Pixels and Weights

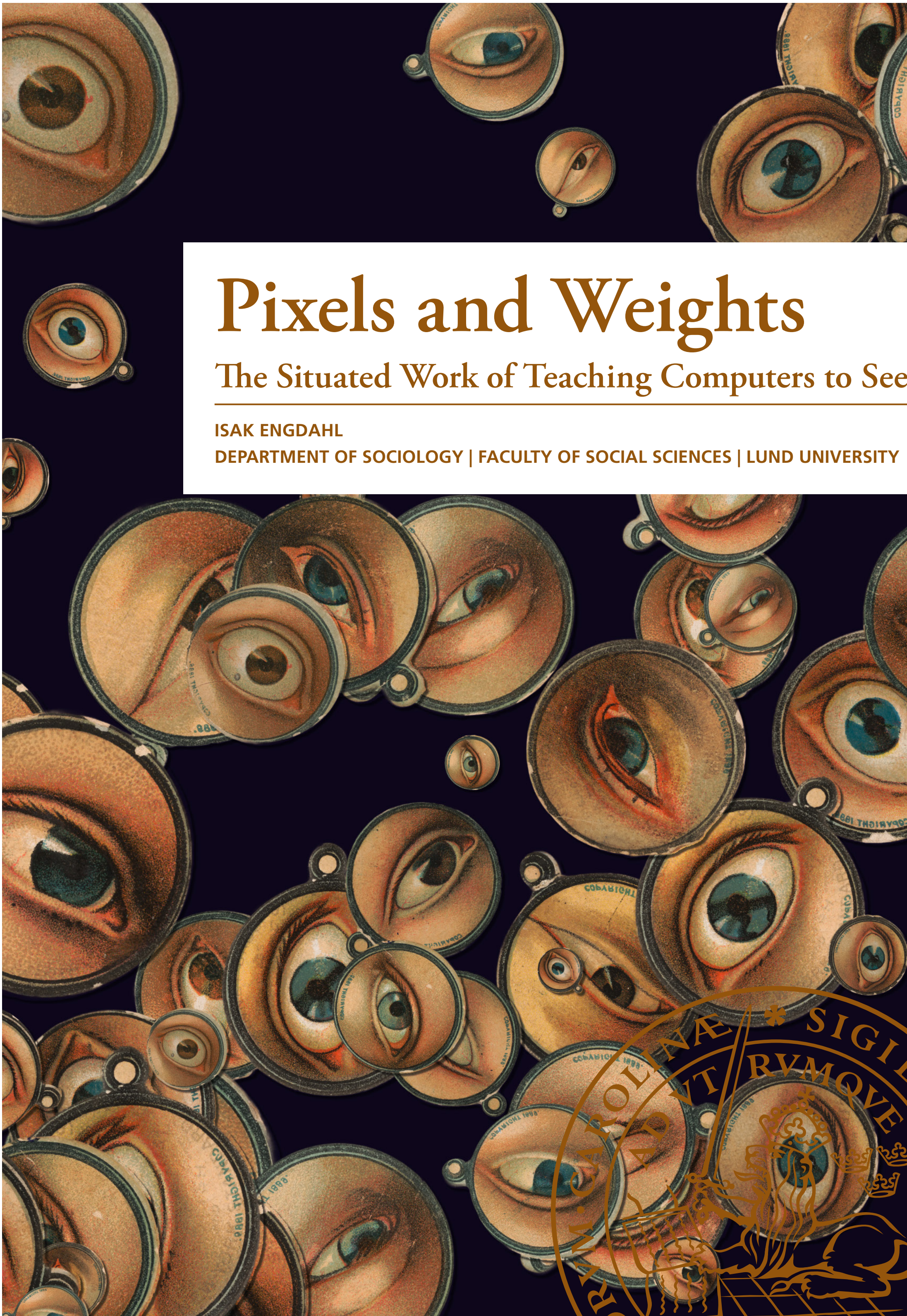
The Situated Work of Teaching Computers to See

Engdahl, Isak

2025

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Pixels and Weights

The Situated Work of Teaching Computers to See

ISAK ENGDAHL
DEPARTMENT OF SOCIOLOGY | FACULTY OF SOCIAL SCIENCES | LUND UNIVERSITY

Pixels and Weights traces the situated work of research and development in a core area of artificial intelligence: computer vision. Drawing on an interactionist perspective, the analysis draws attention to how computer vision scientists produce knowledge and navigate technical production. Based on a multi-sited ethnography of research labs, it explores model development, benchmark dataset construction, and model implementation. The study develops concepts that address how this work is carried out within computational infrastructures and practical arrangements. Its contribution is a grounded account of how research and development in artificial intelligence is performed in practice.

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

Faculty of Social Sciences
Department of Sociology

Lund Dissertations in Sociology 143
ISBN 978-91-8104-691-5
ISSN 1102–4712

LUND UNIVERSITY

PIXELS AND WEIGHTS

Pixels and Weights

The Situated Work of Teaching Computers to See

Isak Engdahl

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Social Sciences at Lund University, to be publicly defended on the 28th of November 2025 at 13.15 in LUX Aula Nedre, Helgonavägen 3, Lund

Faculty opponent: Associate Professor Malte Ziewitz, Cornell University

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation **Date of issue:** 2025-11-28

Author(s): Isak Engdahl

Title and subtitle: Pixels and Weights: The Situated Work of Teaching Computers to See

Abstract: The uses of artificial intelligence are increasing across many areas of contemporary society, including computer vision systems involving machine learning and artificial neural networks. These systems rearrange action possibilities and are often described as "black boxes" that challenge conventional modes of scientific understanding and engineering control. It is important to understand the activities through which such systems are constructed. The aim of this dissertation is to analyze the situated work involved in constructing computer vision components and systems: how research and development take shape in practice. Drawing on ethnographic fieldwork and interviews in computer vision research labs, the dissertation analyzes how computer vision scientists make computers capable of 'seeing'. Adopting a pragmatist–interactionist perspective, the dissertation treats science and engineering as hybrid forms of collective action where scientists interact with computational objects and research apparatuses. The analysis follows three lines of their work: model development, the construction of evaluative benchmark datasets, and the implementation of models in new contexts. The first study argues that daily laboratory activity is organized through articulation work and pipeline welding, where computational components are integrated into operational models. These models become epistemically opaque yet meaningful through shared interpretive work in relationally sustained awareness contexts. The second study analyzes the creation of a benchmark dataset, showing how standards and protocols depend on less-visible alignment work to hold cooperation together. The third study investigates trajectories in which scientists adapt pre-trained models to local contexts, and develops the idea of a social license to capture how researchers coordinate expertise and effort when reconstructing external models for local use. This social license helps actors manage trajectories and navigate the contingencies that arise as models are adapted to new settings. The dissertation advances an interactionist and processual understanding of computational knowledge production, foregrounding the collective work that makes computer vision systems function beyond narratives of exceptionalism, crisis, or autonomy.

Key words: laboratory ethnography, situated work, computer vision, machine learning, model development, benchmark dataset, model implementation, artificial intelligence

Language English

ISSN and key title: 1102–4712, Lund Dissertations in Sociology 143

ISBN: 978-91-8104-691-5 (print), 978-91-8104-692-2 (electronic)

Number of pages: 176

Signature Date 2025-10-12

Pixels and Weights

The Situated Work of Teaching Computers to See

Isak Engdahl

LUND UNIVERSITY

Published by:

Department of Sociology
Faculty of Social Sciences
Lund University

ISBN 978-91-8104-691-5 (print)
ISBN 978-91-8104-692-2 (electronic)

Lund Dissertations in Sociology 143
ISSN 1102–4712

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2025

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Table of Contents

List of Tables

Acknowledgments

I want to thank all the researchers who participated in this study. This dissertation was made possible by your willingness to share your time and experiences, allowing me to follow your work. It has been a privilege to learn from another knowledge community, and I have tried my best to understand the complexity of what you do.

I want to thank my supervisors Alison Gerber and David Wästerfors. I am so grateful for your guidance, patience, and all the conversations that have shaped this work. This dissertation carries many traces of what I learned with you, and so do I. You are truly brilliant, and kind.

To my project colleagues in “Show & Tell”[1]—Alison Gerber, Tobias Olofsson, Alexandra Middleton, Lila Lee-Morrison, and Kenneth Ravn—working with you has been both inspiring and grounding. I have learned so much from us being together, and I already cherish the memories of the other experiences we shared along the way. I am very thankful for your companionship and mentorship throughout our shared explorations.

