weizenbaum
DISCUSSION PAPER

September 2025
50

Lea Schneidemesser, Daniel Schneiß, Jana Heim

Dynamics of Data Work in AI Implementation Processes

AUTHORS

Lea Schneidemesser \\ Weizenbaum Institute, Berlin \ Goethe University Frankfurt am Main
schneidemesser@soz.uni-frankfurt.de

Daniel Schneiß \\ Weizenbaum Institute, Berlin \ Kiel University
daniel.schneiss@weizenbaum-institut.de

Jana Heim \\ Weizenbaum Institute, Berlin \ WZB Berlin Social Science Center
jana.heim@wzb.eu

ABOUT THIS PAPER

This paper was produced as part of the research programme of the research group 'Working with Artificial Intelligence' at the Weizenbaum Institute for the Networked Society. The research group investigates implementation processes, design and use of AI technologies in companies, focusing on the changes in job content, work organisation, qualification structures and co-determination. The research programme includes case studies of companies, analysis of selected issues relating to human-AI cooperation in laboratory experiments, and a company survey.

ABOUT THE WEIZENBAUM INSTITUTE

The Weizenbaum Institute is a joint project funded by the German Federal Ministry of Research, Technology and Space (BMFTR) and the State of Berlin. It conducts interdisciplinary and basic research on the digital transformation of society and provides evidence- and value-based options for action in order to shape digitalization in a sustainable, self-determined and responsible manner.

Weizenbaum Discussion Paper

Dynamics of Data Work in AI Implementation Processes

Lea Schneidemesser, Daniel Schneiß, Jana Heim

Abstract

The discourse on automation and artificial intelligence (AI) highlights the critical yet underexplored role of data work. While previous studies have focused on data work in the production of AI systems, this article investigates its significance in subsequent implementation phases. Drawing on two in-depth case studies from traditional industries in Germany – one in manufacturing and one in administration – we explore the labor and power dynamics inherent in these processes. Our findings reveal that AI implementation necessitates additional data work which cannot be outsourced due to its reliance on organization-specific knowledge. This dependency fosters new labor relations and power dynamics among development companies, organizational management, and workers, often leading to tensions and negotiation challenges. We introduce the concept of *data work facilitation* to describe the unique labor emerging from these dependencies and identify potentials for a new power resource workers could employ – *data work bargaining power*.

Contents

1 Introduction

In discussions on automation and artificial intelligence (AI), critical social research has brought to the fore the criticality, complexity, and conditions of human labor that enable the development of AI-based systems. This is often described using the concept of data work. Inspired by these critical interrogations, we argue that data work does not only play a significant role in the development phase of AI systems but also in their subsequent implementation and customization phases within organizations. These are key moments along the AI supply chain, which, however, remained less empirically investigated so far. The functioning of AI systems is often crucially dependent on the data input from within implementing organizations, necessitating workers' involvement in the production, curation, and analysis of such data.

In our article, we investigate data work in AI implementation processes in companies. More specifically, we explore how the introduction of AI systems into organizational processes and labor practices requires extensive work connected to the production, curation, and evaluation of data and to the facilitation of such data work. We identify emerging dependencies between actor groups involved in implementation phases of AI-based technologies, namely the development companies, the management of the implementing organizations, and workers as new users of AI systems. We also shed light on the power dynamics that emerge in such AI implementation phases. Because the implementation and customization of AI systems within organizations require organization-specific knowledge and skills to integrate, curate, and evaluate data from internal processes, such data work cannot be easily outsourced. Due to this constitutive dependency, the implementing and developing companies rely on workers in implementing organizations to conduct data work. We suggest describing this dynamic as *data work facilitation* and analyze how it is negotiated in practice.

To understand data work in AI implementation processes, our work is guided by two research questions. RQ1: What forms of data-related work are required in AI implementation processes in organizations and what labor relations emerge? RQ2: What power relations and dependencies characterize organizational AI implementation processes?

Empirically, our article adds to a deeper understanding of the dynamics, actor constellations, and conditions of data work in AI implementation processes through an analysis of two in-depth case studies of such implementation processes in companies in traditional industries in Germany. In the first case, an AI system is introduced to a manufacturing work context to identify machine malfunction and measure machine performance. In the second case, an AI system is introduced to an administrative work context, analyzing how processes are conducted and identifying optimization potential. Despite significant differences between the two cases, we identify increasing importance of workers' participation in related data work tasks for the developing company and the implementation companies' management, leading to changing power dynamics between involved actors, and potentials for conflicts in the introduction phase of AI-based systems.

Conceptually, we engage with two strands of research. Firstly, we engage with research that investigates human labor that facilitates the functionalities of AI systems, including data work (Muldoon et al., 2024; Miceli et al., 2022; Newlands, 2021; Gray & Suri, 2019). Secondly, we contribute to research on data work in 'end use' contexts within organizations that increasingly integrate data and AI-based technologies (Jarke & Büchner, 2024; Donatz-Fest, 2024; Pine et al., 2022; Bossen et al., 2019). By focusing on moments of AI implementation in conventional industries in Germany, we add empirically grounded insights into a context of data work less studied so far. In doing so, we contribute to research on automation and the role of AI at the workplace (Krzywdzinski et al., 2024; Jarrahi et al., 2021; Kellogg et al., 2020).

2 Data work from an AI implementation perspective

Contemporary AI systems are ultimately about the processing of (digital) data[1] in increasingly advanced ways. In theory, these forms of processing data are intended to mirror human cognition. In practice, these systems combine highly advanced statistical modeling techniques in which part of the calculative process is given to the algorithmic system itself (Mackenzie, 2017). In their application, AI systems often constitute an add-on technology to an already existing software and hardware ecosystem. In this way, they become part of a 'technology toolbox' (Nitzberg and Zysman 2022: 1757) in which the intertwinement of algorithms, data, and platforms plays a crucial role. The data input remains, nonetheless, one of the most significant chokepoints that influences the functionalities of AI systems. From this perspective, '[d]ata is the critical infrastructure necessary to build Artificial Intelligence (AI) systems. Data largely determines performance, fairness, robustness, safety, and scalability of AI systems' (Sambasivan et al. 2021: 1).

Contrary to repeated calls about the rise of automation, human labor remains a key component in developing, operating, and maintaining data- and AI-based systems (Pink et al., 2022). This is often discussed as data work, defined as the tasks that are concerned with the collection, classification, curation, and labeling of data. In recent years, research has studied the criticality, complexity, and conditions of such human labor and showed how it is distributed among a globally dispersed workforce and structured by power asymmetries and precarious working conditions (e.g. Muldoon et al. 2024; Ahmad 2023; Miceli et al. 2022; Gray and Suri 2019). Research increasingly attends to human labor and data work specifically in 'end use' contexts of algorithmic and AI-based systems (e.g. Jarke and Büchner 2024; Donatz-Fest 2024; Pine et al. 2022; Bossen et al. 2019) and emphasizes that additional data-related tasks emerge in the introduction of new technologies into organizational processes.

1 In this context, we understand 'data' primarily as digital data. For reasons of simplicity, we speak only of data except in those cases in which other forms of data are addressed.

