

RESEARCH PAPER

Work HCI Data: What it Is, Who it Belongs to, and the Need for Regulation

Luis Gonçalves  [University of São Paulo | goncalves.luis@usp.br]

Roseli Figaro  [University of São Paulo | roselifgaro@usp.br]

 School of Media and Arts (ECA-USP), University of São Paulo, Av. Prof. Lúcio Martins Rodrigues, 443, Butantã, São Paulo, SP, 05508-020, Brazil

Abstract. Both the LGPD (General Data Protection Law) and the data protection bills currently pending in the National Congress do not prescribe sufficient guidelines on the use of human-computer interaction (HCI) data from work, whether for monitoring working hours or for application toward other purposes, which allows corporate interests and collective damages to remain inadequately characterized. This article seeks to contribute to filling this gap by discussing the production of workplace HCI through computer-supported cooperative work (CSCW) software and the challenges in implementing public policies that protect workers' know-how. To this end, we start from the empirical focus of the work of media (which includes a datafication experiment), in addition to reviewing parts of the academic literature on the subject and the corporate documentation of socio-digital platforms (terms of use, privacy policies, financial statements, patents, etc.), especially Google, Meta and Alphabet. The theoretical deepening regarding what the data is, its quality, and the purpose of its capture helps clarify the economic functioning of the digital data productive chain [Lastres *et al.*, 2025], and reveals the specificities of non-personal data—above all, data from work processes. The results presented in the analyses aim to support public policies to regulate how companies should proceed when appropriating workplace HCI.

Keywords: Datafication, Sensitive Materialities, Work HCI Data, Media

Edited by: Vagner Figueredo de Santana  | **Received:** 26 March 2026 • **Accepted:** 03 August 2026 • **Published:** 26 September 2026

1 Introduction

Datafication is the process of industrial production, storage, and use of digital data from any source, including human-computer interactions. It is a planned, controlled, and organized process with highly diverse purposes; ranging from those aimed at the interests of public management of social goods to, and above all, those dedicated to profit [Cukier and Mayer-Schoenberger, 2013; Sadowski, 2019; Wong *et al.*, 2025].

The importance of knowledge regarding workplace HCI stems from a historical trajectory that can be traced back to the division of labor in the pin factory analyzed by Adam Smith [2023] in his work “The Wealth of Nations”. According to Pasquinelli [2023], this example inspired mathematicians to organize calculation labor into modular, subsequently mechanizable tasks, which opened up the possibility of imagining, as Ada Lovelace stated, machines through which “*by means of some simple notations, be easy to consign the execution of them to a workman*” [Menabrea, 1843, p. 14]. Under capitalism, human labor has always been an activity conceived and envisioned to be apprehended, controlled, and, whenever possible, automated [Leontiev, 1978]. However, from Babbage to Taylor, Ford, and Ohno, in their creations for capturing labor activity turned into techno-scientific organization, nothing has operated quite like the current invasion of the worker's most intimate gestures by the digital data production chain [Lastres *et al.*, 2025]. The capture of productive human interactions [Agre, 1994] has become a fundamental source for these industries to drive current socioeconomic transformations [Figaro, 2024]. In the case of workplace HCI, this refers to clicks and mouse movements, typing, errors and correc-

tions, where they occur, as well as pauses, screen switching, and even pupil behavior [de Santana *et al.*, 2018], among others, which take place within labor processes and become vital for the performance of approaches such as business process management [Dinkowski *et al.*, 2025]. Within these, HCIs are interpreted in the form of digital data, thereby fueling multiple businesses, especially generative artificial intelligence (AI) operations [Alphabet Inc., 2024], which demands new public policies to regulate increasingly platformized computer-supported cooperative work (CSCW) toolsets.

Current legislation in force and the bills pending in the National Congress regarding the protection of people's information have been limited to what is understood as personal data; regulations are restricted to moral issues; and data from work processes remain unprotected by public policies [Giermindl *et al.*, 2022]. Thus, it is disregarded that corporate research and development structures in technology operate based on the datafication of work activities for the storage and application of this knowledge in platforms such as Microsoft Viva and Google Work Insights, or even in tools like those of the Google News Initiative. These examples claim to offer, respectively, products that would improve business productivity and scalability—but to do so, they produce and seize workers' data in an obscure manner.

Such regulatory gaps are confirmed by the guidelines of GranDHCI-BR 2025-2035, specifically within the challenge focused on Human-Data Interaction [Coleti *et al.*, 2024]. The Brazilian research scenario indicates that transparency and auditability of systems that perform massive data appropriation are imperative to ensure human agency and to avoid power asymmetries exacerbated by the opaque datafication of productive activities.

Accordingly, this article discusses how HCI data from work activities are produced through CSCW software, contributing to the proposal of public policies that protect workers' know-how. To this end, we take the empirical example of how the work of media provides data for sociodigital platform corporations, particularly Meta and Alphabet. This discussion is backed by a multidisciplinary literature review, a conceptual discussion applied to the work of media professionals, a study of corporate documents (financial statements, institutional websites, patents, programs, and service offerings), and a technical experiment involving data mirroring of media professionals' activities. All these procedures make it possible to formulate analyses that demonstrate the need for public policies tailored to how companies have organized their business around the production of people's data.

In the first part of the article, we problematize the objectives of the General Data Protection Law (LGPD – Lei Geral de Proteção de Dados, in portuguese) regarding the protection of personal data and labor-related data. We evaluate that the legislation currently in force falls short of the dimensions involved in the production of data from work activities, leaving corporate actions to a *laissez-faire* approach. In light of the study conducted by the Brazilian team of the International Observatory on AI and Work [AIPOWW Global Tracker (Brasil), 2025], we reaffirm the lack of legislation or bills that address the implications of capturing work activities.

