, analysis, and publication. Therefore, in our case, all project partners went to great lengths to ensure that joint research was conducted with the collected data, resulting in joint publications (Schelenz et al., 2021; Meegahapola et al., 2021, 2022; Assi et al., 2023). This ensures that all stakeholders benefit from the data they help produce and that the data are used in ways that are relevant to different local contexts.

Avoiding paternalism is by far the more difficult part because it cuts most deeply into interpersonal relationships, which are shaped by historically established and normalized inequalities, while also requiring the most far-reaching structural changes. Although the voluntary transfer of own research funds from EU partners to partners in the Global South and their proactive participation in design decisions and research make it possible to avoid extractivist dynamics, avoid replicating the data colonialism of private sector companies, and even promote pluriverse perspectives, all these activities are taking place on the grounds of a relationship of dependency between the parties. This is exacerbated by the fact that the ultimate responsibility for the project's success lies with the EU partners, while the associated partners are only allowed to participate unofficially without being given institutional authority. Such a structure provides an excellent template for perpetuating historically entrenched paternalism. Only if the EU funding instruments themselves are willing to rethink their "Europe-first" policies and see Europe as part of a larger, pluriversal whole will it be possible to challenge structures that, even if not intentionally, effectively confirm rather than redress existing relations of inequality.

### **Conclusion**

Publicly funded and transcontinental technology development projects involve much more than just "exported" or "imported magic" (Medina and Costa Marques, 2014). Rather, the co-production and appropriation of technology by Global South actors can shake up Western notions of innovation, thereby being truly innovative in the sense of social innovation. To unleash this potential, we need to create spaces for collaborative and constructive self-critique. This paper is the result of such an exercise. Triggered by the critical whiteness perspective that embedded ethics research brought to the project, we became aware of some inconsistencies in our own work, which, despite its dedication to preserving and leveraging diversity, in some cases replicated the top-down logic that is prevalent in private sector innovation cultures. Therefore, we analyzed our experience to identify useful counter-strategies that are already available and their limitations. In summary, we can claim that when working towards it, we can indeed make significant differences to the extractive

private sector practices that are criticized under the term data colonialism. Nonetheless, we have also found that it has been almost impossible to overcome paternalistic logics in the scope of our abilities. The reasons for this, we believe, are symptomatic of broader problems: first, the methodological credibility and knowledge production of local researchers from the Global South continues to be undermined by deeply entrenched institutional structures; second, even publicly funded researchers work under the pressure of scalability imperatives, which in turn are influenced by data capitalist models. In the words of Sandra Harding, “We need to rethink what the background is and who and what we are foregrounding” (Harding, 2011). In the current scale-fixated innovation culture, local context is in the background, and large-scale data extraction and analysis is in the foreground, but the latter is meaningless unless more emphasis is placed on the former (Kitchin and Lauriault, 2015). At the same time, publicly funded intercontinental research collaborations are becoming more frequent and important, not least as a response to private sector strategies. To ensure that they are equally beneficial to all participants, we need to tweak funding instruments and support infrastructures toward a more equitable distribution of power and design space to enable pluriversal research and innovation.

### Acknowledgements

This work was funded by the EU Horizon 2020 program for research and innovation under grant agreement no. 823783. We would also like to thank Aalborg University and the IDIAP Institute for additional financial support, and the members of the research colloquium at the Global Digital Cultures RPA at the University of Amsterdam for their valuable feedback, as well as the discussants of our panel during the 2022 Association of Internet Researchers Conference in Dublin. Finally, we thank the valuable feedback by the anonymous reviewers who significantly helped us to sharpen our arguments.






### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Paula Helm  <https://orcid.org/0000-0002-2719-9721>  
 Amalia de Götzen  <https://orcid.org/0000-0001-7214-5856>  
 Luca Cernuzzi  <https://orcid.org/0000-0001-7803-1067>  
 Shyam Diwakar  <https://orcid.org/0000-0003-1546-0184>  
 Daniel Gatica-Perez  <https://orcid.org/0000-0001-5488-2182>

### Note

1. In what follows, we use the terms “Global South” and “Global North.” By doing so, we resort to the North–South dualism, as it helps to mark important power asymmetries while at the same time acknowledging the diversity of the areas addressed.