In this section, we engage with these two strands of literature on human labor along different phases of AI supply chains, in which 'several actors contribute towards the production, deployment, use, and functionality of AI technologies' (Cobbe, Veale, and Singh 2023: 1). Whereas the first strand of literature emphasizes the role of outsourced, invisibilized, and precarious data work conditions, it does not engage further with the data work in the implementation phase. The second strand of the literature that focuses on these implementation processes does not discuss, however, the role of data work within companies in which these systems become embedded. The implementation of new technologies within companies affects existing forms of work organization, labor process, and managerial control, a process that is always conflictual and demands negotiations between the affected actors (Krzywdzinski et al., 2024; Bailey & Barley, 2020; Kellogg et al., 2020). In the following section, we engage constructively with the existing literature on data work and outline three levels of analysis that particularly shape data work in the implementation phase of AI systems in corporate settings. These underline a contested need to mobilize data work responsibilities in AI implementation processes that increasingly transgresses the organizational boundaries of developing and implementing companies. Building upon these research insights and our empirical observations, we make a conceptual contribution that centers the power dynamics that arise in data work within implementation phases of AI-based systems.[2]

2.1 Data work in the production and maintenance of AI systems

AI-based systems depend significantly on the quality of their underlying data sets. Machine learning technologies specifically need to be developed along pre-structured datasets, and the systems' outputs need to be constantly evaluated and improved. In many ways, ensuring the quality of the datasets but also the evaluation of the accuracy of AI models remains a specialized time- and labor-intensive activity. Since many organizations lack the expertise and resources for doing these activities at the scale needed, this has led to the development of a global, complex network of AI supply chains (Cobbe et al. 2023) in which manifold forms of human labor are distributed among a dispersed workforce, mainly via business process outsourcing (BPO) companies and digital platforms (Muldoon et al., 2024).

This includes human labor in the production of AI systems, often described as data work (Muldoon et al., 2024; Miceli & Posada, 2022; Newlands, 2021), in content moderation on social media platforms (Ahmad, 2023; Roberts, 2021), and in facilitating various other AI-based and digital services (Newlands, 2021; Gray & Suri, 2019). While development companies often attempt to invisibilize such labor 'behind' the systems' interfaces (Newlands, 2021), critical research has investigated and brought to the fore the labor conditions, power dynamics, and actor constellations of such work. Researchers have investigated how the various forms of human labor enabling such technologies are embedded in organizational and broader structural power asym-

2 While we conceptually and empirically center implementation phases of AI-based systems and consider them key moments along AI supply chains, we acknowledge the dynamic, iterative, and intertwined modes of developing and deploying such technologies.

metries. These power dynamics determine how outsourced and globally dispersed data work is conducted (Miceli et al. 2022; Ahmad 2023). At the same time, some forms of contestation and resistance within outsourced, platform-organized labor exist (Vilasís-Pamos et al., 2024; Ferrari & Graham, 2021; Irani & Silberman, 2013).

Attending to the complexity of human labor for AI systems, scholars have developed different typologies to make sense of such work along AI supply chains (Muldoon et al., 2024; Posada et al., 2023; Newlands, 2021; Tubaro et al., 2020). Concerning data work in the production of AI systems specifically, Muldoon et al. (2024) distinguish different types of internal and external 'AI data work' (crowdsourced/self-employed, employed) and types of 'AI data work institutions' (companies, platforms, BPO centers) along an 'AI data pipeline,' which they define as 'the set of data processing activities necessary to integrate datasets into the training and testing of machine learning models' (Muldoon et al. 2024: 2). An AI data work institution describes 'an organisation that arranges for AI data work to be undertaken either by employees of the organisation within a designated facility or by geographically dispersed independent contractors' (Muldoon et al. 2024: 4).

Even though Muldoon et al. (2024: 8) discuss 'internal data services' through which 'AI companies may also employ in-house workforce,' the focus still lies on the production of AI systems. Shestakofsky (2024: 11) emphasizes that 'data work is increasingly performed in organizational settings that provide structures designed to increase the consistency and accuracy of output.' However, the implementation of AI systems into already existing work organizations requires adaptation to the organization and its labor process. This alignment and 'fine tuning' process demands additional data work in between the implementing and developing company. Here, both workers of the developing organizations and particularly workers of the implementing organizations interact with different interfaces of an AI system, and tasks related to the production, curation, and evaluation of data often add to their existing workloads. In this customization and adaptation phase of AI systems, which Newlands (2021) describes as 'AI coproduction,' actor constellations and power dynamics emerge that are different from the AI production phase.

In this paper we focus on data-related work tasks, power dynamics, and actor constellations in the implementation of AI systems, an empirically less studied part of the AI data pipeline. Thereby, we conceptually distinguish between demands for customization, the role of data work from within the implementing companies, and the inter-firm relations in evaluating the systems' functionalities.

2.2 Data work in the digital transformation of organizations

While existing research has centered the organization of data work via digital platforms and BPO companies, the dynamics of data work within organizations in digitalization processes remain less studied. Especially researchers in the sociology of organizations have investigated how organizations and organizational practices become increasingly datafied and algorithmized (Jarke & Büchner, 2024; Kostis et al., 2024). The process of datafication, in turn, influences how organizations function, how knowledge is processed, and how labor processes are restructured (Alaimo & Kallinikos, 2022; Alaimo et al., 2020). In contrast to data work organized via digital platforms and BPO companies, data work in digitalizing organizations receives a broader meaning relating to all kinds of practices through which data is made 'useful and meaningful' (Pine et al. 2022: 1). The dissemination of data-based technologies in organizations thereby often leads to a distribution of data work tasks among workers across different roles (Jarke and Büchner 2024: 2). Here, data work encompasses highly diverse tasks and activities that often become an additional responsibility and increase in workload. At the same time, they are crucial in datafication processes of organizational procedures.

Investigating datafication in the public sector, Jarke and Büchner (2024), for instance, develop the notion of 'data care arrangements' to describe the contingent coordination of heterogeneous data work tasks that are often mundane but become crucial for digitalizing organizations. Thereby, they emphasize the different ways in which actors in organizations may take on responsibility for data-related work by ascribing meaning and value to it, for instance by considering data to be self-evident or actionable (Fiore-Gartland and Neff 2015 in Jarke & Büchner 2024: 4). As Jarke and Büchner (2024: 5) point out, datafication processes in organizations can often entail conflicts, as understandings of the aims and implications of datafication may differ across organizational departments and actor groups that are disparately affected by such processes and the work they entail. This points to a need to facilitate these work tasks within and in between organizations that develop and implement AI systems and offers a perspective attentive to conflicts between actor groups involved in data work in AI implementation processes.

Jarke and Büchner (2024) discuss 'data care arrangements' in public sector organizations. In companies, the managerial control of the labor process and work organization is often more distinct (Vidal, 2022). This becomes especially apparent with regard to rationalization and optimization goals underlying the introduction of digital technologies, goals that are also mediated through the use of these systems. In contrast to data work in the public sector, data work here takes place at the intersection between the rationalization and control imperatives of profit-oriented companies (Nies, 2021). This points to recent discussions on algorithmic management systems in which the role of data science expertise remains dependent on domain knowledge from within the implementing companies (Krzywdzinski et al. 2024).

In addition, the development and subsequent integration of domain- and organization-specific AI systems often constitutes a knowledge problem for involved actor groups, as significant information on the systems' potential functionalities can be determined only within the implementation process (Kostis et al. 2024). This relates to potential conflicts stemming from 'ignorance of domain knowledge, domain specificities, and of the availability, features, and meanings of existing data, but also to ignorance of AI possibilities and limitations' (Kostis et al. 2024: 39). Here, organizations are confronted with the challenge of bridging such divides between domain knowledge and technical knowledge (Kostis et al. 2024: 40) in AI implementation and customization phases. To solve this knowledge problem, system developers and the involved actors in the implementation context need to work together to achieve an integration between domain and technical knowledge. In this context, data work, embedded in organization-specific knowledge, constitutes an increasingly important aspect in the integration and adaptation of AI-based systems in organizations. Thinking about data work as part of contested control mechanisms within organizations and the contingent *mobilization* of responsibilities for data work allows us to understand the specific dependencies and power dynamics that arise between actor groups in the implementation of digital technologies.