In the second part, the article discusses the gaps concerning the ownership of non-personal data—especially labor HCI (Human-Computer Interaction) data—and the accounting maneuvers that enable its industrial use for developing products that will not necessarily benefit the individuals who generated them. In part three, the article delves deeper into the discussion on the quality and value of labor HCI data. To this end, we build upon three key inflections: datafication as an epistemological practice [Agre, 1994], the quality of the sensitive materialities of this data [Figaro, 2024]—including examples from our ongoing patent study—and the hidden and spurious relations of production in its creation, namely heteromation [Ekbja and Nardi, 2017].

In the fourth part, we provide a more detailed description of our case study, focusing particularly on the ongoing experiment involving the datafication of the work of media professionals. We present some methodological aspects of this research-in-progress and the evidence that reinforces the central thesis of this article: the urgent need for public policies tailored to the specificities of this industrial matrix that operates directly with data produced from work activities. The final considerations provide an overview of what was discussed throughout the article, point out future research directions, and raise pressing questions to be addressed by public institutions.

2 Non-personal data: the gap between the non-existence and insufficiency of public policies

According to the LGPD (General Data Protection Law), personal data is defined as "information related to an identified or identifiable natural person" [Brasil, 2018] — such as full name, CPF (taxpayer registry number), address, etc. How-

ever, how can we characterize the digital data originating from these individuals' human-computer interactions? In this case, it potentially refers to any digital records of changes produced by humans in computational systems — prompts, keystrokes, mouse movements, the activation (voluntary or involuntary) of sensors and cameras, etc. [Soegaard and Dam, 2012]. Even though, in certain cases and through indirect means, such data can lead to the probabilistic identification of individuals [Lermen *et al.*, 2026], it generally constitutes anonymized information that tends to be combined with data from countless other people and things. Despite this indirect possibility, and due to their primary characteristics and functions, data such as HCI (Human-Computer Interaction) are frequently referred to as non-personal data [Viljoen, 2021].

As it turns out, non-personal data plays a decisive role in the global digital economy due to its relational properties. Combined in the form of big data, this data provides a statistical inference power that transcends isolated individual information [Kitchin, 2014; Arrieta-Ibarra *et al.*, 2018; Posner and Weyl, 2018]. This "data externality" [Acemoglu *et al.*, 2022] means that seemingly harmless information (such as telemetry logs), when turning into inputs for machine learning, acquires economic value because it creates population categories and relationships that can then retroactively act in a "personalized" manner upon each individual [Microsoft, 2025a; 2025b]. Therefore, the distinction between external and internal is misleading, given that the datafication of individuals is pre-ideally integrated into industrial and commercial chains that are often distinct from its origin [Google, 2024; Viljoen, 2021]. As a result, "our data may be used for purposes that we, as moral beings, would not approve if we were aware." [Wong *et al.*, 2025].

On one hand, according to the LGPD, the "natural person to whom the personal data refers" has "informational self-determination" [Brasil, 2018] over it in the form of data ownership. On the other hand, despite being "inherently relational and collective" [Wong *et al.*, 2025], non-personal data finds itself legally in a gray area, both under the LGPD and in the legislation of most other countries [Bednarz *et al.*, 2025]. This situation allows a significant range of corporate interests and collective harms to remain without proper representation in regulations and software terms of use [Viljoen, 2021; Acemoglu *et al.*, 2022]. For instance, the unregulated production and consumption of this type of data can result in, among other consequences, restrictions on rights or credit, as well as discrimination based on race, gender, etc. [O'Neil, 2016], the lack of recognition of originality and compensation for such intangible assets [Posner and Weyl, 2018], and the abusive control over workers [Giermindl *et al.*, 2022].

The extensive research by the International Observatory on AI and Work [AIPOWW Global Tracker (Brasil), 2025] at the University of Essex provides several examples of Brazilian Bill of Law (PL) initiatives regarding the regulation of business practices of socio-digital platform companies. Notably, Bill 2338/2023 proposes a legal framework for AI focusing on rights, labor, and high-risk systems, among other issues. Bill 234/2023 (PLP) aims to establish a "General Data Empowerment Law" and creates a "Brazilian Data Monetization Ecosystem", adjusting existing regulations on data protection, e-commerce, and tax and civil structures. The safe, ethical,

responsible, and human rights-based use of AI is addressed by Bill 1797/2024 and the aforementioned 2338/2023. Bill 4869/2023 seeks to curb the dismissal or reduction of working hours for voice professionals and Brazilian Sign Language (Libras) interpreters. Voluntary contributions to the Worker's Support Fund (FAT) and the granting of the "Employment-Friendly Company" seal to corporations using AI are the subjects of Bill 3423/2023. Striking a harder stance on data leaks is the focus of Bill 1126/2024, reinforcing the argument for moral and material damages resulting from unauthorized use. Bill 4675/2025 proposes the Digital Markets Superintendency within the scope of Cade (Administrative Council for Economic Defense) for big tech supervision, which includes transparency in terms of use and ranking criteria, interoperability, and data portability. Complementarily, the debate on the Digital Real (Drex) and Digital Wallets introduces "tokenization" as a means for citizens to directly monetize their financial data [Banco Central do Brasil, nd].

However, none of these bills specifically recognize and address rights and the industrial use of HCI data. Furthermore, the legislative processes experience push-and-pull dynamics depending on the level of public pressure for transparency and the defense of citizens' rights and national interest. Yet, big tech pressure on the National Congress, aimed at maintaining their respective economic interests, is highly significant. Examples include the events that occurred during the legislative process and near-approval of Bill 2630/2020—the Brazilian Law on Freedom, Responsibility, and Transparency on the Internet, better known as the anti-fake news bill—which faced fierce attacks from platform companies like Google and Meta in early 2023 [Dias, 2023].