To my colleagues and friends at the Department of Sociology, the research environment Social Studies of Science and Technology, and the Digital STS Hub, thank you for creating an inspiring and generous academic home. I am grateful for all of the occasions where sociological discussions have often, but not always, been central. To the faculty members who contributed to my sociological development over the years, particularly Chares Demetriou and Bo Isenberg, thank you.

Everyone that has been part of the PhD collective—getting to know you on this journey has been a privilege. This includes Freja Morris, Colm Flaherty, Daniel Karlsson, David Bazan Royuela, Linn Alenius Wallin, Jens Westerskov Andersen, Tove Gustavsson, Kristoffer Berglund, Anton Petersson, Lyudmyla Corney, Anika Binte Habib, Yvonne Jila, Aube Tollu, Anna Kallos, Naja Yndal-

[1] This project was supported by the H2020 European Research Council, grant number 949050.

Olsen, Jesse Wiles, Anna Wiemer, Helena Andersen, among others. I wish to thank Staffan Edling in particular for his extraordinary and warm friendship. I am also grateful to Charlotte Högberg, Alicja Ostrowska, and Camilo Castillo-Estupiñan for being such generous fellow STS travelers.

I am grateful to the readers, who provided valuable feedback on this dissertation: Kristian Bondo Hansen, Christofer Edling, Bertil Rolandsson, and Lisa Flower. I also thank the opponent, Malte Ziewitz, for traveling to Lund in person and the committee members, Francis Lee, Nanna Bonde Thylstrup, and Lisa Flower once more, for your engagement with this study.

I am also grateful to the researchers at other places who took time to organize events, invite me into conversations at conferences, and engage with my work. Francis Lee, thank you so much for your welcoming and excellent mentorship in all of the different settings over the years. And Jacob Gaboury, thank you for hosting me at UC Berkeley. Thanks also to Anders Hylmö, Steve Woolgar, Florian Jaton, Angèle Christin, Marion Fourcade, Nick Seaver, David Ribes, Antony Puddephatt, and Gabriele de Seta, for your input in various stages of this work. Thank you, Paul Edwards, for sharing your idiosyncratic bibliography of IT-related items through around 2007, in chronological order. Thanks to Göran Ahrne for scanning and sending me a copy of 'Teknikkritik. En översikt av en kritiktradition*'* from 1979 and Robin Mansell for doing the same with a book from 1994 so I could read Cooper and Woolgar's, 'Software quality as community performance'.

And to my friends outside academia: thank you for keeping me grounded, encouraging me, and reminding me that there is joy to be found at other places too.

Finally, to Saga, and to my family. Thank you for your love and unwavering support.

Isak
Malmö, 2025, 12 October

List of Papers

Paper I

Engdahl, I. (2025) 'Learning by Doing: An Ethnographic Account of Computer Vision Laboratory Work'. Unpublished manuscript.

Paper II

Engdahl, I. (2024) 'Agreements 'in the Wild': Standards and Alignment in Machine Learning Benchmark Dataset Construction', *Big Data & Society*, 11(2). doi:10.1177/20539517241242457.

Paper III

Engdahl, I. (2024) 'Wuthering Weights—Localisation Trajectories of Machine Learning Models for Local Ends', in Fors, V., Berg, M. and Brodersen, M. (eds.) *The De Gruyter Handbook of Automated Futures: Imaginaries, Interactions and Impact*. Berlin/Boston: De Gruyter, pp. 321–336. doi:10.1515/9783110792256-020.

1 Introduction

Artificial intelligence systems based on artificial neural networks are now integrated into many areas in society, setting new baselines for what computational systems can do. The development processes shape how these systems operate and how they influence action. This study explores the worlds of the scientists who are developing these systems, reshaping how we carry out actions and sometimes enabling new ones.

Artificial intelligence researchers organize their field in specializations like vision or language, and this study specifically concentrates on computer vision. Recent advances in the pursuit of computerizing vision have engendered new action possibilities and rearranged what seeing involves, especially considering that the capability for vision has historically been attributed particularly to living organisms. The prospect of transforming vision into a computational capability is a remarkable technological feat. Beyond achievements like processing printed text or automatically scanning barcodes, computer vision now includes systems that recognize objects in images from cameras or on computer screens and other devices, support self-driving vehicles, detect movements and emotions in videos, and generate images and realistic three-dimensional reconstructions of scenes resembling human visual experiences.

These contemporary advancements have been made possible by the success of machine learning, especially deep learning (LeCun, Bengio and Hinton, 2015; Torralba, Isola and Freeman, 2024). Recipients of the Nobel prize (2024) and Turing Awards (2018, 2024), research teams are celebrated, and computational objects are widely adopted in industries embracing the "innovation imperative" (cf. Pfotenhauer, Juhl and Aarden, 2019). Artificial intelligence has entered a stage where consequential computational approaches are now in use and actively being extended, which commentators sometimes call *the scaling era* (e.g., Patel, 2025). Patel describes the scaling era (2019–2025) as the phase during which progress in artificial intelligence originate from enlarging models, data, and computing power, guided by the assumption that scale equates to better technical performance.

Meanwhile, a central factor separates artificial intelligence from conventional computing technology (cf. Narayanan and Kapoor, 2025). Advanced forms of artificial intelligence encapsulated in concepts such as ‘artificial general intelligence’ and ‘superintelligence’ are embedded in what Jasanoff and Kim (2009:120) call ‘sociotechnical imaginaries’ or collective visions of the role of technology in society. These often draw on the speculative figure of superhuman intelligence, packaged within an ‘AI-first’ society where computing becomes a central constituent of the social order. This serves to show the centrality of technology in social change, and how this relationship is imagined. It is a hotspot of fascination, anxiety, and investment. Relatedly, Marres et al. (2024) call artificial intelligence a “super-controversy” because it spans more than disputes over a single technical claim, typical of scientific controversies (Engelhardt and Caplan, 1987). The term describes a dense knot in which epistemic questions and technical transgressions become evidence of social problems, and technical, ethical, and political arguments merge into a continuous stream of debate with high societal stakes.

While the build-out of new computational arrangements relies on many different actors (Becker, 1982; Kling, 1991; Gray and Suri, 2019), they are fundamentally constructed in research labs. Research labs are sites where computational affordances are constructed, which is why they are central to social change. These locations concentrate the actors and infrastructures that produce and circulate knowledge that computational systems mobilize. This dissertation takes it as worthwhile to analyze the research practices that such imaginaries and debates relate to. Research and development structures what systems become and what they enable or constrain.

This study analyzes how artificial intelligence scientists conduct research and development in computer vision within such research labs, focusing on the processes that structure their work. It takes interest in the situated work involved in constructing components and systems, focusing on how research and development take shape in practice. Drawing inspiration from the tradition of laboratory ethnography (Latour and Woolgar, 1979/1986; Knorr Cetina, 1981; Lynch, 1985; Traweek, 1988; Fujimura, 1996; Hess, 2001; Sismondo, 2010; Lynch, 2024) and following Lynch’s (Ziewitz and Lynch, 2018:382) invitation to “look at what scientific practice involves”, the study focuses on the interactions through which computational knowledge, components, and systems are produced. By engaging ethnographically with such activities, it seeks to advance a sociological understanding of how contemporary computational knowledge is enacted in practice and shows why such attention is necessary to avoid mystifying the systems that shape social life, and to learn about the worlds in which they originate.

I have embedded myself ethnographically in research labs, where researchers spend their days teaching computers to see in different capacities. As alluded to in the title of this study, researching how computing worlds operate involves paying attention to concerns central to the activities of computer vision researchers, such as the numerical data of pixel values and the learned parameters or weights distributed across layers of artificial neural networks (cf. Blumer, 1969; Latour, 1987). Computer vision is an interdisciplinary field rooted in computer science, and seeks to computerize perception, recognition, and reconstruction processes (Malik et al., 2016). It draws on mathematical frameworks, including geometric and probabilistic methods, statistical learning, and theories from cognitive science (Torralba, Isola and Freeman, 2024). A core assumption is that it is possible to observe a scene from different angles or moments, represent it in images, with each view offering more information that can be processed to understand the scene (ibid.). Pixels are the basic data of digital images, forming the numerical foundation for the sophisticated representations computer vision systems rely on when processing visual information (Dobson, 2023). As they perform their work, objects and instruments like these and many others inhabit a central position in how they deal with knowledge and stabilize systems. This often involves technical trade-offs and infrastructural decisions shaping the practical development of seeing computers, similar to what Star (1999) provocatively calls “boring things”. She uses the term deliberately: to outsiders, these things may seem dull, but they actively shape what people can do in specific settings, which is why Star and others were and continue to be drawn to studying them. Attending to these boring things in Star’s sense shows how such possibilities are shaped by the various work arrangements through which research and development occurs: arrangements that makes 'seeing' computers possible to claim with the authority of science and engineering (cf. Knorr Cetina, 1999:1).