2.3 Data work in the implementation of AI systems in organizations

The functionalities of algorithmic systems are significantly influenced by the quality of their underlying data and the meticulousness with which this data is produced and curated, while 'poorly curated data may magnify biases or inaccuracies in later stages' (Muldoon et al. 2024: 8). In many cases, context-specific customization of AI systems is required to adapt them to the unique requirements and nuances of an organization, which underlines the importance of organizational context-specific data production and integration of such data into AI systems. Hence, while digital data can increasingly transgress organizational boundaries and 'decenter organizations' (Alaimo & Kallinikos, 2022), we underline that in the development and deployment of AI systems, organization-specific data remains crucial and its production and curation oftentimes labor-intensive. Pre-trained AI models need to be further trained on data sets from the particular organization to create outputs that are meaningful and accurate for that organization. This becomes even more crucial as complex AI-based systems often require ongoing careful oversight of both inputs and outputs, since their functioning differs from more stable algorithmic systems with predefined rules. Research on data work in organizations that introduce data- and AI-based systems has increasingly attended to the particularities of such data work, but with a primary focus on the public sector, such as education, social work, and policing (Jarke and Büchner 2024; Donatz-Fest 2024; Pine et al. 2022). Less empirical attention has been placed on the dynamics of data work in the introduction of AI systems in private sector companies in Germany.

As we argue, such work requires extensive resources, coordination efforts, and mobilization of organization-specific knowledge. Moreover, data work tasks increasingly transgress organizational boundaries in AI implementation phases, where new dependencies within and between the development companies and the implementing organizations arise. To more thoroughly make sense of the forms, complexities, and dynamics of organizational data work (Jarke et al., 2022) in the context of AI implementation in companies, we offer three conceptual contributions.

Firstly, in line with the notion of the AI data pipeline, we suggest that such an understanding of the complex and interwoven 'set of data processing activities necessary to integrate datasets into the training and testing of machine learning models' (Muldoon et al. 2024: 2) can also inform an understanding of data-related work tasks in implementation phases of AI-based systems, while attending to the different labor conditions, power dynamics, and dependencies of involved actor groups in this context. Thus, we suggest expanding this concept to identify and contextualize human labor along a *data pipeline in AI implementation*.

Secondly, we suggest that the particular *inter- and intra-firm relations and actor constellations* of such data work in AI implementation require more conceptual and empirical attention. In our case studies, this includes workers of the development companies of these systems, the management of the implementing organizations, and workers in the implementing organizations. Here, the specific modes of collaboration in and between developing and implementing companies can significantly influence data production and quality, affecting the customization and functionality of AI systems. The relations and actor groups we identify differ significantly from other AI data work institutions (Muldoon et al., 2024), such as BPO companies and digital platforms in earlier phases of the AI supply chain.

Thirdly, we show that in the implementation phase of AI-based systems, crucial parts of data work cannot be easily distributed but have to be performed by specific employee groups within implementing organizations, as these tasks require organization-specific knowledge and expertise of workers. This stands in contrast to outsourceable, globally dispersed modes of specialized data work in phases of the development and maintenance of AI systems. This creates a new dimension of dependency of organizations on their workers, as they are required to be involved in data work practices, for instance in curating data, evaluating AI model outputs, and in communicating with machine learning engineers. This constitutive dependency results in two dynamics. On the one hand, we suggest understanding this dependency on workers' specific knowledge and data work can offer spaces for a potential new power resource (Schmalz & Dörre, 2014; Silver, 2003), which workers might draw on to negotiate the particular conditions of performing these new tasks – which we suggest to understand as *data work bargaining power*. On the other hand, for development companies and the management of implementing organizations this dependency results in a necessity to undertake what we describe as *data work facilitation*. This constitutes time-intensive work with the aim to produce, curate, and evaluate data within organizations.

These three conceptual aspects structure the subsequent presentation and discussion of our empirical cases of AI implementation processes, particularly focusing on the (inter-)organizational settings and dynamics between different involved actor groups of both developing and implementing companies and on shifts in power dynamics in between these groups.

3 Research design and research methods

Empirically, our analysis is based on two in-depth case studies (Priya, 2021). We follow and reconstruct the implementation process of two AI systems from development to use. More specifically, we comparatively look into their implementation processes in two different work contexts in Germany. The case studies come from two different industries, differ in their technical functionalities, interfaces, underlying data, and scenarios of use. In the first case, the AI system is introduced to a manufacturing context to monitor machine failure by measuring machine vibration and temperature. In the second case, the AI system is introduced to administrative and taxation work contexts to conduct data-based productivity and optimization analyses.

The AI systems used in the two cases belong to the same category of applications: Monitoring, analysis, and optimization of processes. Within this subfield they represent most different cases, regarding types of processes and data, the domain knowledge required in the industry, and the professional background of involved workers in the implementing companies. In the manufacturing case study, the AI system monitors manufacturing processes, collecting data from physical machines and digitalizing the gathered information with the help of sensors and Internet of Things (IoT) technologies. The domain knowledge of this industry revolves around tissue paper production processes and machine maintenance, and workers have a technical background. For the administrative service case study, data does not need to be digitalized but is already available in digital formats. The AI system is applied to the handling of business process data. Workers involved in the projects have domain knowledge in back-office administrative tasks such as taxation, quotation processes, or payroll management. Finally, the cases differ in the time period of implementation. In the manufacturing case, the system is intended to be used long-term, while in the administrative services case, the system is only implemented for a limited project phase. Choosing such different cases allows us to evaluate the data work, interorganizational relations, and power dynamics regarding their specificity or potential generalizability.

Our original qualitative empirical data consists of two elements for each case study: Semi-structured interviews and field observations gathered in a workshop with the development company in the administrative service case study and three company visits at the implementation company of the manufacturing case study. Our empirical data was supplemented by an analysis of publicly available documents to triangulate the data. Semi-structured interviews are the primary method of data collection due to the explorative nature of this study and the lack of previous research on data work in AI implementation in traditional industries (Brinkmann, 2020).

In both cases, we reconstructed the functionality and scenarios of use of the AI system through extensive interviews (1–2h) with involved actors. We interviewed data scientists, machine learning engineers, machine maintenance experts, and project and customer relations managers of the development companies and technicians, plant and company management, and works council members in the implementing companies. The point of entry and main focus of the interviews differ for both case studies. For the manufacturing case study (CS1) the users of the AI system in the implementing company – the machine maintenance teams – are the main focus of our interviews because we identified them as the main actor group involved in data work tasks. For the administrative service case study (CS2), the founder team of the development company was the focus of the interviews, since they, as technical experts, are strongly involved in data work tasks, separate from data production which takes place in the implementing company. Even though our data show greater depth for one of the two organizational contexts, AI development or implementation, for both case studies we consider different perspectives on the respective AI system and its implementation process, conducting interviews with all relevant actor groups in development and implementation companies.