On the other hand, the executive branch has concentrated efforts on developing plans and policies, such as the 2021 Brazilian AI Strategy and the 2024 Brazilian Artificial Intelligence Plan. These address the funding of some key sectors and discuss the issue of sovereignty, but fail to reach data regarding labor activities. Other Brazilian national plans, such as Redata—which aims to use public funds to finance the private (international) data center sector—and the Government Sovereign Cloud Plan, cannot detach themselves from the concentrated economic power of socio-digital platform companies.

Thus, there is a noticeable lack of political understanding of what the capture of labor activity for data production actually means. This discussion connects to the Ethics and Responsibility challenge of GranDHCI-BR. The Brazilian HCI scientific community advocates that systems design must be centered on social justice, which implies recognizing that value production from HCI data cannot occur at the expense of worker over-exploitation and the invisibilization of the appropriation of their know-how. Therefore, the core focus should be on legislation that regulates these new relations of production and a profound overhaul of the terms of use and privacy policies of CSCW software. This is the central and crucial element of contemporary machinery within the current configuration of platform capitalism [Srnicek, 2017; Pasquinelli, 2023] that ought to be regulated.

Our contribution to this discussion and to public policies is a study situated within the field of HCI, specifically in the area conventionally known as computer-supported cooper-

ative work (CSCW) [Grudin, 2002; Kuutti, 1991]. In this case, we investigate which HCI data from work activities are produced through CSCW software, how this occurs, for what purposes, and what the implications of this datafication are for these workers and their professions—specifically in the case of media professionals. We refer to clicks, typing and its content, pauses, errors, screen switching, heuristics, etc., generated by work-related HCI, which contain rich information about the labor process, its products, and its producers [Giermindl, 2022; Microsoft, 2025a, 2025c; Shah et al., 2025]. Thoroughly pre-designed by the companies that own this software, these labor gestures (*gestes du travail*) [Schwartz, 2011, 2019] and the inferences made about them are then formalized into logs and other requests (in addition to cookies, trackers, and scripts, among others), and immediately transferred to distinct private data centers [Google, 2023].

3 Who owns non-personal data?

Datafication is often accompanied by terms such as "collection," "mining," or "extraction." Historically and technically, these expressions belong to a computing jargon [Piatetsky-Shapiro, 2000] that denotes the existence of data in more or less accessible repositories (the internet, etc.) in more or less legal ways [Comitê Gestor da Internet no Brasil (CGI.br), 2023] so that they can be accessed and used. The reality, however, is that ontologically, this data does not pre-exist in the world (like oil or minerals), nor does it grow on trees to be collected, mined, or extracted [Sadowski, 2019]. Before being appropriated by Big Tech, such data must be industrially produced [Amorim *et al.*, 2022; Russell and Norvig, 2021].

The primary form of data production is HCI—astronomers datafy galaxies, quantum physicists datafy atoms, public health experts datafy populations, IT workers datafy, automate datafication, and analyze data (including sensor data), journalists label audiences [Caliari et al., 2026], labelers label data, and even users label images when they like and comment on social media [Trant, 2009]. Therefore, the widespread use of extractivist expressions to refer to the objectification of data elides its social origin, facilitating problematic forms of production, appropriation, and use, while rendering its costs and implications invisible and externalized—that is, transferring them to others [Zuboff, 2021].

With the same naturalness with which data "sprouts" in corporate datacenters, it can come to be treated as capital through monopsonies—that is, legal and socio-technical mechanisms, along with market constraints, that impose a single data "customer," granting this single agent high market power to dictate prices and conditions [Perry, 1978; Arrieta-Ibarra *et al.*, 2018]. To understand these monopsonies, we highlight three examples. First, in the case of personal data, it falls under the responsibility of operators and controllers who can use it—including for other purposes [Brasil, 2018; Google, 2024]—based on the consent we give them in contracts we sign without reading [Costa *et al.*, 2025].

In the case of user creations (texts, images, etc.), as Google's terms of service explain (the wording of which is similar across platform companies), its services "allow you to generate original content," although the company states that

it "will not claim ownership of that content" [Google, 2024]. However, to "operate and improve the services" of Google (which includes creating "new features and functionalities" for third parties), the company requires the user to grant it a "worldwide," "exclusive," and "royalty-free" (free of charge) license to "host, reproduce, distribute, communicate, and use your content" and even "modify or create derivative works based on your content" [Google, 2024]. Note that in both cases, Google does not need to own the data it will use as input (capital) for developing products that will not necessarily benefit the users who originated them.

In turn, HCI data (such as workplace data) undergoes a mysterious form of appropriation. All HCIs produce specific data which, for example in Google's case, besides fulfilling their primary functions (identifying software bugs, etc.), are stored in Google Cloud Storage (GCS) to be immediately converted into tensors (multidimensional arrays). These, in turn, are processed to serve as inputs for products that can be commercialized with third parties [Google, 2024]. According to Alphabet's Annual Report (2024), HCI telemetry and tensor production must be carried out practically within the same industrial plant to achieve zero latency and to properly leverage Tensor Processing Units (TPUs). The report further notes that when we use Gemini in workplace software to suggest text, formatting, categories, and structures, the AI system prompts the user to input and organize their data in a consistent and predictable manner. In this preference extraction work [Silva, 2009]—also called feature extraction or feature engineering [Harper and Randall, 2024]—while the worker interacts, they are simultaneously standardizing the data input, which will facilitate the training of AI models [Google, 2023].

In Microsoft's case, similar processes transform workplace telemetry data into productivity and engagement metrics that feed labor management software such as Microsoft Viva [Hess *et al.*, 2023]. In these and other ways, the goal is not only to "strip all mystery from the productive gesture and transfer all work knowledge to (...) the 'inoperative knowledge' of machinery [Schwartz, 2011, p. 26], but also to offer labor management itself as a service [Gerber, 2019].