As I analyze these arrangements, this dissertation mobilizes a pragmatist perspective of action with sociological roots as central to understanding science and engineering from the interdisciplinary field of science and technology studies or STS (Strauss, 1993; Clarke and Star, 2008). A central assumption is to consider computational knowledge production as a social activity and technical systems as social and material products. In adopting this perspective, technological and scientific projects become a site of action, which this dissertation brings into play at the edge of computational research and development. I renovate some of the sensibilities associated with this framework by referencing theoretical elaborations on the role of materiality in action and meaning generation (Puddephatt, 2005), so as to arrive at a stronger theory capable of addressing the complex activities of science and engineering, heavily reliant upon human–computer interaction. Centering the processual doings in social life, science and technology is understood as a series of actions performed by and between actors, necessitating infrastructure, maintenance, adjustment, negotiation, and contingency (Clarke and Star, 2008).

Based on this framework, I approach artificial intelligence and computational knowledge production as a process in which diverse ways of knowing converge from the perspective of action. Knowledge and systems gain stability and coherence through interaction (Clarke and Star, 2008). Instead of seeing computer vision work as one uniform process, I focus on what I call *situated work*, inspired by Strauss (1993): the local activities and interactions actors engage in to get their work done. It draws attention to the decisions, adjustments, and negotiations scientists make as they deal with ambiguous problems, unexpected challenges, and complex tools. It also shows the quieter work, such as the ways actors sustain collaboration within projects that stabilize computational objects. Understanding how collaborative progress is enabled is important precisely because it moves research and development work along.

Because the dissertation explores specialized activities shaped by specific forms of expertise, it is helpful to first lay out how I understand the premises of the research questions and why they are important.

The ethnographic fieldwork that forms the basis of this study includes a diverse range of computer vision research and development efforts. This includes work on core technical tasks such as three-dimensional reconstruction methods, human motion estimation, and detailed modelling of hand–object interactions from first-person video streams. These are foundational computer vision tasks that make operational systems possible. For instance, if a computer can reconstruct a three-dimensional scene from images, much like piecing together a room by looking at photographs taken from different angles, it can start to operate in environments that were not pre-mapped. Human motion estimation involves teaching the computer to follow how people move. If it can estimate how bodies move, it becomes possible to recognize actions and assist in physical tasks. And if it can understand how hands interact with objects, it can support fine-grained activities, such as guiding a robot in manipulating tools, remembering where one put their keys, or interpreting what a person is doing in a laboratory setting. The fieldwork also includes efforts often grouped under the label 'AI for science', where computer vision techniques are applied to problems such as monitoring bird behavior, measuring car emissions, or supporting radiotherapy planning. Moving between these projects provides an entry point for understanding how particular ways of seeing the world are computationally enacted.

I draw attention to three focus areas in the three studies that make up this dissertation. I focus on the work involved in what scientists call 'model development', 'benchmarking', and 'model implementation': central clusters of activity in making computers see. I refer to the performance of work involved in the foundational stage of model development, meaning the micro-level dynamics of the laboratory, exploring how scientists interact with layered computational infrastructures and each other to build and refine new model architectures, such as those used for three-dimensional reconstruction and generative models. While tuning models and adjusting parameters affect how accurately a system performs, a model architecture must first be developed.

The second focus area consists of a practice central to the field: the creation of benchmark datasets. The well-known Turing Test from the 1950s evaluated computers on their linguistic abilities and whether they could pass as human (e.g., Collins, 1990; Collins, 2021). Contemporary artificial intelligence research, however, relies on other kinds of testbeds. Benchmarks, sometimes called the Common Task Framework (Donoho, 2017), are the internal mechanism through which scientists evaluate new models, defining what constitutes a shared problem and how performance is measured (Raji et al. 2021). Benchmarking and evaluation metrics typically define the standards by which success is measured, and baseline implementations that provide reference performance levels. Benchmarking consists of shared tasks, metrics, and data that allow models to be compared with existing ones under similar experiment-like conditions. Datasets determine what a model can learn and what kinds of problems it can address. Benchmarks shape research priorities by specifying tasks and guiding performance standard evaluations, and their outcomes influence both how technical work is recognized and labs' reputations. Benchmarking may thus be understood as evaluating a model's performance on a constructed test.

This is particularly relevant in technical knowledge production (e.g., Vincenti, 1990) from the perspective of the sociology of testing, where assessing prototypes and performance establishes a technology's workability; what Pinch (1993) calls "prospective testing". Following Mulkay's (1979) view, testing "cuts into the heart of the content of a technology", and is the basis for how its performance is understood. Moreover, in STS, 'data' is understood as a product of inscription (Latour and Woolgar, 1979), where phenomena must be translated into a form that can be stored. Creating data involves representational work (Lynch and Woolgar, 1990; Coopmans et al., 2014), which refers to the processes through which particular aspects of the world are categorized, carved, structured, and encoded into representations. Datasets are products of representational work, encoding selective aspects of the world into forms that can be processed. My approach to benchmarking in computer vision builds on this perspective: benchmarks are sites where workability is realized, requiring careful scaffolding and ongoing maintenance. They are knowledge objects, central to technical production, constructed rather than given, through labor that becomes infrastructural. Benchmarks are thus an evaluative framework which is constructed through a collaborative, and often complex, organization of work.

The third focus area concentrates on what happens after a model has been developed and benchmarked; specifically, its implementation in new settings. In contemporary artificial intelligence, it is common practice to take existing 'pre-trained' or 'foundation' models and adapt them to address new, specific problems. These models circulate through public repositories or model zoos, where they can be accessed and repurposed by different groups. Research teams copy, retrain, and integrate these models into their own infrastructures in order to solve locally defined problems. The stabilization of scientific and technological products depends on their ability to circulate between sites and on the coordinated work required to transfer and align both understandings and material systems (Latour, 1988; Puddephatt, 2004). Understanding how such models travel across settings is key to understanding how contemporary science and technology become stabilized in practice.

Based on specific case studies of local settings, the analysis targets key areas of activity in the era of deep learning, analyzing the situated work involved in three central clusters of activity: making models (developing, training, and refining architectures like so-called transformers and neural radiance fields); making benchmark data (constructing datasets for model evaluation); and implementing models (adapting trained convolutional neural network models and making them work for specific ends). Local and often mundane processes, such as deciding which images are useful and debating what the relevant signals are, are central to how computer vision techniques acquire functionality in practice. This makes visible how computational systems are produced through social practices, even those that become attributed as autonomous. In other words, the technology thrives on human assumptions and values.

1.1 Research Aim

This dissertation aims to analyze the situated work involved in constructing computer vision components and systems, focusing on how research and development take shape in practice. It explores how computer vision products achieve stability by tracing the agency of research and development teams and the interactions that push their activities and systems forward. Rather than treating these technologies as autonomous or deterministic, the purpose is to add to our sociological understanding of the activities forming computational systems which permute society in a grounded fashion.

1.2 Research Questions

To achieve this aim, this dissertation presents three research questions answered in three case studies, focusing on model development, benchmark dataset construction, and model implementation. It analyzes specific instances of these activities, which offer a partial but grounded and generative perspective of how computer vision systems gain stability in practice. The research questions are accompanied by an analytical interest in how actors make sense of and enact their work. This dimension is developed over the course of answering the following research questions:

1. How do computer vision researchers perform their activities when developing models within the layered infrastructure of the lab?
2. How do computer vision researchers perform their work when creating benchmark datasets that serve as evaluation tools?
3. How do computer vision researchers perform their activities when implementing models in new settings?

The first question deals with how laboratory work during model development is performed. This research question focuses on how actors interact with their computational and interpersonal workspace. The research question draws attention to the ways model development is accomplished through performances of human and human–computer interaction. It focuses on the micro-level dynamics of laboratory work, including how scientists manage their activities, interact with tools and infrastructures, and address challenges in model development.

The first research question is answered in a case study, investigating the activities within a laboratory primarily focused on model development. It assumes a research landscape where several artificial neural network architectures are consolidated in the field. The lab members' work involves improving existing model architectures, specifically, for so-called three-dimensional reconstruction methods and transformer architectures, the latter being a common 'backbone' through which generative models are built.