We conducted nine interviews for the manufacturing case study and five interviews for the administrative service case study. The interviews were transcribed or documented as protocols and the transcriptions were integrated with our field notes from company visits and the workshop with the development company of case study 2, additional notes from the interviews themselves, and document analysis. For the data analysis, we followed a mix of inductive and deductive coding inspired by qualitative content analysis (Mayring, 2014). In all interviews, we looked for closer descriptions of the interaction with the AI system, its perceived functionalities, types of data-related tasks, and the different ways the actors were involved in the implementation process.

Table 1: Overview of interviews

No	ID	Date	Mode	Company Pseudonym	Interviewee Role	In
				Case Study Manufacturing		
1	CS1_1	22.03.2023	In Person	TissueCo	Machine Vibration Analyst 1, Location 1	Germany
2	CS1_2	22.03.2023	In Person	TissueCo	Technician; Machine Vibration Analyst 2	Germany
3	CS1_3	21.05.2023	Virtual	ManufacturingAI	Customer Success Manager	Netherlands
4	CS1_4	11.07.2023	Virtual	ManufacturingAI	Technical Success Manager	USA
5	CS1_5	06.05.2024	In Person	TissueCo	Head of Works Council; Technicians Location 2	Germany
6	CS1_6	06.05.2024	In Person	TissueCo	Production Manager Location 2	Germany
7	CS1_7	26.09.2024	In Person	TissueCo	ManufacturingAI's 'Champion' Location 3	Germany

8	CS1_8	26.09.2024	In Person	TissueCo	ManufacturingAI's 'Champion'; Technician Location 3	Germany
9	CS1_9	26.09.2024	In Person	TissueCo	ManufacturingAI's 'Champion'; Maintenance Manager Location 3	Germany
			Case Study Administrative services			
10	CS2_1	06.09.2023	Virtual	AdminAI	CTO	Germany
11	CS2_2	20.02.2024	Virtual	AdminAI	CEO	Germany
12	CS2_3	02.05.2024	Virtual	AdminAI	CEO	Germany
13	CS2_4	06.06.2024	Virtual	AdminAI	CEO; CTO	Germany
14	CS2_5	21.08.2024	Virtual	TaxationCo	Managing Partner; Head of IT	Germany

4 Case studies: Data work in the implementation of AI systems in manufacturing and administrative services

The following chapter will introduce our two case studies of AI implementation processes. Thereby, we will first describe the cases and highlight different customization processes and their dependency on respective forms of data work. We reconstruct and visualize these as a *data pipeline in AI implementation*, corresponding to the *AI data pipeline* in the production of AI-based systems (Muldoon et al., 2024). This helps us identify different kinds of data work facilitation and a changing division of labor that emerges between the AI system developer, the implementing company's management, and the workers involved in the implementation and use of the system.

4.1 Data work in manufacturing – conflicts about mobilizing expertise

The manufacturing process has been identified as a vast source of data in recent years that can be used for optimization and cost saving. At the same time, manufacturing environments are a difficult field to collect and make sense of data when external service providers are involved, because companies often consider data on production processes competition-relevant and therefore sensitive information (Butollo & Schneidemesser, 2022). Even once the data privacy issue is resolved, the heterogeneity of machine types and generations which record data in inconsistent data formats requires additional work to integrate them (CS1_6) or data has to be collected anew by installing external sensors.

Our first case study is a development company which we call ManufacturingAI[3]. ManufacturingAI has developed a predictive maintenance system that performs machine and process monitoring to reduce incidences of machine breakdowns and therefore loss of production capacity. The company employs machine learning techniques for an initial identification of anomalies in the data transmitted from the machines. For the monitoring, physical process data, specifically machine vibration and temperature data, are collected via sensors from the manufacturer's production equipment. Based on this data, ManufacturingAI generates alerts about anomalies in the machine operation and suggestions on how to remedy them to optimize the production volume. The implementing company is a globally operating hygiene company, which we call TissueCo, with production locations in Germany and Europe.

TissueCo has implemented ManufacturingAI's software to monitor the condition of its tissue machines. ManufacturingAI offers its customers a comprehensive service, from the installation of sensors on the machines and other hardware, to data collection, analysis, and evaluation. Real-time data from TissueCo's machines is compared to the data already collected by ManufacturingAI across all of their projects, using machine learning to detect anomalies in machine functioning. ManufacturingAI's business model and AI-based software therefore rely on access to customers' physical process data and are dependent on high quality of that data.

Machine data, more specifically data from TissueCo's machines, is the critical component for the machine monitoring software offered by ManufacturingAI. To arrive at the intended result – less machine downtime – a number of data-related tasks and processes need to be performed by workers of both organizations along the data pipeline to guarantee the necessary quality of data. In this process, the AI system becomes customized to the specific goals of the implementing company.

At ManufacturingAI, a customer success manager (CSM), a technical success manager (TSM) and machine vibration analysts (VA) are involved. They work remotely on a global scale and do not visit TissueCo's plants regularly. At TissueCo the company and plant management are involved, as they invested in the system and expect a return, yet the central actors and users of the system are the maintenance staff responsible for the machines monitored with ManufacturingAI's software.

The implementation of the machine monitoring system can be divided into the following steps that require the execution of specific data work tasks (see also figure 1). First, sensors on the machines and IoT nodes in the plant are installed to make the temperature and vibration of the machines available in digital form. To digitize this information, ManufacturingAI sends an installation team to TissueCo's plants to install the required hardware (CS1_1; CS1_3). Then, data is produced and collected automatically and sent directly to ManufacturingAI's servers, where it is analyzed by a machine learning algorithm, comparing the incoming data to the historical data from all similar machines ManufacturingAI monitors worldwide (CS1_4). When the software detects a machine anomaly, it sends a notification to the machine vibration analysts of ManufacturingAI who evaluate the model output: They interpret the detected anomaly and decide whether this is an incident that the customer should be informed about. If the alert is forwarded, the maintenance team and plant management at TissueCo receive a notification, inform-

3 All company names are pseudonyms (see also Table 1) to ensure the anonymity of our interview partners.

ing them about the detected issue, its location, potential causes, and suggestions for action. A maintenance worker then goes to the machine and evaluates the AI model output, investigates the issue, potentially performs repairs or schedules it for the next regular machine shutdown. The worker then returns to the software system and provides feedback whether the detected anomaly corresponded with an actual fault of the machine and how they acted on it. This feedback curates the AI model output.

The information on the actual situation on the shopfloor provided by the maintenance staff is used to improve the data produced by the machines and to retrain the machine learning algorithms (CS1_4). Thus, ManufacturingAI is 'blind' to the actual situation on the shopfloor and depends on TissueCo's workers' assistance and knowledge to correctly interpret the AI outputs, as explained by a TSM from ManufacturingAI:

'We'll see the machine shaking, but we won't see that the valve has closed. I'll ask the customer, "Hey, did anything happen in the process to cause this machine to vibrate this much?" Then he'll say, "Yes, the valve was closed by somebody mistakenly." Then we put that into the data and into the platform' (CS1_4).

The sketch of the implementation process of ManufacturingAI's software shows that the engagement of TissueCo's maintenance staff is crucial to refine the automatically collected data and further improve the AI system's accuracy, as ManufacturingAI's workers do not see what is actually happening with TissueCo's physical machines. In the eyes of ManufacturingAI, the ideal customer is one 'that's engaged, involved, responding, and attending meetings [...], believes in the system, embraces it, responds to the alerts, provides us input and insight into what he's seeing out there, and what he's hearing' (CS1_4). In-depth knowledge about machine vibration analysis on the side of the customer is negligible, according to ManufacturingAI (CS1_4). This constellation makes data work facilitation an important component of the data pipeline, mainly performed by the CSM and TSM of ManufacturingAI and Manufacturing AI's 'champions'[4] in the implementing organizations (CS1_7). It has the aim to engage the relevant employees to interact with the AI system so that they perform the necessary data work. In the implementation of ManufacturingAI's software in TissueCo's plants, data work facilitation is important mainly at two moments in the data pipeline: In the beginning of the project, even before the installation of the hardware, and during the data curation process when the verification of analysis results by TissueCo's workers is required.