4 Heteromation and the capture of the sensitive materialities of labor

Advancing the understanding of the quality and value of labor HCI (Human-Computer Interaction) data, as well as understanding the forms and implications of its appropriation by platforms and the consequent need for its regulation, requires three inflections. First, according to AI scientist Philip Agre (1994), within the context of HCI, it is useful to qualify the concept of capture and separate it from the ideas of surveillance or tracking. The idea of surveillance relates to the loss of privacy based on visual metaphors of discrete and centralized observation by the State, corporations, etc. To a large extent, protecting privacy and preventing the surveillance of individuals in liberal societies is what guides current data legislation and the very limitation of its scope to personal data [Viljoen, 2021], such as the LGPD in the Brazilian case.

On the other hand, Agre [1994] refers to "capture models" as dispossessive epistemological practices mediated by

computer systems that seek to produce and appropriate knowledge about human activities, recursively acting back upon them in five phases. The analysis phase (1) studies "an existing form of activity and identifies its fundamental units in terms of some ontology (e.g., entities, relations, functions, primitive actions, and so forth)" [Agre, 1994, p. 746]. In the articulation phase (2), such units are reconnected as a grammar in order to formalize the rules and data structure that, in subsequent cycles, will represent the activity to be captured and modeled. In the imposition phase (3), users must be induced or forced to organize their actions according to the established grammar—which involves network effects, affordances, choice architectures, etc. [Gonçalves, 2023]. This causes the original activity to be altered so as to become legible by computational models in the instrumentation phase (4), where sensors, interfaces, etc., operate in parallel with the user's original activity, datafying every change of state. Finally, in the elaboration phase (5), the data is analyzed and used by the datafying agents, enabling them to develop a representation scheme "to fully, accurately, or 'cleanly' express particular semantic notions or distinctions, without reference to the actual taking-in of data" regarding human activities [Agre, 1994, p. 744].

The second inflection refers to the nature of what is represented in HCI data. Patents and corporate documentation confirm Agre's capture model, but we can find even deeper historical connections in the datafication of labor. While in the past, the so-called "scientific management of labor" [Taylor, 1911] studied and manipulated labor gestures in experimental environments, the datafication of labor allows such knowledge to be "produced through social behavior, acquiring a fine granularity that was absent from the aggregate statistical data of yesteryear." [Ekbj and Nardi, 2017, p. 135, emphasis in original].

For example, a joint analysis of patents US9672232B1 [Auerbach *et al.*, 2017], US9608892B2 [Ding and Zivkovic, 2017], US9952882B2 [Kuscher *et al.*, 2018], US8667013B1 [Lawrence and Wang, 2014], and EP3803737B1 [Milvaney *et al.*, 2024] reveals how the logging of clicks could produce an indexing of the relevance of each event. This allows calculating the measure of association between documents, actions, and content elements, tagging the entire workflow and the contextual strength of the implicit connections in collective labor. Patents CN108605053A [Zhang *et al.*, 2018], US10387512B2 [Battle *et al.*, 2019], and US12160481B2 [Wu *et al.*, 2024] demonstrate how event logs could be used for workflow mining, reconstructing activity graphs and interaction profiles, as well as identifying logical models and productivity deviations in real time. All of this would allow this knowledge and the labor process to be visualized and managed through platforms such as Microsoft Viva [Hess *et al.*, 2023].

Thus, in a way, it is now the workers themselves who, involuntarily and unconsciously, reveal the work secrets "hidden in the body" [Schwartz, 2014, p. 265], which may include motivations, volitions, heuristics, and individual and collective know-how itself. Work gestures, whether voluntary (mouse movement, etc.) or involuntary (pupil behavior, etc.) [de Santana *et al.*, 2018], have acquired a new importance for companies, workers, and for the field of HCI itself. Therefore, by drawing from the theoretical-methodological approach of

the communication and work binomial [Figaro, 2024] and Cultural-Historical Activity Theory [Blunden, 2010], we highlight the importance of actions and operations within activities by calling them, in our research, sensitive materialities. These would be the part of social materiality exclusive and immanent to the human body in activity; the socially transformative gestures and expressions of the use of the body-self [Schwartz, 2019], whether voluntary or not, their errors and heuristics which, due to new technologies, have now become susceptible to datafication [Figaro, 2024].

The third issue refers to the social relations that ensure the capture of sensitive materialities in platformized work. The absence of labor regulations and other public policies allows this process to occur stealthily, without the worker's knowledge, consent, recognition, or any form of compensation. This is not an isolated situation, but one that takes part in the larger context of heteromation—that is, the concealment, exploitation, and absent or unfair compensation of human labor which, in doing so, makes computer systems appear fully automated [Ekbia and Nardi, 2017]. Heteromation involves the precarious and poorly paid human participation of data annotators and content moderators in countries of the Global South [Viana Braz, 2021], tasks imposed on users (such as CAPTCHA), the use of their personal and non-personal data for AI training, the unregulated use of sensors [Agre, 1994], and, in our case study, the use of workers' sensitive materialities for these same purposes.

Therefore, the emergence of new productive forces, such as platformized and datafying CSCW software, is accompanied by new relations of production, where the human labor indispensable to the functioning of AI is devalued and acquired through spurious means. On the other hand, the withholding of information (feedback) and compensation to workers by technology companies not only allows them to enable business models that concentrate the means of production and social wealth in their hands, but also grants them unprecedented control over labor. The heteromated datafication of work HCIs implies asymmetric relations at multiple levels. The workforce is doubly exploited—by the direct employer and also by the software owner—and the company using these tools also has its work processes and knowledge captured, in addition to becoming increasingly dependent on companies like Alphabet, Microsoft, or Meta.