The second research question moves to another central part of computer vision, benchmarks, a central device through which computer vision scientists evaluate new models. Prior to any contestation of quantified technical performance by benchmarking is possible, practical scaffolding of it must occur, which leads me to ask the fundamental question of how this work is organized.

The second research question is explored in a case study of a consortium, which is a temporary organization of research labs working together toward a shared goal. In this project, the labs collaborate to create a new benchmark dataset, along with metrics and computer vision tasks. The dataset is created 'in the wild', an insider term meaning it captures naturally occurring conditions. It consists of videos of daily activities from a first-person perspective, with participants wearing cameras on their heads in their homes and similar settings. The study analyzes how the creation of a multifaceted object like a benchmark dataset is accomplished in practice, focusing particularly on how actors involved in this large and complex organization interact and seek alignment throughout their work.

The third research question starts from the idea that scientific and technical products become stable when they circulate and can be used in different settings. A common case in computer vision is model adaptation: researchers often take an existing model and adjust it so it works for new tasks or environments. For example, an object detection model trained to recognize certain objects can be modified to detect new categories or to work better with a specific camera setup. Object detection is a computer vision task involving automatically identifying and labeling objects in images, and researchers commonly fine-tune these so-called 'pre-trained' detectors. Doing so requires active interventions such as collecting and labeling new data, adjusting parameters, tuning architectures, and calibrating to sensors. This question points to this trajectory and examines how different research teams repurpose and adapt existing models for their local purposes. Following these activities makes visible how circulation is accomplished and how actors together shape what models are able to do.

This dissertation is primarily based on fieldwork conducted within research and development projects in academic labs. While it incorporates observations from conferences with academic and corporate actors, industry collaborations, visits at companies, and interviews with corporate actors, the primary focus on academic settings may be a limitation to the study. Furthermore, this research concentrates on the interactional knowledge dynamics within research and development processes. Dimensions related to political economy, governance, and surveillance are bracketed to foreground a more sustained analysis of research and development activities. Similarly, the construction of data centers and other infrastructural and commercialization components that play an important role in stabilizing this technology fall outside the central scope of this study.

Significant attention is currently directed at corporate labs and large language models. This research takes interest in academic computer vision labs operating outside that immediate spotlight, whose projects are no less consequential despite the lack of limelight. In industry, high computing power is concentrated in large technology companies, which have in-house clusters and access to proprietary hardware and datasets. In academia, only a few research labs achieve comparable capacity, often through exceptional funding, national supercomputing centers, or joint industry–academia projects, and such cases are rare. In academia, most university labs operate with limited resources, relying on shared clusters, grant-based access, and have little capacity for large-scale experiments. Conducted during a period of remarkable momentum in the field, I attempt to trace how actors wield and utilize recent advancements within academic research settings, occasionally in close proximity to industrial research labs and language models.

Moreover, as seen in the variety of neural network architectures developed over the past decade, computer vision systems can be created in different ways (e.g., Maslej et al. 2025). In other words, there is not one single path that merits focused examination. It is therefore more accurate to consider these paths as plural and diverse. For this reason, I focus on the activities that shape research and development rather than on a single model or application.

1.3 Outline

Following this introduction, the next chapter provides an analytical background to situate the contemporary approach to computer vision and the current conditions for research and development. The study then presents a literature review, and develops the theoretical framework, drawing on the pragmatist tradition. The subsequent methods chapter details the ethnographic fieldwork and analysis strategies used. Methodologically, the research uses a multi-sited ethnographic approach to study research and development across different phases of work and institutional settings. This framework informs the analysis presented in the following overview of the three analytical texts.

Through the ethnographic analysis, this dissertation develops concepts from situated data that help describe how computer vision scientists navigate research and development work processes across central clusters of activity. These concepts bridge empirical observations with theoretical insights about technological development and scientific work.

The dissertation concludes by synthesizing the findings across the articles and the contribution to our understanding of how computer vision technologies are shaped through situated work, and suggests avenues for future research. The findings have implications for efforts to develop more accountable and responsible computer vision systems by directing attention to points at which values and trade-offs are enacted within situated practices in research and development settings. The findings contribute to science and technology studies, while engaging with vocabulary from artificial intelligence communities, creating bridges between these areas. The extensive work that underlies these computational systems challenges narratives of technological autonomy, pointing to the activities required to make these systems function.

2 A Glimpse into the Worlds of Computer Vision

This dissertation emerges from an immersive engagement with the computer vision community during a period of significant transformation, 2021–2025, a time that felt like the eye of the storm. This chapter situates the research within that context, analyzing the conditions that influence work on the ground.

This chapter interprets the emergence of the present juncture by analyzing how some developments gained their significance: how deep learning became the “right tool for the job” (Clarke and Fujimura, 1992) and the contested political arguments that arose in close pursuit, behaving similar to an “arena” (Strauss, 1993).

Discussions with scientists were crucial to identifying the developments chronicled here. This offers insight into how some insiders consider the forces fueling what Hughes (1983) termed "technological momentum" within these worlds, motivating the discussion of technical and mathematical dimensions. Meanwhile, a sociological perspective on these developments is served well by a critical stance vis-a-vis the “rhetoric of progress” (Woolgar, 1985), similar to what recent scholarship terms “hype” (e.g., Bender and Hanna, 2025), often drawing on purely technical narratives, stemming from an “engineering model” of change isolating information processing capabilities as important drivers that explain social action and change (Kling, 1991); or mobilizing heroic figures (Haraway, 1989) such as ‘grandfathers’ and ‘grandmothers’, which tends to be incorporated in internal histories of science and technology (Bimber, 1990). It is preferable not to overstate the role of technical developments in explaining social change. However, their importance should not be undermined, either (see Radder, 1992, for a discussion). Attending to field developments is a way of approaching the ‘contextual conditions’ (Strauss, 1993) of computer vision practice. The chapter therefore contributes further commentary to parse out these developments, focusing on the material knowledge dynamics that come into play. Understanding these central developments helps us comprehend how the pursuit of teaching computers to see is approached in contemporary practice.

2.1 Deep Learning as the Right Tool for the Job

Computer vision as a discipline and practice emerged in the 1950s, involving various approaches to the prospect of teaching computers to see as part of bringing forth artificial intelligence (Dobson, 2023). Historians of computing sometimes trace the origins of this discipline to figures like McCarthy, who coined the term “artificial intelligence” in the 1950s and defined it as the science of creating intelligent machines. Samuel defined machine learning in 1959 as enabling computers to learn without explicit programming. The trajectory of artificial intelligence as a field is associated with a seasonal narrative drawing on ‘summers’ (funding and advancement) and ‘winters’ (lack of funding and advancements). Stories about early attempts to supposedly ‘solve vision’ over the course of a summer in 1966 as part of creating artificial intelligence are often evoked. And analysts draw attention to the discipline’s close and persistent ties to the military and warfare applications (Suchman, 2022). Computer vision is in the proximity of this narrative.

One of the earliest articulations of attaching computers to cameras was conceived by Rosenblatt's in his 1957 work on the Perceptron, which, in some sense, laid the groundwork for machine learning in computer vision by introducing a neural network model based on image sensors. His probabilistic framework involved learning from experience, enabling machines to recognize objects despite variations in attributes such as size, orientation, or background. Although overshadowed by symbol-based artificial intelligence research in later decades, the principles of the Perceptron saw a resurgence with advancements in neural networks in the late 20th century, including backpropagation and gradient descent algorithms (Olazaran, 1996), also influencing the renewed uptake of neural networks of today. Another significant contribution came in the 1960’s, when Roberts (1963) published “Machine Perception of Three-Dimensional Solids”. Roberts' work established essential principles in computer vision, connecting geometric computation with visual perception and laying the foundation for three-dimensional modeling, which continues to be central to modern computer vision. From what I gather from journal contributions and conference proceedings over time, computer vision techniques have occasionally splintered into and taken on scientific rubrics like image processing, signal processing, pattern recognition, remote sensing, geometry, machine learning, and recently, deep learning, or simply just artificial intelligence.

Pixels are one of the most central objects in computer vision. Pixels consist of numerical values, corresponding to color intensity in a fixed range of values, making up the atoms of quantified imagery (Dobson, 2023). Recent artificial intelligence techniques have served to deepen the significance of pixels, and contemporary computer vision models rely on them in their role as image data. Pixels are significant units of work, and often serve as the stepping-stone to more sophisticated representations.