The CSM of ManufacturingAI described that the initial planning phase is a crucial moment to involve the actor groups that need to support the system in the long run, to guarantee what they understand as the successful implementation of the software. Otherwise, in their perspective, 'it can take years to overcome [...] resistance and to build trust' (CS1_3) which would affect the quality of their interactions with the system and the quality of data. Therefore, the CSM conducts workshops with the involved actor groups, explains the system's functionalities and possible goals, and trains them in using the platform: How to interpret the alerts and how to provide feedback in the system (CS1_3).

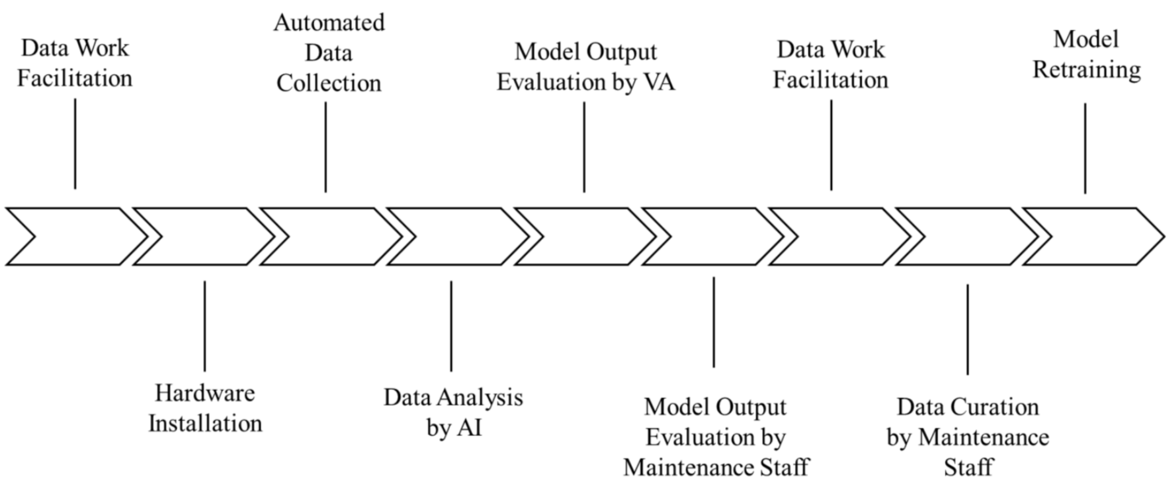
---

4 In both case studies, the role of an internal 'champion' describes an employee in the implementing organization who is the main contact person for the developing company and responsible for the implementation process. This organizational role of a 'champion' is not specific to our case studies but a common strategy in technology implementation processes within organizations.

One of TissueCo's manufacturing plants (location 1) that employs internal machine vibration analysis experts in addition to their regular maintenance teams demonstrates how excluding relevant stakeholders in the initial phase of the project and a critical stance on the system can impact the interactions with it. The dependency on the mobilization of the workforce might lead to conflicts in the implementation. The machine vibration analysts in location 1 have used a different software for monitoring their machines for many years, which they find helpful in their work. They experienced the implementation of ManufacturingAI's system as a top-down decision, since they were not consulted in the process, and they are interacting with ManufactuingAI's system only sporadically, mainly still relying on and preferring the old system. They have developed significant expertise in analyzing machine vibration and temperature data, deriving indicators for emerging machine breakdowns from it (CS1_1; CS1_2). This is a task ManufacturingAI is taking on for their software system, making workers' expert knowledge in this field less relevant in the operation of their system.

The process in which data work facilitation is most important in the long term is the curation of the data by TissueCo's maintenance staff after receiving an alert from ManufacturingAI. In this step of the data pipeline, the maintenance staff is required to validate or correct the alerts sent by ManufacturingAI's VA's based on the actual situation on the shop floor. Data work facilitation at this point is undertaken by ManufacturingAI's TSM. They meet with the maintenance team and the operators of each machine and each plant in regular intervals. These meetings are scheduled to discuss problems and questions that might hinder TissueCo's workers from engaging with the AI system and to collect further information on the maintenance activities on the shopfloor which might not have been communicated via the platform (CS1_4). Even for those maintenance workers who do not object to the system fundamentally, interacting with it constitutes additional work: This includes meetings with CSM and TSM, monitoring the condition of the sensors, observing the AI system, and evaluating alerts sent by ManufacturingAI's VA. Having a tight schedule in general and emergency repairs to be performed can lead to decreased priority for engaging with ManufacturingAI's workers and systems (CS1_2; CS1_5).

Figure 1: Data pipeline in AI implementation: Manufacturing[5]



5 This is a schematic representation which takes into account conceptual simplification. The actual process is more complex and not as linear, including iterations, overlaps, and feedback loops.

The case study on the cooperation between ManufacturingAI and TissueCo shows the dependency on new data work and facilitation tasks in the implementation process. Data work ranges from the installation of hardware to model output evaluation by ManufacturingAI's VA as well as data curation by the maintenance staff of TissueCo. The circumstance that a specific group of TissueCo workers – the maintenance staff – has to make a critical contribution for the system to function effectively results in data work facilitation, which attempts to ensure that the maintenance workers perform the required evaluations and engage with the system. Requiring work related to the evaluation and input of data and domain and organizational knowledge from within the implementing organization also results in interactions between ManufacturingAI as the system developer and TissueCo as the implementing company that are characterized by a dependency on the maintenance teams' interactions with the system.

4.2 Data work in administrative services – conflicts about mobilizing acceptance

The monitoring and evaluation of administrative processes and productivity-increasing changes are routine in many companies. Such evaluations are traditionally performed by internal or external consultancy teams who observe workers' desktop work process and conduct interviews and workshops, followed by the analysis of this qualitative data by consultants, deriving suggestions for improvements from the results. In the last years, software companies have started to develop systems to perform such time and work-intensive process evaluations based on data, promising to save the implementing company time and costs (CS2_1).

The development company in our second case study, AdminAI, is a software company that is exploring this field. Using process mining techniques, the company collects data on their customers' employees' work processes, which they report in anonymized form (CS2_4). The data is produced by workers in their digital work processes, for example through mouse clicks and the use of software, e.g., during a tax filing process, the calculation of a quotation for a certain service, or the processing of an incoming order. The software company analyzes the data, identifies variations in process execution, defines which approach is the most efficient, suggests which parts of a process could be automated to save time, and evaluates how effectively software licenses are used. Based on the data analysis, they provide suggestions for process optimization. For implementation projects, responsibilities are divided between customer relations, responsible for the organizational and contractual implementation of customer projects and technical relations, responsible for the technical implementation and customization of AdminAI's software for the customer (CS2_4). From the implementing companies' side, AdminAI's 'champion' within the management, team leads, and most importantly, the employees in administration departments are involved, as the expected users of the software.

AdminAI's software is not used long-term within organizations but in optimization projects that run for a couple of weeks. Once the data collection is completed, the software is uninstalled, and after the data analysis is finished and suggestions for process optimization are conveyed, AdminAI is no longer involved. To achieve the promised results, AdminAI needs to have access to data that is produced by workers in their everyday work routines. It is therefore dependent on their consent to participate in the data production process.