This new context, which potentially reaches all types of platformized work, has already been suggested in the Technological Sovereignty and Autonomy challenge of GrandHCI-BR. Dependence on foreign infrastructures for labor mediation—such as the Alphabet and Microsoft ecosystems—threatens not only individual privacy but the very capacity of the Brazilian State to regulate digital productive chains that appropriate sensitive materialities of national labor without social compensation. These issues have not yet been debated more specifically in academia, labor negotiations, or by the State, thus creating a vacuum exploited by tech companies in a way that eludes the very exchange relations of economic liberalism.

5 A multiple case study: the datafication of media work

These new forms of labor exploitation have been identified, initially and primarily, by academic literature that has characterized them under the broad concept of heteromation [Ekbia and Nardi, 2017]. However, in many cases, their description has been chiefly theoretical and based on secondary research data—with the exception, for instance, of Viana Braz [2021]. Nevertheless, studies on the platformization of media labor in Brazil have been the first to produce and analyze primary data regarding the datafication of CSCW, as well as the implications of its protective legislative vacuum. Since 2015, we have been researching how the dominance of Big Tech and its associates—both in providing tools for journalism production and dissemination, and in funding independent media arrangements—has profoundly transformed these professional activities [Figaro, 2018].

In the current cycle of this broad research, our objective is to obtain empirical evidence of the datafication of media workers' labor. Through methodological triangulation, the multiple case study of alternative media professional arrangements has allowed us to monitor how these professionals perform their media activities in total dependence on the tools and affordances of socio-digital platform companies, particularly the Google Workspace suite. These arrangements are initiatives by micro-enterprises and social organizations organized to produce quality information originated from and directed toward specific audiences, often addressing "social causes." These professionals work using their own resources, invest in training, dedicate significant time to managing the "arrangement," and seek an increasingly less diversified range of funding sources. Public grants from national and foreign foundations are among the ways to secure sustainability. However, funding from specific Google and Meta programs, alongside Google AdSense advertising, has grown significantly. Consequently, the research has found that alternative arrangements independent of traditional media conglomerates [Figaro, 2018, 2021; Figaro *et al.*, 2026] have become increasingly dependent on the socio-digital platform companies Google and Meta.

This dependency relationship is constituted, however, in a highly obscured manner for the media professionals within these arrangements. It is not merely a matter of funding ties through programs, training for the use of digital tools, or the steering of editorial themes and priorities. It is about the capture of these medias' productive activities, whose datafied labor produces information for proprietary digital platforms regarding journalistic and communicational know-how. This information is transformed into highly valuable data for Big Tech operations. Demonstrating how the labor activities of medias are datafied is the central objective of this ongoing research project, supported by Fapesp.

The triangulation methodology comprises (1) an academic literature review from different fields of knowledge, (2) a review of corporate documentation—terms of service and privacy policies, financial documents, patents; and (3) the application of empirical instruments such as interviews, self-reporting, and technical observation of the work performed by professionals in the media arrangements. For the empirical stage of technical observation, we developed a technical

device to assess how the datafication of labor activity occurs within platformed tools. The device is applied in fixed-period rounds, following a protocol that includes informing participants about the research objectives, securing approval via an Informed Consent Form, and providing technical support and monitoring for questions or failures in handling the investigation protocol by the participants. We also organized meetings to present partial results and clarify the testing stage, which includes adherence to the self-reporting protocol of labor activities.

Specifically in our datafication experiment, our research is based on automated methods for mirroring network requests and other logs, recording interactions in real-world work contexts, with the use of Google Docs software—part of the Google Workspace platform—as the object of study. The measurement is based on the temporal correlation of two different data streams, aiming to identify causal links between user actions and the generated data. On one hand, a detailed log of human-computer interaction events—clicks, scrolling, and typing—is executed, utilizing DOM (Document Object Model) manipulation to extract coordinates and semantic context. Simultaneously, network monitoring (Network Telemetry) mirrors HTTP/HTTPS requests sent by the browser to Google Docs servers during platform use.

Our device offers control over data granularity. Instead of being limited to high-level metrics, the system records low-latency events which, when cross-referenced, allow for the analysis of the transformation of sensitive materialities into telemetry appropriated by Google. The experiment follows strict ethical and LGPD compliance protocols, with approval from the Ethics Committee of Plataforma Brasil. To address the privacy risks associated with network and keystroke logging, a source-side sanitization mechanism via Regular Expressions (Regex) was implemented, preventing the persistence of sensitive information in the database.

Our experiment is further distinguished by allowing workers to visualize the datafication of their activities in real time through a dedicated dashboard. This functionality aligns with the fundamental HCI principle of feedback—defined by Norman [2008] as the continuous return of information about the results of executed actions. Considering that concealing the datafication of labor is a central feature of heteromation, making this flow public becomes a primary objective of the research, aiming to reduce information asymmetry between the worker and the automated system. This also echoes the concerns highlighted in the Human-Data Interaction, Ethics and Responsibility, and Technological Sovereignty and Autonomy challenges of GrandHCI-BR.

In the current phase, thousands of communications records between the client (Google Docs) and Google's servers are produced and stored for subsequent analysis. The preliminary mapping focuses on identifying the purposes of each communication unit through the analysis of request destination URLs. For example, we found a significant alignment between workflow mining as conceptualized in the business process management literature [van der Aalst *et al.*, 2012], US Google patent 8688499 B1 [Bose and Landa, 2014], and the Google Docs network requests and URLs related to the "People Stack Intelligence Service". This URL ranges from critical operational functions—such as document and user

profile synchronization, saving, and font management—to requests intended for advertising, A/B testing, and people analytics metrics. Google Analytics URLs ending in "collect" and "analytics/web/browser-channel" can be used to map complete workflow journeys, interactions with interface elements (DOM), and engagement levels in real time [Google, nd]. Indications like these suggest that workers may be providing, without their knowledge and consent, information that is unrelated to the specific functioning of the software and that may not benefit them.