In the current era of artificial intelligence, the central object is the artificial neural network. From a mathematical perspective, grounded in the canonical Universal Approximation Theorem, artificial neural networks promise users the ability to model complex data distributions and to approximate their underlying function (Cybenko, 1989). This theorem powers the ambition of developing models that successfully generalize to new data. Practitioners typically refer to 'models' as specific artificial neural networks trained on data. During inference, it is the trained model (rather than the individual algorithms used during training) that generates predictions at so-called 'test-time' (Szeliski, 2010:222).

In their computational construction, artificial neural networks operate through nodes organized into layers, each node processing input data and passing outputs to subsequent nodes (Alpaydin, 2020). An important aspect of learning in neural networks is weight adjustment, which modifies the strengths of connections between nodes to effectively process data. The process involves dividing data into training and validation sets, initializing the network's weights randomly, and updating these weights. By propagating errors backward from the output layer to the input layer, values march iteratively through the network (forward pass, backward pass). This adjustment is typically achieved through an algorithm that computes gradients called 'backpropagation', passing them along within the network so as to adjust its weights; and an optimization algorithm like Adaptive Moment Estimation, which adjusts learning rates (Szeliski, 2010). This recursive updating enhances the network's ability to converge, getting estimates closer to the 'right answer' that corresponds to a dataset's specific distribution of values. In this sense, core computer vision architectures, such as artificial neural networks, make up the backbone of modeling efforts, while supporting algorithms, such as optimization routines, serve as infrastructural scaffolds enabling their application (cf. Szeliski, 2010).

The versatility of approximating data distributions is a key explanation for why artificial neural networks have found extensive reach and are understood as being capable of addressing a wide array of problems. As data distribution becomes seized under mathematical control, the updated weights of neural networks, encapsulating learned patterns of a dataset, become increasingly treated as central repositories of contemporary knowledge. These capabilities of artificial neural networks represent how deep learning can be considered central to knowledge processes in our societies. This tendency extends beyond computer vision and throughout academia, the private sector, and public institutions, with high-profile projects like AlphaFold and Nobel Prize-winning developments, indicating their force. And, not least, an unrelenting series of initiatives with less media coverage attests to the same.

From the perspective of deep learning's computational structure, algorithms govern key processes such as gradient descent for optimization, backpropagation for updating weights, and matrix operations for forward propagation. Dean (2022:60) describes that "the computations needed for deep learning algorithms are almost entirely composed of different sequences of linear algebra operations on dense matrices or vectors, such as matrix multiplications or vector dot products".

Dean (2022:60–61) explains that graphics processing units or GPUs, optimized for graphics processing and parallel 'floating-point' operations such as matrix multiplications, were found to be well-suited for deep learning workloads. Dean discusses how GPUs are optimized to perform numerous calculations in parallel, thereby increasing computational efficiency. While 'high precision' refers to a more accurate and detailed numerical representation (more numbers), 'low-precision' computation uses fewer bits to represent numbers (fewer numbers). In deep learning, extensive numerical precision is often considered unnecessary because the process relies on approximations made at a large scale rather than exact calculations at every step. This arrangement can be compared to navigating with a compass, where the general direction, rather than precise accuracy at each computation, is desirable. This capability, he notes, spurred the development of specialized 'accelerators' such as Tensor Processing Units, and devices designed to orchestrate multiple chips for large-scale computational workloads.

Bengio, LeCun, and Hinton (2021:63) note a dependency in the foundation of machine learning models: 'the i.i.d. assumption', short for 'independent and identically distributed', which presumes that test data will mirror the distribution of the training data. This assumption, they argue, contributes to the problem of "out-of-distribution generalization", where artificial neural networks struggle to perform when faced with new data scenarios that differ from the distribution of values within the dataset used in their training arrangements.

Artificial intelligence scientists tend to attribute the recent success and the boom of deep learning to events occurring in 2012. Li and Krishna (2022:90) describe it as a "watershed moment", which "inspired other researchers...to shift to deep learning approaches". More precisely, the announcement of the computational model, AlexNet: a "convolutional neural network" (Krizhevsky, Sutskever and Hinton, 2012). Despite early and foundational progress (Fukushima, 1969; LeCun et al. 1995), this architecture remained relatively niche until this event. This model achieved significant success in classifying images into various categories in the ImageNet Large Scale Visual Recognition Challenge. It performed better than its competitors in predicting the right (semantic) category for images, counted through the top five predictions with the highest probability scores. This accomplishment is considered a breakthrough for data-centric machine learning in the field of computer vision, leading up to what Dean (2022) calls the "golden decade of deep learning".

There is wide agreement in the computer science community that the key contribution was the role of model depth in achieving such performance. In machine learning, the term 'model depth' refers to the number of layers containing data-processing nodes inside a neural network, and underlies the phrase 'deep learning'. This depth came with significant computational costs, but the utilization of GPUs, instead of central processing units (CPUs) during the training process, rendered the technical approach feasible. This event tends to be sourced as the key lever that actuated social change: a catalyst that made further action align in a shared direction. For instance, this event is sometimes recounted as spawning the ongoing rush to increase compute budgets through amassing GPUs.

Insiders sometimes emphasize the massive training data that underlie the success of the model as another central factor, commonly expressed as training the network 'at scale'. As the name of the competition indicates, the model was related to the dataset, ImageNet. From a knowledge perspective, it is worth noting the following regarding ImageNet. It was the first massive-scale dataset designed to collect and catalog images of tangible physical objects (Li, 2023:146). The ambition behind ImageNet was monumental at the time. Commenting on the prospect of collecting such massive data in her memoir, the principal scientist of the project, Li (2023:140), reflects:

> My thoughts raced as I imagined the wealth of visual cues an algorithm trained on such a dataset might internalize. The hard edges of plastic, the sheen of lacquered wood, the texture of an animal's fur, the reflections on the surface of an eye, and so much else—perhaps everything else. I envisioned our algorithms growing ever more flexible in their ability to separate foregrounds from backgrounds, to distinguish where one object ends and another begins, and to separate light and shadow from surface and volume…What if the secret to recognizing anything was a training set that included everything?

The ambition to catalog 'everything' indicates a desire to capture the pixel relationships of all objects, enabling computational systems to accumulate enough exposure to data to learn their structure across the board. To label millions of images, a practice known as 'data annotation', Li and her colleagues engaged a loose network of annotators through crowd-sourced labor via the 'click-work' platform Amazon Mechanical Turk. This approach provided the workforce needed to annotate each image. Social scientists and computer scientists have since criticized ImageNet, as annotations have been shown to include harmful language, raising ethical concerns, and exploiting cheap and menial labor (Denton et al. 2021).

The construction required strenuous efforts to collect and store millions of images downloaded from digital platforms, showing the research laboratory's extending its activity to external sites that stored image files captured elsewhere. From a knowledge perspective, society, represented by the data annotators, entered the lab under limited conditions. This demonstrates how the laboratory extended into society, transforming it—at least temporarily and at a distance—into part of the laboratory's infrastructure.

This reflects Latour's (1987) perspective on the actions typical of "centers of calculation", where data sourced from elsewhere are processed within the laboratory to produce knowledge. ImageNet shows how knowledge production transcends the boundaries of the laboratory by creating networks of collaboration and delegating work to various actors. Li's lab employed a distributed group of annotators, researchers, and engineers to advance the pursuit of computer vision, transforming disparate sites and individuals into components of scientific work. This mirrors Latour's (1988) analysis of how Pasteur enrolled proxy sites and collaborators to expand his efforts.

In contrast to conventional natural science laboratories where physical experiments, specialized instruments, and controlled environments often mark the boundaries between scientific work and the outside world, computer vision labs operate with a different kind of spatial logic. This shrinkage of distance manifests in multiple ways. Their products are not confined to the site of their creation but are built to function remotely, often without direct oversight from their developers. A model trained in an academic lab may be integrated into an industrial application halfway across the world; another, optimized for a specific dataset, may soon be running on millions of smartphones. And data collection from cameras in car cabins may lead to the fruition of a new safety mechanism turned into a product offering. This capacity for deployment at a distance collapses traditional separations between research and application, making computer vision labs more entangled with external areas of society than their equivalents in experimental science. Research labs sometimes share their models in online repositories to advance the field and garner their reputation, making them accessible to developers and practitioners elsewhere. The lab, in this sense, is both a site of production and a relay point in a distributed network of inputs and outputs; some more than others.