The data needed to implement AdminAI's software – digital process data produced by workers in administrative departments of the implementing company – shapes the data pipeline, the constellation of involved actors, and the required data work in the implementation process. The project implementation starts with the installation of AdminAI's software on the computers of all workers taking part in the process evaluation project. In the initial phase of a project, AdminAI runs the software for two weeks without including this data in the analysis. According to AdminAI, data during the first two weeks shows a certain bias as workers are more aware of the data collection in the background. In their work, they tend to adhere more to instructions than usual (CS2_4). This data is used to determine what kind of software workers use in their work. Together with the team lead of the respective project team, AdminAI then determines which of those programs should be included in the data collection. During a second phase of data collection (about one week), AdminAI develops an anonymization format that claims to make sure that no company internal information is included in the data production and that no personal information about workers can be inferred from the data (CS2_4). In a third phase, the actual data collection starts and runs for several weeks. According to AdminAI, the data collection does not produce additional work for workers, as they produce the data by performing their daily work (CS2_4). Subsequently, AdminAI analyzes the data, using unsupervised machine learning techniques to identify processes. The model output evaluation and deriving of suggestions for improvements is done by AdminAI's team members manually: 'The current situation is that a human has to look at it, we need human intelligence, [...]. We have this AI that has found the process here. Now the human has to look at it and say, okay, what conclusions can I draw from this?' (CS2_4). In the final phase of the project, the results of the analysis and suggestions for changes are presented to the participants and the management of the implementing organization during a workshop. AdminAI is not involved in the following process of implementing changes.

In this case, the customization process in the implementation of the system is less about the data production and its control but more about the development of meaningful optimization recommendations. A technology-centered description of the data pipeline misses important aspects of work that are crucial for the implementation process. This is work attempted to facilitate workers' participation in the project and work that is necessary to develop criteria to adapt the system's functionality and scope for organization-specific purposes (see also figure 2).

For the implementation of AdminAI's software, attempts to facilitate workers' participation are made primarily before the start of the project. In AdminAI's perspective, finding workers that agree to take part in such a project is the greatest challenge. In this context, one of the founders of AdminAI states, 'as crazy as it sounds, workers are the stakeholder number one' (CS2_3). This statement echoes the dependency on the digital data produced by workers that has crucial value for them and demonstrates the dependency on workers participation in the data production.

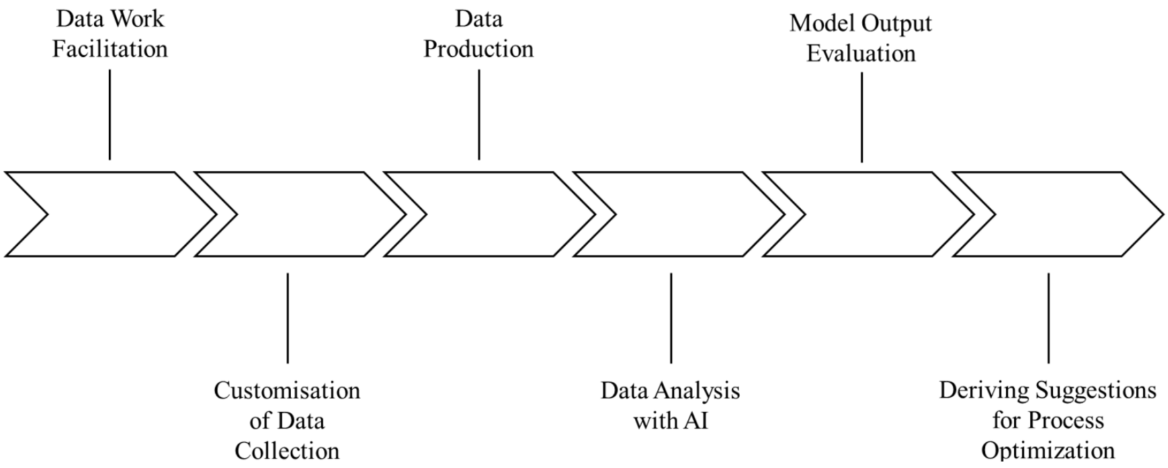
However, the participation of workers constitutes a conflictual process. One explanation for workers' reluctance to participate in such projects is the nature of the data. What is collected is their personal work process data which creates the possibility for surveillance. The data creates transparency about their work routines which might not always conform to companies' guidelines and allows benchmarking workers' efficiency against each other (CS2_5). While AdminAI emphasizes that in practice they disable this possibility of analyzing workers' specific data through anonymizing and aggregating data, this constitutes a serious concern by workers in the initial planning phase of the project and a reason for refusing to participate in a project (CS2_3).

Additionally, there is a concern that AI-based process optimization will intensify the workload or eliminate the jobs of the employee groups that participate (CS2_5). Much of the data work facilitation in this case is performed by the implementation company's management, particularly by AdminAI's 'champion' within the implementing organization. This includes obtaining the required approval from data protection officers and, if it exists, the works council. However, this only creates the basis for workers' engagement with the project. In addition, company management has to invest significant time in gaining workers' trust. According to AdminAI's founders, this can be a lengthy process and projects can be terminated in this stage or even before workers are approached because the higher-level management believes the project to bear too big a risk of losing workers: 'the management sometimes blocks it, [..] [and] says, 'yes, we think it's really good, but we're afraid that the employees will quit' because the labor shortage is so extreme' (CS2_4).

Once the management of the potential implementing company has found workers to participate in the project, AdminAI gets involved in attempts of data work facilitation. They conduct an informational workshop in which they explain the software, the scope of the data collection, and the project's timeline. Effectively motivating workers to participate in the project and integrating their conditions for participation is crucial for AdminAI and the management of the implementing company, as the software for data collection can be switched off by users at any point in time during the data collection period (CS2_4).

In addition, the customization of the process recommendations of the AI system leads to a de-centering of the implementing organization, as they are dependent on the data analysis by the development company. This results in the middle management gaining decision making power about process design and optimization goals. As described above, the evaluation of the model output data and the development of possible recommendations still relies on significant human input. However, this input cannot come only from the data science team by AdminAI but needs insight from the organization-specific demands of the implementing company. Formal process and domain expertise from within the implementing company remains crucial to develop meaningful recommendations. This is performed mainly by the middle management that defines optimization goals in exchange. The employees whose data is used for these analyses might lose influence once the system is customized.

Figure 2: Data pipeline in AI implementation: Administrative services



The case study shows that work by both AI data work institutions – AdminAI and the implementing company – and by different actors is required along the data pipeline to implement AdminAI's process evaluation software. This includes the management of the implementing company and AdminAI defining the conditions and scope of data collection in an iterative process, workers of the implementing company producing the data, and AdminAI evaluating the model outcome and developing suggestions for changes in the process. As administration workers are the data producers, extensive time and resources are invested to motivate them to take part in the project. This is work mainly performed by the implementing companies' management, and AdminAI has to rely on their ability to win workers' trust.

The availability of digital data in this case study shifts the main focus of data work towards the developing company while data work facilitation on the level of the employees is mainly about the manufacturing of acceptance. Once legal conditions and acceptance are secured, the implementation of the AI system and relies less on the direct involvement of the employees. Finally, there is a tendency of de-centering evaluation practices towards the developing company, which might result in the implementing company losing oversight over sensitive data. The integration of the management by determining process goals remains, however, an important step of the customization of the system.