In the next steps of the research, we hope to create or utilize already available workflow and task mining tools to deepen and conclude the analysis of requests and their URLs, in order to eventually prove the hypothesis that knowledge about the know-how of media workers can be extracted from the analysis of Google Docs software logs.

6 Conclusion

The discussion and preliminary findings presented here reinforce the GrandHCI-BR agenda for the next decade, which advocates for an HCI more engaged with public policies. The regulation of the production and commercial use of labor data is not merely a technical issue of information security, but a fundamental step toward developing technologies that promote decent work and digital sovereignty in the Global South.

The experiment with media professionals has been validating the hypothesis that CSCW software (such as Google Docs) functions as sensors that capture workflow for purposes potentially unrelated to the user's benefit. Overall, our investigation demonstrates that datafication is not an incidental process, but an industrial activity where productive gestures—clicks, pauses, and heuristics—are formalized as inputs for AI training, management systems, and more. This practice, which we characterize as heteromation—the concealment and devaluation of human activity that makes computers appear as complete automatons—radicalizes power asymmetries where the worker provides the knowledge that, ironically, can be used to replace or control their own activity.

An examination of the Bills of Law and national strategies reveals a critical gap: the lack of recognition regarding the industrial use of labor HCI data and its consequences. While foreign platform companies operate under a *laissez-faire* regulatory logic, appropriating "sensitive materialities" without any reciprocity, the Brazilian State is challenged to guarantee technological sovereignty and labor rights. Data protection must, therefore, expand its scope beyond individual privacy, reaching the social and economic dimensions of labor practice. It must, for instance, ensure the implementation of real-time feedback mechanisms so that workers can visualize how their data is produced, valued, and utilized. It must clearly and detailedly inform users of the platforms' intended purposes for this data, allowing an easy opt-out for these possibilities. Likewise, it must postulate the revision of terms of use and privacy policies to ensure self-determination over HCI data as well. And, finally, scientific research must be fostered alongside public policy agendas that promote system design centered on social justice and decent work.

6.1 Limitations

Our research encountered some limitations. As data analysis is still in its early stages, we do not yet know if the data produced will be sufficient for a more exhaustive investigation. Some information contained in the requests is encrypted, rendering it unavailable for analysis. Although the patent review has been reinforcing our hypotheses, we do not yet know if and how they are effectively implemented in the software of our case study. Finally, while our team includes professionals from computing and other fields, our research group is predominantly from the field of Media, meaning we face particular difficulties in knowing and mastering certain theoretical frameworks that could increase the research's assertiveness.

6.2 Ethical Issues

This article is the result of a collective research effort involving approximately 25 researchers from different states across Brazil, as well as researchers from Colombia and Canada. One of its fronts interviewed professionals from 17 of these arrangements and observed their work through a technical device installed on the computers of 10 professionals within these labor arrangements. All empirical activities involving interviews and work observation were conducted with Informed Consent signed by the participants. The research was approved by Plataforma Brasil under CAAE 79199924.2.0000.5390.

Declarations

Acknowledgements

We thank all the other researchers at CPCT for the collaborative work that made the writing of this article possible.

Funding

We acknowledge the support and funding received from the Fapesp 2022 Lincar call through the approval of our project Datafication of media activity and work of media arrangements: clashes with platform companies, Fapesp process 2022/05714-0. We also thank CNPq for the PIBIC grants received over the past few years, which contribute to the training of young researchers and the advancement of our research.

Authors' Contributions

Luis Gonçalves: Conceptualization, data curation, formal analysis, research, drafting of the original document, and revision and editing. Roseli Figaro: Conceptualization, data curation, formal analysis, research, drafting of the original document, and revision and editing, securing funding, methodology, project management, supervision, and validation.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Research development data is being made available in our public database. Cycle of demands regarding data work, tools, and professionals: <https://zenodo.org/records/14743976>. Survey and analysis report on Google News Initiative tools: <https://doi.org/10.5281/zenodo.18370987> and <https://doi.org/10.5281/zenodo.18371016>. Conceptual presentation of the data mirroring experiment: <https://doi.org/10.5281/zenodo.18370892>. Additional data and information: <https://comunicacaoetralho-datificacao.ea.usp.br/>.

Further relevant information

The authors utilized the Gemini 3 Pro AI to assist in information searches within documents, for the organization of bibliographic references, and for spell-checking and translation, with human review in all instances. The Manus 1.6 Lite and Gemini 3 Pro AIs were used to organize the submission process.

The authors declare that they are aware of and comply with the Citation Diversity Statement in this article.

References

- Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4):218–256. DOI: <https://doi.org/10.1257/mic.20200200>.
- Agre, P. E. (1994). Surveillance and capture: Two models of privacy. *The Information Society*, 10(2):101–127.
- AIPOWW Global Tracker (Brasil) (2025). Essex AI policy observatory for the world of work. University of Essex. <https://www.essex.ac.uk/research-projects/ai-policy-observatory-for-the-world-of-work>. Accessed on September 3, 2026.
- Amorim, H., Cardoso, A. C. M., and Bridi, M. A. (2022). Capitalismo industrial de plataforma: Externalizações, sínteses e resistências. *Caderno CRH*, 35:e022021. DOI: <https://doi.org/10.9771/ccrh.v35i0.49956>.
- Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., and Weyl, E. G. (2018). Should we treat data as labor? moving beyond “free”. In *AEA Papers and Proceedings*, volume 108, pages 38–42. DOI: <https://doi.org/10.1257/pandp.20181003>.
- Auerbach, D. B., Khan, O. H., and Lawrence, S. R. (2017). Systems and methods for selectively storing event data. U.S. Patent US9672232B1. <https://patents.google.com/patent/US9672232B1/en>. Accessed on September 3, 2026.
- Banco Central do Brasil (n.d.). Drex. <https://www.bcb.gov.br/estabilidadefinanceira/drex>. Accessed on September 3, 2026.
- Battle, A. J., Cohn, D. A., Grimes, C. E., and Lamping, J. O. (2019). Deriving and using interaction profiles. U.S. Patent US10387512B2. <https://patents.google.com/patent/US10387512B2/en>. Accessed on September 3, 2026.
- Bednarz, Z., Lewis, K., and Sadowski, J. (2025). ‘it’s not personal, it’s strictly business’: Behavioural insurance and the impacts of non-personal data on individuals, groups and societies. *Computer Law & Security Review*, 56:106096. DOI: <https://doi.org/10.1016/j.clsr.2024.106096>.
- Blunden, A. (2010). *An Interdisciplinary Theory of Activity*. Brill.
- Bose, A. and Landa, E. (2014). System and method for generating business process models from mapped time sequenced operational and transaction data. U.S. Patent 8,688,499. <https://www.lens.org/lens/patent/044-947-992-271-338/frontpage?l=en>. Accessed on September 21, 2026.
- Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. lei geral de proteção de dados pessoais (lgpd). Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Accessed on September 3, 2026.