Another major development in deep learning occurred with "The Attention is All You Need" (Vaswani et al., 2017) paper, which introduced transformer architecture in 2017, which was originally developed for natural language processing tasks. Other important architecture advancing at around this time include generative adversarial networks (Goodfellow et al., 2014). These consist of two networks, a generator and a discriminator, which compete against each other to improve the quality of the output, sometimes used as a tool to generate synthetic data (see Castells and Roberge, 2021). However, the advent of transformers has had a massive influence in areas beyond language, particularly in computer vision tasks that require long-range dependencies and contextual understanding. Transformers changed how computers process different text types of data, like images, which makes it relevant to computer vision research. Reducing the boundaries between these data formats, transformers encode these data into continuous vector representations, called tokens, in a latent space. These break down input data, like images, into smaller, manageable units. Instead of representing an image with all its pixel data, it is represented by a vector of numbers. The paper introduced 'self-attention' as a design choice to capture the most relevant relationships between tokens in a sequence. The numbers in the vector thus capture some aspect of the token's features in relation to other tokens.

Many have since interacted with a version of the transformer architecture through OpenAI's platform, ChatGPT, the introduction of which commentators attribute as the 'iPhone moment' in the field, making it broadly accessible to millions of users (Forman, 2023). Over time, ChatGPT incorporated computer vision capabilities. Since the original paper and ChatGPT, transformers have become the backbone of breakthroughs in many areas, like natural language processing, generative text-to-image, and video models that deal with sequenced image frames, as well as multimodal models combining both language and vision capabilities in a shared latent space. In 2025, ChatGPT introduced 'visual reasoning' where these capabilities are combined, iteratively analyzing images by crossing over visual and semiotic registers. The transformer transformed the world, and contributed to stirring a cacophony about 'Generative AI'.

The open-sourcing of key technologies accelerated the development. Some suggest that Google's decision to make certain technologies accessible to the wider community stemmed from the perception that they were primarily basic research, lacking direct applicability for immediate competitive advantage. This shows the ongoing tension between basic and applied research in science and technology, and how this tension can influence the flow of information and resources between different actors (e.g., Friedland, 2023). Moreover, it demonstrates how research and engineering advances occur in industry and academia and the motion in-between.

While the spotlight in deep learning for computer vision often shines on novel model architectures, the less visible, yet crucial, foundation upon which contemporary research and development is built consists of infrastructural tools and frameworks. Borrowing Star's (1999) notion of "boring" things, these are epistemically relevant, structuring the possibilities of knowledge production in the field. These 'infra' and 'development and operation' or DevOps tools, as they are sometimes termed within the community, are active constituents of the development process. This infrastructure helps manage the complexity of deep learning projects. Beyond specialized tools for deep learning, others relevant to the broader development community, like GitHub, are sites for collaborative code storage and version control. These platforms support collaborative and networked engineering practices, shaping the situated work through which deep learning models are developed. A far-reaching mediatization ecology concerning deep learning and artificial intelligence is present. This ecology involves extensive conventional news coverage, but also podcasts, tech influencers, newsletters, forums, and lectures, all accessible online, further accelerating knowledge circulation. There are therefore many spaces through which engaged actors track developments.

The notion of so-called ‘scaling laws’ is a central hypothesis, or perhaps an imperative, guiding recent deep learning research; particularly within corporate labs, but academic labs relate to it, as well. An artificial intelligence scientist named Sutton (2019) wrote a blog post called “The Bitter Lesson”, and in it, he posited that the most effective way to achieve artificial intelligence is through methods that can leverage ever-increasing amounts of computation. He also argued that while clever engineering and algorithmic innovations might offer temporary gains, they ultimately pale in comparison to the relentless power of scale. In 2020, researchers at OpenAI published a paper, “Scaling Laws for Neural Language Models” (Kaplan et al., 2020), in which they identify an empirical trend, seemingly validating Sutton's theory. The paper compares how well different versions of the Transformer architecture fare on language tasks. Akin to Moore's Law, which regards generational performance increase in processing units widely popular in computer science, they observe that the performance increase in artificial intelligence models is directly tied to increases in "training time, context length, dataset size, model size, and compute budget" (Kaplan et al., 2020:2). This convergence of theoretical prediction and empirical observation solidified the scaling imperative within the machine learning community. This emphasis on scale is not new, with precedents in earlier research and development efforts like AlexNet's focus on model depth and ImageNet's emphasis on dataset size.

The fieldwork at the heart of this dissertation ended at the beginning of 2025. At that point, the scaling paradigm, while still influential, had begun to be questioned. Analysts increasingly emphasized the energy demands required by processing massive troves of data, raising concerns regarding environmental sustainability. Meanwhile, OpenAI research opened up a new area ripe for scaling, shifting the focus from targeting computation-intense efforts at pre-training, to so-called test-time: when inferences are performed by the model. This scaling at inference involved a new avenue for further performance gains, further entrenching the scaling logic. However, clever engineering by the research lab DeepSeek showed the possibility of tweaking the model architecture so that the ‘compute’ demands otherwise necessitated in the scaling imperative, both during training and inference, radically decreased in volume. A key takeaway amongst analysts was that DeepSeek's work showed that architectural innovation offered a path to improved performance without solely relying on brute-force scaling, thus questioning the massive billion-dollar investments made by major research labs in their attempts to scale up existing models (e.g., Mok, 2025).

Against this backdrop, the current trajectory of deep learning-driven artificial intelligence remains open-ended: its future course is shaped by these conditions. The end products of these ostensibly modernist systems, promising intelligent capabilities distilled from data and rooted in Enlightenment ideals of calculability, exhibit significant variation. We may conclude this discussion by pointing to the specific characteristics such end products become attributed to by looking at an influential definition of an ‘AI system’ developed as part of the recent Artificial Intelligence Act in the European Union (2024):

> a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments

As seen, autonomy is a central characteristic attributed to these systems (cf. Winner, 1977). These computational systems use artificial neural networks trained on vast datasets, resulting in models that act as instruments for further action: technically achieved through inference from an often-complex data distribution. An artificial intelligence system, therefore, can be understood as comprising computational models and automated processes that generate actions from data to perform further actions in situations within specific institutional contexts—the details of which are contingent upon the situated research and development work explored in the following chapters.

2.2 Arena: Contested Dimensions

The computing world is experiencing turmoil for several reasons. Some view recent artificial intelligence developments as standing upon the cusp of an unprecedented technological shift. This is rooted in the belief that a shift to ‘artificial general intelligence’ or ‘AGI’ is on the way, suggestive of a tectonic change while echoing the deepest layer of the modernist vision of technological progress. The attribution of historical agency to ‘AGI’ reflects a narrative in which technological systems are a transformative agent and a repository for collective hopes and anxieties about the future of human civilization. This transformation is represented as a Kuhnian (1962) “paradigm shift”, where existing sense-making categories struggle to latch onto the new capabilities of computing. This status trades on its potential to unlock new forms of knowledge, reorganize labor, and reframe societal priorities. Even without the consolidation of artificial general intelligence, deep learning propels this vision of unparalleled impact in its own right: neural is not neutral.

Common among these groups is a propensity for ‘rupture talk’, a term used since the 1990s to describe claims focused on profound societal shifts due to technology (Hecht, 2002). Such talk is often deconstructed or critiqued through the lens of technological solutionism, a critique that deconstructs these tech-driven applications and fixes as instances of inadequate or exploitative change theories (e.g., Morozov, 2013). In artificial intelligence, some actors circulate compelling stories of human redundancy in the labor market and existential extinction. Dire warnings from alarmists, or ‘doomers’ as termed by their optimistic counterparts known as ‘accelerationists’ mix with more confident and idealist claims. Another group engages in attributing human-like qualities to technology, such as personality or human rights (Collins, 2018; see McDermott, 1976).

As one of my interlocutors said, the current mobilization, urgency, and large-scale efforts around artificial intelligence resemble a contemporary ‘Manhattan Project’ reaching backwards in time to the organization around the emergence of the nuclear bomb. Nation-states and venture capital firms are channeling vast resources into the development of artificial intelligence, which shows the strategic importance and competitive stakes at play. This may be interpreted as part of the core machinery of “empire” (Crawford and Joler, 2023), with the relevant technologies refurbishing its “centers of calculation” (Latour, 1987). We need not introduce politics into this world; it is already there.