5 Discussion

In both case studies, the AI systems are implemented with the aim to monitor, analyze, and optimize processes. The systems differ, however, in their availability and form of data and in the digital interfaces through which interactions with the systems are mediated. The AI system in the manufacturing case study collects data from the manufacturing process via sensors, from manufacturing equipment involved in the production process. Here, the translation of machine data into digital data constitutes a key process. The AI system in the administrative service case study, on the other hand, collects data from digital administrative processes. Data is scraped by the system's access to desktop processes and is therefore already available in a digital format. The case studies show that data work plays a crucial role in AI implementation processes in organizations. The responsibilities for this new data-related work are divided between different actors in the two key AI data work institutions (Muldoon et al., 2024), the developing and implementing organizations, while the distribution of data work tasks differs across the two cases.

For both case studies, data work facilitation is performed already before the start of the project to motivate workers to participate in the project. In the administrative services case, it only takes place at this phase, since workers automatically produce data after they have agreed to participate in their regular work processes. In the manufacturing case, data work facilitation is also crucial in the continuous operation and improvement of the software, requiring the CSM and TSM to stay involved and maintenance workers to engage with the software, evaluating and correcting its outputs. The importance of workers' participation in data work for both use cases indicates a certain dependency of the developing and implementing company on these groups of workers. If data work is evaluated as contributing to a possible rationalization of the tasks performed by the involved workers, this can lead to refusal to participate or to a reluctance to work with the systems. Yet, in these two AI implementation cases, workers seem to not have started leveraging their newly gained structural power grounded in their crucial data work by negotiating their potential involvement in and the conditions of AI system uses in their workplace.

The two cases demonstrate how implementing companies make a significant contribution to the data work necessary to implement ManufacturingAI's and AdminAI's software. The division of labor along the data pipeline in AI implementation is more complex and interactive in the manufacturing case study, with maintenance workers involved in the model output evaluation of ManufacturingAI's software, resulting in an additional workload. In the administrative service case study, the data pipeline in AI implementation and related data work tasks are more sequential. Workers in administration departments are involved only as data producers, which they, if agreeing to work with the software, become automatically by performing their daily desktop-based work, presumably without causing additional workloads. AI model outcome evaluation and sensemaking of data is performed by AdminAI together with the management of the implementing company. How workers of the implementing organizations are involved in data work furthermore influences attempts of facilitating such data work.

Consequently, in contrast to outsourced and globally distributed data work in earlier phases along AI supply chains, the contexts, conditions, and power dynamics that structure data work in our cases differ. While data work in AI production is often characterized by precarious working conditions with data workers predominantly located in the Majority World, workers confronted with new data-related tasks in AI implementation in our cases in Germany are better protected by industrial relations institutions. Their expertise is more recognized and correspondingly remunerated, and they are members of established organizations that often have strong works councils. Through the intended implementation and subsequent integration of AI systems into organizational processes and work practices, they are now confronted with new, additional requirements and tasks related to data production, evaluation, and curation, along with their other responsibilities, which can further add to workers' existing workloads.

Our findings are related to three levels of analysis. On the level of inter-firm relations between the developing and implementing companies, we see an organization-centric process of data facilitation in the first case, as the data production and evaluation depends heavily on insights from the shopfloor. In the second case, if gaining acceptance of the workers, the processes of data production and analysis take place together with middle management, sidestepping further workers' involvement. Secondly, the data work tasks in the first case are more visible in both data production and analysis. This is due to non-routinized tasks that emerge in tending to the AI model in- and outputs and stands in contrast to the second case, where data production is largely automated, while it can be manually turned on and off by workers. Finally, this leads to different forms of responses by the workers within the implementation process. For the management, the mobilization of workers' participation is important in both cases. In the second case, this mobilization focuses on workers' acceptance to participate in the data production. Acceptance is also important in the first case, although here the mobilization of workers to add data and evaluate AI model outputs constitutes an additional element.

6 Conclusion

Our empirically grounded analysis of AI implementation processes in manufacturing and administrative services investigates different forms of data work and work that aims to facilitate such data work that arise in the implementation of AI-based systems in companies. Thereby, our perspective on AI implementation processes extends present understandings of human labor along the AI data pipeline (Muldoon et al., 2024) by centering data-related tasks along what we describe as a corresponding *data pipeline in AI implementation*.

In the two contrasting case studies, workers of the developing and the implementing companies are confronted with new data work tasks in their interactions with the interfaces of the respective AI system. These tasks are increasingly coordinated and conducted across established organizational boundaries of the two companies, with implementing companies actively participating in the production, curation, and evaluation of the data the developing company requires to run its software. Our analysis shows that emerging dependencies arise from such work, as certain data work tasks in AI implementation require organization-specific knowledge and skills and can thus only be performed by a selected group of workers in the implementing organization. For the developing companies and the implementing companies' management, these dependencies that underlie and structure AI implementation processes in organizations bring rise to a new set of tasks which we describe as *data work facilitation*, work with the aim to produce, curate, and evaluate data within organizations. For workers who are indispensable in AI implementation processes, this dependency on their participation in data work can open potential new spaces for negotiation and a new form of workers' structural power, which we describe as *data work bargaining power*.

Our research adds to work that has studied the criticality and conditions of data work along the development and use of AI systems. In contrast to data work in the production of AI systems, data work in AI implementation phases differs in several aspects. This form of data work requires organization- and domain-specific knowledge and expertise, hence it cannot be easily outsourced or distributed to a global workforce of data workers. In addition, it establishes a changing division of labor between the AI implementing and developing companies. Thus, the coordination of data work increasingly transgresses organizational boundaries, but the required data work expertise is closely linked and embedded in knowledge of organization-specific processes and practices. Furthermore, data work often adds to existing workloads of employees, and it has to be embedded into existing forms of work organization and labor processes.

Interestingly, even though some workers in our two empirical cases refuse to work with the software (administrative services case) or do not engage with it extensively enough for the developing company to be able to improve its AI models (manufacturing case), workers have not yet acknowledged and realized their potential *data work bargaining power* as such, which can also be linked to the fact that these workers, even if better protected than outsourced data workers, are also dependent on their employee status. Due to the constitutive dependency on data work that requires organization- and role specific knowledge of workers, workers and works councils, on a company level, and trade unions, on a sectoral level, could leverage such power by negotiating the specific conditions under which workers are willing to engage in data production, curation, and validation and therefore participate in the training and improvement of AI systems. Topics for negotiation could include the scope of secondary use of data produced in implementing contexts, workers' financial share in productivity gains through their data work, and co-determination of technology implementation beyond legal stipulation. The precondition for leveraging such data work bargaining power is that workers and their representatives recognize data work in implementation as a powerful new area of work.

Investigating AI implementation in companies comes with certain limitations, as the analytical and empirical focus of this study lies on particular 'moments in time' in ongoing digital transformation processes in organizations. As this work focuses on the implementation process of AI-based systems, it cannot draw conclusions regarding the long-term use and effects of these systems. From our perspective, this phase along AI development and use is, however, especially significant, as it provides moments of bargaining that workers might be able to use to influence AI uses in their interest. Nonetheless, even though we identify and outline the dependency on new data work tasks, the further development and use of such systems can lead to forms of automation and job substitution. It is therefore important to develop adequate strategies to shape the digital transformation of organizations. Research across different organizational contexts of such data work and corresponding potentials and conditions of new data work bargaining power can add to a more nuanced understanding of these power dynamics. Similarly, the particularities and respective effects of AI systems mediating such data work, especially with regard to workers' (dis)empowerment, changing roles of expertise, and potentials for co-determination at the workplace are further aspects to take into consideration in research on the introduction of AI systems to organizations.