- Coleti, T. A., Divino, S. B. S., Salgado, A. d. L., Zacarias, R. O., Saraiva, J. d. A. G., Gonçalves, D. A., Morandini, M., and Santos, R. P. d. (2024). Grandhci-br 2025–2035 – gc5: Interação humano-dados, alfabetização de dados e privacidade acessível. In *Anais do XXIII Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais*. DOI: <https://doi.org/10.1145/3702038.3702058>.
- Comitê Gestor da Internet no Brasil (CGI.br) (2023). Sistematização das contribuições à consulta sobre regulação de plataformas digitais. Technical report, NIC.br|CGI.br. <https://cgi.br/publicacao/sistematizacao-das-contribuicoes-a-consulta-sobre-regulacao-de-plataformas-digitais/>. Accessed on September 21, 2026.
- Costa, G. C. d., Rebecchi, C. N., Gonçalves, L. H. N., Raslan Filho, G. S., Boava, A. Z., and Valente, V. C. P. N. (2025). A política de privacidade da plataforma x em análise. *Lumina*, 19(1):10–29. <https://periodicos.ufjf.br/index.php/lumina/article/view/46560>. Accessed on September 3, 2026.
- Cukier, K. and Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*, 92(3):28–40.
- de Santana, V. F., Ferreira, J. J., de Paula, R. A., and de Gusmão Cerqueira, R. F. (2018). An eye gaze model for seismic interpretation support. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/3204493.3204554>.
- Dias, T. (2023). Big techs abraçaram bolsonarismo em ofensiva contra pl das fake news. *Intercept Brasil*. <https://www.intercept.com.br/2023/04/25/big-techs-abracaram-bolsonarismo-em-ofensiva-contra-pl-das-fake-news/>. Accessed on September 3, 2026.
- Ding, Z. and Zivkovic, D. (2017). Client-side click tracking. U.S. Patent US9608892B2. <https://patents.google.com/patent/US9608892B2/en>. Accessed on September 3, 2026.
- Dinkowski, T. G., Avila, D. T., and Thom, L. H. (2025). Fifteen years of process mining in brazil: Current contributions, most used techniques, and challenges. In *Simpósio Brasileiro de Sistemas de Informação (SBSI)*, pages 409–418. SBC. DOI: <https://doi.org/10.5753/sbsi.2025.246519>.
- Ekbja, H. R. and Nardi, B. A. (2017). *Heteromation, and other stories of computing and capitalism*. MIT Press.
- Figaro, R. (2018). *As relações de comunicação e as condições de produção no trabalho de jornalistas em arranjos econômicos alternativos às corporações de mídia*. ECA-USP; Centro de Pesquisa em Comunicação e Trabalho. https://comunicacaoetrabalho.eca.usp.br/publicacoes_cpct/as-relacoes-de-comunicacao-e-as-condicoes-de-producao-no-trabalho-de-jornalistas-em-arranjos-economicos-alternativos-as-corporacoes-de-midia-2/. Accessed on September 3, 2026.
- Figaro, R. (2021). *Discurso jornalístico e condições de produção em arranjos econômicos alternativos às corporações de mídia*. ECA-USP; Centro de Pesquisa em Comunicação e Trabalho. https://comunicacaoetrabalho.eca.usp.br/publicacoes_cpct/discurso-jornalistico-e-condicoes-de-producao-em-arranjos-economicos-alternativos-as-corporacoes-de-midia/. Accessed on September 3, 2026.
- Figaro, R. (2024). Datificação das materialidades sensíveis: captura das atividades cotidianas de trabalhadores da comunicação. *Chasqui: Revista Latinoamericana de Comunicación*, (157). DOI: <https://doi.org/10.16921/chasqui.v1i157.5039>.
- Figaro, R. et al. (2026). Trabalho jornalístico, plataformas digitais e captura de dados: os limites da independência nos arranjos alternativos. In *Anais do 35º Encontro Anual da COMPÓS*, volume 35. <https://publicacoes.softaliza.com.br/compos2026/article/view/15423>. Accessed on September 3, 2026.
- Gerber, C. (2019). Old power in digital garb. In Butollo, F. and Nuss, S., editors, *Marx and the robots: Networked production, AI and human labour*, pages 199–214. Pluto Books.
- Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., and Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, 31(3):410–435. DOI: <https://doi.org/10.1080/0960085X.2021.1927213>.
- Gonçalves, L. H. N. (2023). *Máquinas Automáticas: trabalho, mente estendida e subjetividade sob o fetiche da mercadoria*. Tese de doutorado, Pontifícia Universidade Católica de São Paulo. <https://repositorio.pucsp.br/jspui/handle/handle/36263>. Accessed on September 3, 2026.
- Google (2023). Cloud data processing addendum. Google LLC. <https://cloud.google.com/terms/data-processing-addendum>. Accessed on September 3, 2026.
- Google (2024). Termos de serviço do Google. Google LLC. <https://policies.google.com/terms?hl=pt-BR>. Accessed on September 3, 2026.
- Google (n.d.). Measurement protocol para o Google Analytics 4. Google Developers. <https://developers.google.com/analytics/devguides/collection/protocol/ga4?hl=pt-br>. Accessed on September 3, 2026.
- Grudin, J. (2002). Computer-supported cooperative work: History and focus. *Computer*, 27(5):19–26. DOI: <https://doi.org/10.1109/2.291294>.
- Harper, R. and Randall, D. (2024). Machine learning and the work of the user. *Computer Supported Cooperative Work*, 33:103–136. DOI: <https://doi.org/10.1007/s10606-023-09483-6>.
- Hess, D., Schot, A.-J., and Van Der Schyff, T. (2023). *Getting Started with Microsoft Viva: An End User Guide to Business Transformation*. Apress.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). DOI: <https://doi.org/10.1177/2053951714528481>.
- Kuscher, A. F., Messerly, K. J., and Chen, J. S.-M. (2018). Integrated task items launcher user interface for selecting and presenting a subset of task items based on user activity information. U.S. Patent US9952882B2. <https://patents.google.com/patent/US9952882B2/en>. Accessed on September 3, 2026.
- Kuutti, K. (1991). The concept of activity as a basic unit of analysis for CSCW research. In *Proceedings of the Second European Conference on Computer-Supported Cooperative*