An extensive actor ecology, including grassroots movements, advocacy groups, technology journalists, and critical computer scientists (e.g., Kalluri et al. 2023), has been instrumental in bringing controversial and contentious dimensions of artificial intelligence to light. Several recent documented cases of the use of these technologies provoke direct unease. The repository managed by the independent organization AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC, 2025) contains hundreds of inflammatory instances, including racist and sexist algorithmic outputs in a host of other cases (see also Buolamwini and Gebru, 2018). For instance, a Hollywood union representing artists, actors, and other creative professionals conducted the first collective action in 2023 in response to artificial intelligence technology, protesting a particular strain of generative models. These models are using copyrighted material without compensating creators, and generating content. To combat this, artists engage in data poisoning, adding imperceptible pixels to artworks before uploading them online, aiming to protect their work from unauthorized use (Heikkilä, 2023). This legal action is one example of numerous recent controversies in artificial intelligence. Marres, Castelle, and Tripp (2024) describe the wave of contestations as a "super-controversy" because heterogenous concerns converge into one line of debate, often in mediatized settings, against the background of more conventional scientific controversies taking place in academic seminar rooms (see also Suchman, 2023).

One characteristic of the current controversy is that several parallel discourses exist in parallel. Arenas are often controversies, following Strauss, which tend to appear similarly to the spiraling movement of a 'whirlpool'. What Strauss (1993:227) refers to as arenas or "interaction by [members of] social worlds around issues" of mutual concern provides an analytical handle on the collective indetermination of the broader situation. The arena of artificial intelligence is rife with extraordinary statements circulating meanings on fundamental societal considerations, directly reproducing modernist fantasies, while others aim at governing the technology in a more grounded fashion. Some researchers articulate these modernist ideals in a reflexive mode, conscious of its limitations and dangers. This mirrors Beck's (1992) thesis on reflexive modernization, in which societies confront and respond to the consequences of their technological foundations, such as power concentration and environmental hazards, exemplified in the recent debates around the energy demands of data centers, and in an epistemological register, the language models' mimicry in the 'stochastic parrots' debate, and on the ethics of the technology (e.g., Bender et al., 2021).

3 Reviewing Research Directions in the Social Study of Technology and Computation

The purpose of this literature review is to establish a foundation for the inquiry. Readers are invited to engage with this chapter as a contextualizing supplement, compared to the more focused and targeted literature reviews featured in the individual analytical texts. This review focuses on how existing scholarship has researched technology and computation as objects of study. The purpose is to organize the contours of social research of a complex object with numerous incarnations, which is why I review social studies of science and technology, social theory, and interdisciplinary fields.

There is a massive amount of empirical research that addresses processes relevant to science, engineering, computation, technology, data, instruments, and knowledge production. I distinguish between waves that have characterized some of this research, informing present social research of computation. The wave metaphor is an effort to organize analytical dispositions vis-à-vis similar objects of study. The distinctions include a programmatic wave including ethnographic lab studies, a forked wave, and two strands of social scholarship on artificial intelligence.

The programmatic wave, exemplified in the proliferation of acronyms such as LTS, SSK, SCOT, and ANT, imported frameworks developed in the sociology of science as the interdisciplinary field of social studies of science and technology (STS) emerged. The programmatic wave went beyond the study of technology as a second-order social processes involving industrialization and rationalization; envisioning it mainly as products related to technical rationality, alienation, and deskilling, related to the genre of workplace and impact studies, to a first-order process. From then and onwards, a forked wave erupts, where increasingly diverse technology applications and increasingly specialized lines of research become entrenched. In one of these lines of research, I trace and outline key debates and themes in the study of computation, drawing on sociology of algorithms as well as the interdisciplinary fields of critical data, algorithm, and artificial intelligence studies. I refer to STS-inflected research as social studies of artificial intelligence, and further distinguish two generations following the distinction between two kinds of artificial intelligence: the symbolic, rule-based kind, and the machine learning orientation.

3.1 The Programmatic Wave

In social theory, the problem of technology tends to be ambiguous because of the many terms that are used to designate it. Numerous sociologists have explored technology in its various forms. In the 1970s, social scientists begin to elaborate on the concept of technology; specifically, how to theorize and research it. Several research programs, typically known by their acronyms, were introduced in line with a set of narrower traditions such as the frameworks emanating from the sociology of science.

This took place in the interdisciplinary field that crystallized under the banner of Science, Technology, and Society (STS), which is now called 'social studies of science and technology', or just by its abbreviation, STS. This field is a diverse enterprise, including philosophers, anthropologists, sometimes media studies scholars, and critical technologists. Its origin stories describe its emergence as a fusion of sociological and historical studies of science—post-Merton and post-Kuhn—with activist research on issues such as nuclear weapons and sexual reproduction technology in the 1960s; other forerunners mentioned in such origin stories sometimes include the study of ideologies and knowledge following Marx and Mannheim, and Fleck's thought collectives (e.g., Sismondo, 2010).

In this field, the understanding of science and technology as central constituents of the social order is common. Technological systems have been described both as 'missing masses' in social analyses of society (Latour, 1992) as well as a 'shadow constitution' that contributes to shaping social action and relationships lurking in between more overt forms of power (Smits, 2001).

In the 1980s, sociologists brought concerns and frameworks developed in the sociology of scientific knowledge to the study of technology. Distinct from earlier Mertonian sociology of science, which analyzed science as a social institution, the sociology of scientific knowledge housed different theories that claimed to explain scientific knowledge in its own right. The convergence between the sociology of science and technology is articulated most clearly by Woolgar's (1985:558) article

"Why Not a Sociology of Machines? The Case of Sociology and Artificial Intelligence" in which he asks, "How can a machine (science) be the object of fruitful sociological inquiry?" and is moreover indicated in texts such as Pinch and Bijker's (1984) "The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other", and later ones such as "The Turn to Technology in Social Studies of Science" (Woolgar, 1991). The focus of analytical emphasis is shifting towards examining the contents of technology and scientific knowledge as such, rather than its organizational structures or the functions of science proposed by Merton (1973). Relatedly, the ASA Section on Science, Knowledge, and Technology (SKAT) was founded in 1987. The first chair, Etzkowitz (ASA, 1988:4) argued, "[a]s the action moves to science, knowledge and technology; so will sociologists" as a response to critical sociologists who at the time stated that "too much attention to such 'dependent variables'" (i.e. technology) would push the scholar outside the bounds of what is acceptable areas of study for sociologists (ASA, 1988). In this vein, Law (1987:406) differentiated the "new sociology of technology" in its focus on how technology is created, unlike related fields that study technology's effects on workplace social dynamics, sometimes referred to as 'impact' or 'workplace' studies (see Heath, Knoblauch and Luff, 2000). Wajcman (2002) comments that what was once called the "new sociology of technology" became referred to simply as "constructionist studies of technology". Actor network theorists contributed to this approach as well (e.g., Callon, Law, and Rip, 1986; Latour, 1990). Scholars presented concepts such as "inscriptions", and processes through which designers allocate specific roles, what Akrich (1992) termed "scripts", while allowing for user resistance or "antiprograms" that can reconfigure these allocations (Latour, 1992).

In sum, a central problem within this strand of sociology and social theory is the issue of how science and technology become produced: moving science and technology from a second-order process to the first-order. A common move is to study science and engineering practices as being responsible for the emergence and integration of technological objects and orders into the fabrics of society.

Several research programs were developed during the programmatic wave. A central reference is Winner's (1980) analysis, which asks whether technological artifacts have politics. Pinch and Bijker's (1987) Social Construction of Technology (SCOT) considered the constitution of technology through competing social groups that were establishing varying object definitions that circumscribe their meanings but retain interpretive flexibility among its users. The Large Technical Systems (LTS) perspective shifted focus to actors as system builders responsible for assembling and maintaining vast infrastructural networks, thus tracing the socio-technical dynamics at play in their evolution (Hughes, 1983, 1987). Technical objects of various kinds are implicated in social accomplishments and rearrangements of sociality—through and through (Latour, 1993; Mead, 1932). Another vein of research takes an ethnomethodological stance, directing attention to the everyday practices through which actors collectively produce order, drawing attention to the tacit, procedural work, foregrounded by Button (1993) in analyses of software development. Kling and Scacchi's (1982) theorized people and computing as an intertwined "web", noting the interdependencies among organizational routines, institutional norms, and technical configurations. Other foundational anchors, especially in social studies of science and technology, include technoscience, a neologism often attributed to Bachelard (1953), intended to better encapsulate the crisscrossing of science or knowledge production and technology by fitting them into a single category. Bachelard realized that the production of scientific knowledge became dependent upon technological instruments and vice versa, making up a dense intersection. Material instruments are the means of knowledge production itself. Engineers use knowledge and artifacts to achieve their objectives, whereas scientists turn to engineering techniques to advance their knowledge. As an indicator of social change, this reformulation helped illuminate the social experience of a blur in the perceived division between science and technology.