\\ References

Ahmad, S. (2023). Who Moderates My Social Media? Locating Indian workers in the global content moderation practices. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds), *Challenges and Perspectives of Hate Speech Research* (pp. 111–125).

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science, 33*(1), 19–37. https://doi.org/10.1287/orsc.2021.1552

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and Value. In S. Nambisan, K. Lyytinen, & Y. Yoo (Eds), *Handbook of Digital Innovation* (pp. 162–178). Edward Elgar Publishing. https://doi.org/10.4337/9781788119986.00022

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization, 30*(2), 1–12. https://doi.org/10.1016/j.infoandorg.2019.100286

Bossen, C., Chen, Y., & Pine, K. H. (2019). The emergence of new data work occupations in healthcare: The case of medical scribes. *International Journal of Medical Informatics, 123*, 76–83. https://doi.org/10.1016/j.ijmedinf.2019.01.001

Brinkmann, S. (2020). Unstructured and semi-structured interviewing. In P. Leavy (Ed.), *The Oxford Handbook of Qualitative Research* (2nd edn, pp. 277–299). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190847388.001.0001

Butollo, F., & Schneidemesser, L. (2022). Who runs the show in digitalized manufacturing? Data, digital platforms and the restructuring of global value chains. *Global Networks, 22*(4), 595–614. https://doi.org/10.1111/glob.12366

Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1186–1197. https://doi.org/10.1145/3593013.3594073

Donatz-Fest, I. (2024). The 'doings' behind data: An ethnography of police data construction. *Big Data & Society, 11*(3), 1–12. https://doi.org/10.1177/20539517241270695

Ferrari, F., & Graham, M. (2021). Fissures in algorithmic power: Platforms, code, and contestation. *Cultural Studies, 35*(4–5), 814–832. https://doi.org/10.1080/09502386.2021.1895250

Fiore-Gartland, B., & Neff, G. (2015). Communication, Mediation, and the Expectations of Data: Data Valences Across Health and Wellness Communities. *International Journal of Communication, 9*, Article 0.

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.

Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–620. https://doi.org/10.1145/2470654.2470742

Jarke, J., & Büchner, S. (2024). Who cares about data? Data care arrangements in everyday organisational practice. *Information, Communication & Society, 27*(4), 702 – 718. https://doi.org/10.1080/1369118X.2024.2320917

Jarke, J., Zakharova, I., & Breiter, A. (2022). Organisational Data Work and Its Horizons of Sense: On the Importance of Considering the Temporalities and Topologies of Data Movement When Researching Digital Transformation(s). *Historical Social Research, 47*(3), 142 – 170. https://doi.org/10.12759/hsr.47.2022.29

Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society, 8*(2), 1 – 14. https://doi.org/10.1177/20539517211020332

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals, 14*(1), 366 – 410. https://doi.org/10.5465/annals.2018.0174

Kostis, A., Sundberg, L., & Holmström, J. (2024). Data work as an organizing principle in developing AI. In I. Constantiou, M. P. Joshi, & M. Stelmaszak (Eds), *Research Handbook on Artificial Intelligence and Decision Making in Organizations* (pp. 38 – 57). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781803926216/book-part-9781803926216-10.xml

Krzywdzinski, M., Schneiß, D., & Sperling, A. (2024). Between control and participation: The politics of algorithmic management. *New Technology, Work and Employment*, 1 – 21. https://doi.org/10.1111/ntwe.12293

Mackenzie, A. (2017). *Machine Learners: Archaeology of a Data Practice*. MIT Press.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*.

Miceli, M., & Posada, J. (2022). The Data-Production Dispositif. *Proceedings of the ACM on Human-Computer Interaction, 6*, 460:1-460:37. https://doi.org/10.1145/3555561

Miceli, M., Posada, J., & Yang, T. (2022). Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? *Proceedings of the ACM on Human-Computer Interaction, 6*, 1 – 37. https://doi.org/10.1145/3492853

Muldoon, J., Cant, C., Wu, B., & Graham, M. (2024). A typology of artificial intelligence data work. *Big Data & Society, 11*(1), 1 – 13. https://doi.org/10.1177/20539517241232632

Newlands, G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data & Society, 8*(1), 1–14. https://doi.org/10.1177/20539517211016026

Nies, S. (2021). Eine Frage der Kontrolle? Betriebliche Strategien der Digitalisierung und die Autonomie von Beschäftigen in der Produktion. *Berliner Journal für Soziologie, 31*(3 – 4), 475 – 504. https://doi.org/10.1007/s11609-021-00452-8

Nitzberg, M., & Zysman, J. (2022). Algorithms, data, and platforms: The diverse challenges of governing AI. *Journal of European Public Policy, 29*(11), 1753 – 1778. https://doi.org/10.1080/13501763.2022.2096668

Pine, K., Bossen, C., Holten Møller, N., Miceli, M., Lu, A. J., Chen, Y., Horgan, L., Su, Z., Neff, G., & Mazmanian, M. (2022). Investigating Data Work Across Domains: New Perspectives on the Work of Creating Data. *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3491101.3503724

Pink, S., Berg, M., Lupton, D., & Ruckenstein, M. (Eds). (2022). *Everyday Automation: Experiencing and Anticipating Emerging Technologies*. Routledge.

Posada, J., Newlands, G., & Miceli, M. (2023). Labor, Automation, and Human–Machine Communication. In A. L. Guzman, R. McEwen, & S. Jones (Eds), *The SAGE Handbook of Human–Machine Communication* (pp. 384–391). SAGE. https://doi.org/10.4135/9781529782783.n48

Priya, A. (2021). Case Study Methodology of Qualitative Research: Key Attributes and Navigating the Conundrums in Its Application. *Sociological Bulletin, 70*(1), 94–110. https://doi.org/10.1177/0038022920970318

Roberts, S. T. (2021). *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445518

Schmalz, S., & Dörre, K. (2014). Der Machtressourcenansatz: Ein Instrument zur Analyse gewerkschaftlichen Handlungsvermoegens (The power resource approach: An instrument to analyze trade union action capabilities). *Industrielle Beziehungen - Zeitschrift Für Arbeit, Organisation Und Management – The German Journal of Industrial Relations, 21*(3), 217–237.

Silver, B. J. (2003). *Forces of Labor: Workers' Movements and Globalization Since 1870*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615702

Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society, 7*(1), 1–12. https://doi.org/10.1177/2053951720919776

Vidal, M. (2022). *Management Divided: Contradictions of Labor Management*. Oxford University Press.

Vilasís-Pamos, J., Pires, F., Grohmann, R., & Fernandes Araujo, W. (2024). Social media and platform work: Stories, practices, and workers' organisation. Convergence: *The International Journal of Research into New Media Technologies, 30*(1), 410–427. https://doi.org/10.1177/13548565241227391

\\ Imprint

Lea Schneidemesser, Daniel Schneiß, Jana Heim
Dynamics of Data Work in AI Implementation Processes

Weizenbaum Discussion Paper #50
September 2025

ISSN 2748-5587 \ DOI 10.34669/WI.DP/50

Weizenbaum-Institut e. V.
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de

COORDINATION: Dr. Moritz Buchner \ **TYPESETTING:** Luisa Le van, M.A., Atelier Hurra

weizenbaum institut

With funding from the:

Federal Ministry of Research, Technology and Space

BERLIN

Senate Department for Higher Education and Research, Health and Long-Term Care