- Work (ECSCW'91), pages 249–264. Springer. DOI: https://doi.org/10.1007/978-94-011-3506-1_19.
- Lastres, H. M. M., Cassiolato, J. E., and Dantas, M. (2025). *Economia política de dados e soberania digital: Conceitos, desafios e experiências no mundo*. Editora Contracorrente.
- Lawrence, S. R. and Wang, N. (2014). Systems and methods for determining an article association measure. U.S. Patent US8667013B1. <https://patents.google.com/patent/US8667013B1/en>. Accessed on September 3, 2026.
- Leontiev, A. N. (1978). *Activity, consciousness, and personality*. Prentice-Hall.
- Lermen, S., Paleka, D., Swanson, J., Aerni, M., Carlini, N., and Tramèr, F. (2026). Desanonimização online em larga escala com llms. In *35º Simpósio de Segurança da USENIX (USENIX Security 26)*, pages 1947–1966. DOI: <https://doi.org/10.48550/arXiv.2602.16800>.
- Menabrea, L. F. (1842). Sketch of the analytical engine invented by Charles Babbage. Translated by Ada Lovelace. Original work published 1842. https://johnrhudson.me.uk/computing/Menabrea_Sketch.pdf. Accessed on September 3, 2026.
- Milvaney, D. L. et al. (2024). Document status management system. European Patent EP3803737B1. <https://patents.google.com/patent/EP3803737B1/en>. Accessed on September 3, 2026.
- Norman, D. A. (2008). *O design do dia a dia*. Rocco.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
- Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso Books.
- Perry, M. K. (1978). Vertical integration: The monopsony case. *American Economic Review*, 68(4):561–570. <https://ideas.repec.org/a/aea/aecrev/v68y1978i4p561-70.html>. Accessed on September 21, 2026.
- Piatetsky-Shapiro, G. (2000). Knowledge discovery in databases: 10 years after. *SIGKDD Explor. Newsl.*, 1(2):59–61. DOI: <https://doi.org/10.1145/846183.846197>.
- Posner, E. and Weyl, E. G. (2018). *Radical markets: Uprooting capitalism and democracy for a just society*. Princeton University Press.
- Russell, S. J. and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition.
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). DOI: <https://doi.org/10.1177/2053951718820549>.
- Schwartz, Y. (2011). Conceituando o trabalho, o visível e o invisível. *Trabalho, Educação e Saúde*, 9:19–45.
- Schwartz, Y. (2014). Motivações do conceito de corpo-si: Corpo-si, atividade, experiência. *Letras de hoje*, 49(3):259–274.
- Schwartz, Y. (2019). O enigma do corpo no trabalho. *Ergologia*, 22.
- Smith, A. (2023). *A riqueza das nações*. Nova Fronteira.
- Soegaard, M. and Dam, R. F. (2012). *The encyclopedia of human-computer interaction*. The Encyclopedia of Human-Computer Interaction.
- Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. Harper & Brothers, New York.
- Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1). <https://jodi-ojs-tdl.tdl.org/jodi/article/view/269>. Accessed on September 3, 2026.
- van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., and Baier, T. (2012). Process mining manifesto. In *Business Process Management Workshops*, pages 169–194, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-642-28108-2_19.
- Viana Braz, M. (2021). Heteromação e microtrabalho no brasil. *Sociologias*, 23(57):134–172. DOI: <https://doi.org/10.1590/15174522-111017>.
- Viljoen, S. (2021). A relational theory of data governance. *Yale Law Journal*, 131:573. DOI: <https://doi.org/10.2139/ssrn.3727562>.
- Wong, W. H., Duncan, J., and Lake, D. A. (2025). Why data about people are so hard to govern. *Regulation & Governance*, 19(1):236–252. DOI: <https://doi.org/10.1111/rego.12591>.
- Wu, X., Wu, E., and Klebanov, M. A. (2024). Dynamically configurable client application activity. U.S. Patent US12160481B2. <https://patents.google.com/patent/US12160481B2/en>. Accessed on September 3, 2026.
- Zhang, F., Zhuang, P., and Meng, Y. (2018). Optimize UI data cached for future actions. Chinese Patent CN108605053A. <https://patents.google.com/patent/CN108605053A/en>. Accessed on September 3, 2026.
- Zuboff, S. (2021). *A era do capitalismo de vigilância*. Editora Intrínseca.