### References

- Anderson W (2002) Introduction: Postcolonial technoscience. *Social Studies of Science* 32(5-6): 643–658. DOI: 10.1177/030631270203200502
- Anwaar A, Qadir J, Rasool Ru, et al. (2016) Big data for development: Applications and techniques. *Big Data Analytics* 1(1): 2. DOI: 10.1186/s41044-016-0002-4
- Armenteras D (2021) Guidelines for healthy global scientific collaborations. *Nature Ecology & Evolution* 5(9): 1193–1194. DOI: 10.1038/s41559-021-01496-y
- Arora P (2016) Bottom of the data pyramid: Big data and the Global South. *International Journal of Communication* 10. <https://ijoc.org/index.php/ijoc/article/view/4297>
- Arora P (2019) General data protection regulation—A global standard? Privacy futures, digital activism, and surveillance cultures in the Global South. *Surveillance & Society* 17(5): 717–725. DOI: 10.24908/ss.v17i5.13307
- Assi K, Meegahapola L, Droz W, et al. (2023) Complex daily activities, country-level diversity, and smartphone sensing: A study in Denmark, Italy, Mongolia, Paraguay, and UK. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394215, pp. 1–23. DOI: 10.1145/3544548.3581190
- Belhabib D (2021) Ocean science and advocacy work better when decolonized. *Nature Ecology & Evolution* 5(6): 709–710. DOI: 10.1038/s41559-021-01477-1
- Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity. ISBN 978-1-5095-2640-6.
- Bruns A (2019) Filter bubble. *Internet Policy Review* 8(4). DOI: 10.14763/2019.4.1426
- Castells M (2009) *The Rise of the Network Society: The Information Age: Economy, Society, and Culture Volume I*. Chichester, West Sussex; Malden, MA: Wiley-Blackwell. ISBN 978-1-4051-9686-4.
- Costanza-Chock S (2020) *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: MIT Press. ISBN 978-0-262-04345-8.
- Couldry N and Mejias UA (2019) *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press. ISBN 978-1-5036-0366-0.
- Dalton CM, Taylor L and Thatcher J (2016) Critical data studies: A dialog on data and space. *Big Data & Society* 3(1). DOI: 10.1177/2053951716648346
- Daniels J, Nkonde M and Mir D (2019) Advancing racial literacy in tech: Why ethics, diversity in hiring & implicit bias trainings aren't enough. [https://datasociety.net/wp-content/uploads/2019/05/Racial\\_Literacy\\_Tech\\_Final\\_0522.pdf](https://datasociety.net/wp-content/uploads/2019/05/Racial_Literacy_Tech_Final_0522.pdf)
- Deldjoo Y, Jannach D, Bellogin A, et al. (2023) Fairness in recommender systems: Research landscape and future directions.

- User Modeling and User-Adapted Interaction*. DOI: 10.1007/s11257-023-09364-z
- Dworkin G (1972) Paternalism. *The Monist* 56(1): 64–84. DOI: 10.5840/monist197256119
- Escobar A (2017) Response: Design for/by [and from] the ‘Global South’. *Design Philosophy Papers* 15(1): 39–49.
- Escobar A (2018) *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. New ecologies for the twenty-first century. Durham: Duke University Press. ISBN 9780822370901.
- European Commission (1999) The EU values. <https://ec.europa.eu/component-library/eu/about/eu-values/>
- European Commission (2019) General Data Protection Regulation (GDPR) – Official Legal Text. <https://gdpr-info.eu/>.
- European Commission (2022) The EU’s global approach to research and innovation. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733550/EPRS\\_BRI\(2022\)733550\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733550/EPRS_BRI(2022)733550_EN.pdf)
- European Parliament (2021) *Regulation (EU) 2021/695 of the European parliament and of the council*, volume 170. <http://data.europa.eu/eli/reg/2021/695/oj/eng>
- Felt U (2022) *Making and Taking Time*. London: Routledge. p. 204–217. DOI:10.4324/9781003129424-17
- Fry T (2017) Design for/by “the Global South”. *Design Philosophy Papers* 15(1): 3–37.
- Fry T (2020) *Defuturing: A New Design Philosophy*. London: Bloomsbury Publishing USA. ISBN 1350089532.
- Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599. DOI: 10.2307/3178066
- Harding (2011) *The Postcolonial Science and Technology Studies Reader*. Durham, N.C.: Duke University Press. ISBN 978-0-8223-4957-0.
- Harding S (1995) “Strong Objectivity” : A response to the new objectivity question. *Synthese* 104(3): 331–349.
- Helm P, Michael L and Schelenz L (2022) Diversity by design: Balancing protection and inclusion in social networks. In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, AIES ’22*. New York, NY, USA: Association for Computing Machinery, pp. 324–334.
- Hilbert M (2016) Big Data for development: A review of promises and challenges. *Development Policy Review* 34(1): 135–174. DOI: 10.1111/dpr.12142. <https://onlinelibrary.wiley.com/doi/abs/10.1111/dpr.12142>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/dpr.12142>
- Jasanoff M (2006) *Edge of Empire: Lives, Culture, and Conquest in the East, 1750–1850*. New York: Vintage. ISBN 978-1-4000-7546-1.
- Kitchin R and Lauriault TP (2015) Small data in the era of Big Data. *GeoJournal* 80(4): 463–475. DOI: 10.1007/s10708-014-9601-7
- Madianou M (2019) Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises. *Social media + society* 5(3): 205630511986314.
- Medina E, Costa Marques Id and Holmes C (2014) *Beyond Imported Magic: Essays on Science, Technology, and Society in Latin America Inside Technology*. Cambridge, MA: MIT Press.
- Meegahapola L, Droz W, Kun P, et al. (2022) Generalization and personalization of mobile sensing-based mood inference models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6(4): 1–32. DOI: 10.1145/3569483
- Meegahapola L, Ruiz-Correa S, Robledo-Valero VdC, et al. (2021) One more bite? Inferring food consumption level of college students using smartphone sensing and self-reports. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 5(1).
- Mignolo W (2012) *Local Histories/global Designs: Coloniality, Subaltern Knowledges, and Border Thinking*. Princeton: Princeton University Press.
- Milan S and Trere E (2019) Big Data from the South(s): Beyond data universalism. *Television & New Media* 20(4): 319–335. DOI: 10.1177/1527476419837739
- Mill JS (1859) *On Liberty*. Independently published. ISBN 978-1-70682-085-7.
- Müller R (2012) Collaborating in life science research groups: The question of authorship. *Higher Education Policy* 25(3): 289–11. DOI: 10.1057/hep.2012.11
- Musa M and Rodin J (2016) Scaling Up Social Innovation (SSIR). [https://ssir.org/articles/entry/scaling\\_up\\_social\\_innovation](https://ssir.org/articles/entry/scaling_up_social_innovation)
- Pariser E (2012) *The Filter Bubble: What the Internet is Hiding From You*. London: Penguin Books. ISBN 9780241954522.
- Pfotenhauer S, Laurent B, Papageorgiou K, et al. (2021) The politics of scaling. *Social Studies of Science*. 03063127211048945. DOI: 10.1177/03063127211048945
- Ricaurte P (2019) Data epistemologies, the coloniality of power, and resistance. *Television & New Media* 20(4): 350–365. DOI: 10.1177/1527476419831640
- Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6(1): 2053951718820549. DOI: 10.1177/2053951718820549
- Schelenz L, Busso M, Bison I, et al. (2021) The Theory, Practice, and Ethical Challenges of Designing a Diversity-Aware Platform for Social Relations. In: *Proceedings of the 4th AAAI/ACM Conference on Artificial Intelligence, Ethics and Society*. DOI:10.1145/3461702.3462595
- Schelenz L and Pawelec M (2022) Information and communication technologies for development (ict4d) critique. *Information Technology for Development* 28(1): 165–188. DOI: 10.1080/02681102.2021.1937473
- Schultz T, Abdulla D, Ansari A, et al. (2018) What is at stake with decolonizing design? A roundtable. *Design and Culture* 10(1): 81–101.
- Staunton C, Barragàn CA, Canali S, et al. (2021) Open science, data sharing and solidarity: Who benefits? *History and Philosophy of the Life Sciences* 43(4): 115. DOI: 10.1007/s40656-021-00468-6
- Taylor L and Broeders D (2015) In the name of development: Power, profit and the datafication of the Global South. *Geoforum; journal of physical, human, and regional geosciences* 64: 229–237. DOI: 10.1016/j.geoforum.2015.07.002
- Thatcher J, O’ Sullivan D and Mahmoudi D (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space* 34(6): 990–1006. DOI: 10.1177/0263775816633195
- Tsing AL (2004) *Friction: An Ethnography of Global Connection*. Princeton: Princeton University Press. ISBN 978-0-691-12065-2.
- Tsing AL (2012) On nonscalability: The living world is not amenable to precision-nested scales. *Common Knowledge* 18(3): 505–524. DOI: 10.1215/0961754X-1630424
- United Nations (2021) *Digital Economy Report 2021: Cross-border data flows and development*. United Nations. ISBN 978-92-1-113022-5.

- Varon J and Peña P (2021) Artificial intelligence and consent: A feminist anti-colonial critique. *Internet Policy Review* 10(4).
- Vos Ad (2020) The problem of colonial science. Conservation projects in the developing world should invest in local scientific talent and infrastructure. *Scientific American*. <https://www.scientificamerican.com/article/the-problem-of-colonial-science/>
- Wetzling T, Sarkesian L and Dietrich C (2021) Solving the transatlantic data dilemma. [https://www.stiftung-nv.de/sites/default/files/snv\\_solving\\_the\\_transatlantic\\_data\\_dilemma.pdf](https://www.stiftung-nv.de/sites/default/files/snv_solving_the_transatlantic_data_dilemma.pdf)
- 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* 48(1): 1–13. DOI: 10.1080/17439884.2021.2010094
- Zeni M, Zaihrayeu I and Giunchiglia F (2014) Multi-device activity logging. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct. New York, NY, USA: Association for Computing Machinery. ISBN 9781450330473, pp. 299–302.