Woolgar and Grint (1997) describe the problems of essentialism and determinism as negative counterpoints that characterized the programmatic wave. The stability associated with technological objects, assuming some essence, reifies the very object. Analysts risk contributing to essentializing those artifacts when describing their makeup as a stable thing. Technology undergoes significant transformations in several directions, especially concerning the extent to which sociological analyses take an interest in technology as a social process. Technologies are designed and attributed with capacities, but their use differ significantly. Woolgar and his colleagues reconceptualized technology as a form of text, drawing parallels between authors and readers, and thus the possible readings of technological objects. In an ethnographic study researching the development of personal computers, engineers configured users' actions through textual intervention (Woolgar, 1991).

Woolgar and Neyland's (2013:17) more recent articulation highlights the local contingency and occasioned character of science and technology, probing how objects and technologies are contextually shaped and understood—or, in their words, become "contingently enacted". The authors strive to move beyond hidden biases toward latent realism or constructivism, sometimes instantiated in contemporary concepts such as 'affordances'. Their inquiry reveals that the characteristics and existence of these objects are entwined with interpretations and uses, with stability being an uncommon outcome (Woolgar and Neyland, 2013:46). The shortcoming of this approach is that preoccupations with local accomplishments create a slippery type of situationism, in which unique, relevant relationships imbuing the scene only appear momentarily by actors in locally situated sites. While attempting to do so, they also acknowledge the difficulty in completely avoiding notions of essentialism and determinism. Woolgar and Neyland suggest that while their view acknowledges governance and accountability as emerging from practical ontologies, or how the object is enacted in a local context, this perspective subtly alters determinism by focusing on the emergence or appearance of such outcomes (cf. Winner, 1980). Despite their careful treatment, they concede that burdened traces of essentialism and determinism remain in their approach, asking "is it ever possible to be rid of these?" (Woolgar and Neyland, 2013:261).

More generally, sensibilities associated with the ethnographic lab studies of knowledge production conducted in science laboratories is central here (Latour and Woolgar, 1979; Knorr-Cetina, 1981; Collins and Pinch, 1982; Gilbert and Mulkay, 1984; Lynch, 1985; Traweek, 1988). The regime of 'scientific programming' in science, like research on processing images for scientific purposes (e.g., Lynch, 1991; see also Vertesi, 2015), draws together knowledge production, computation and engineering. Studies of computing in science and technology studies focused on issues such as scientific representations, simulations, and visualizations highlighting digital work practices (e.g., Carusi et al. 2014; Turkle, 2009; Amann and Knorr Cetina, 1988), and the related and broader engineering-centric literature on social studies of technical work and organization (e.g., Latour, 1978; Whalley, 1986; Orr, 1996; Henderson, 1991; Kunda, 1992).

In sum, this expansion of sociological terrain proved productive. It destabilized conventional distinctions between technology and society, which proved fruitful for better understanding the production of science and technology (e.g., Latour, 1999). This tradition made sociologists consider technology as more than 'applied science', and not as an entity that 'impacts society' as an external or independent factor (Mackenzie and Wajcman, 1999; cf. Winner, 1977). In more recent research in social studies of algorithms, Seaver (2017:10) extends these insights by suggesting that algorithms tend to be understood as causal and independent variables acting upon culture, or, more advantageously, as components of culture, or society. Lee and colleagues (2019:1) express a similar temperament as they note that "algorithms are in society, they do not control society". Aligning with this sociological approach, this dissertation focuses on the activities performed in computer laboratories following subsequent findings in social studies of science and technology, suggesting that research labs indeed are 'where the action is' (cf. Dourish, 2001).

3.2 The Forked Wave

Technical doctrines multiplied at an overwhelming pace in the 1990s onwards, and their objects appear ever more fine-grained. A cascade of specialized terms suggests computational diversification: machine, computer, program, data structure, algorithm, software, internet, Web 2, Web 3, platform, digital application, digitization, digitalization, big data, cloud, digital twins, automated decision-making, machine learning, deep learning, artificial intelligence, humanoid—the terms proliferate. The concept of digital technology seems to have been stretched to the point of it almost becoming devoid of meaning. Koselleck (1982:87) suggests that situations such as these are anything but random; on the contrary; "transformation in the meaning of words and of things, change of situation, and impulse to rename things, all of these correspond diversely one with another". A central perspective is the material expansion of computational technology, which these terms hint at. In this regard, this proliferation is linked to a strong corporate drive which distinguishes 'new' technical products from earlier ones (cf. Winner, 1986). Engendering products viable for commercialization, the "innovation imperative" (Pfotenhauer, Juhl and Aarden, 2019) reinforces the status of 'the groundbreaking' in connection to technology.

Social scientific discourse has proved itself readily available to adopt these distinctions. From the end of the 20th century onwards, the multiple co-existing complexes of differentiated and specialized technologies appear foundational to such characterizations of society as ‘post-industrial’, ‘knowledge’, or ‘late modern’. In other words, technology becomes recruited to designate dominant social processes such as digitalization or datafication in societies with material apparatuses reflected in the main theses themselves.

Relatedly, Sehr undertook the task of perusing relevant sociological nomenclature appearing during recent decades, having himself popularized the term "knowledge society", and more recently developed his thesis through "knowledge capitalism". He (2022:142-143) rightfully asks and lists:

> What justifies the designation of the emerging society as a knowledge society rather than, as is frequently the case, a science society (e.g. Kreibich, 1986), an information society (e.g. Nora and Minc, 1980; Duf et al., 1996; May, 2000; Webster, 2002), a network society (Castells, 1996; Schiller, 1999), a post-industrial society (Bell, 1964), as technological civilization (Richta, 1969), Risk Society (Beck, [1986] 1992), post-capitalist society (Drucker, 1994), Audit Society (Power, 1997), cognitive capitalism (Vercellone, (2007), digital capitalism (Nachtwey and Staab, 2015), finance capitalism (e.g. Fraser, 2017), “digital civilization”, “Artificial Intelligence Society” (Inglehart, 2018), or the “algorithmic society” (Schuilenburg and Peeters, 2020)”…“platform capitalism” (Smizik, 2017), a “surveillance society” (Zubof, 2019), or “digital colonialism” (Kwet, 2021).

This collection of concepts is like a firework bursting into small fragments, each illuminating different dimensions of the consequences of science and technology in the last decades. Castell’s (1996) treatise on the network society, first coined by van Dijk (1991), addresses the social organization and implications of information and communication technology, which he argues restructure basic units such as space and time and bring about new types of inequalities. Relevant concepts include “knowing capitalism”, which highlights how data analyses are wound up in capital interests (Thrift, 2005); the symbolic interactionist notion of the “terminal self” living among screens (Gottschalk, 2018); or Farrell and Fourcade’s (2023:232) "high-tech modernism", which aims to capture how computational “tools and their designers categorize and order things, people, and situations” reminiscent of bureaucratic and market processes. Some important touchstones in social theory that may be further invoked to designate social tendencies relevant to this dissertation are “algorithmic society” (Schuilenburg and Peeters, 2020; Burrell and Fourcade, 2021), “information age” (Castells, 1996), or "datafication", broadly meaning "taking information about all things under the sun—including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge—and transforming it into a data format to make it quantified" (Mayer-Schönberger and Cukier, 2013:15). In response, Sehr (2022) argues that knowledge makes up the cohesive glue amidst numerous processes recently introduced in the literature. Meanwhile, computational systems are part of world-making efforts in numerous disparate areas, from knowledge production to creative processes in and across bureaucracies, markets, and other areas of social life (Fourcade and Burrell, 2021). These conceptual contours reveal the salience of technological processes and expertise in society (see Collins and Evans, 2002), and provide us with a scattered sense of direction regarding recent social experience.

An institutional precursor traces back to the emergence of engineering and science-based industry in the late 19th century, which Noble defines as an "industrial enterprise in which ongoing scientific investigation and the systematic application of scientific knowledge to the process of commodity production have become routine parts of the operation"(1979: 5). Noble sheds light on how professional engineers integrated, and seemingly got co-opted, into corporate worlds—harvesting the fruits of "the wedding of science to the useful arts"—and made empirical and experimental practice part of product development and thus the capitalist economy (Noble, 1979:24).

These trends are related to political-economic arrangements such as commodification, assetization, and rentiership (Birch, 2013). Technology, for instance, is sometimes wound up in several intellectual property rights, including copyrights, trademarks, patents, trade secrets, and industrial designs, all of which are covered by the 1995 Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement, administered by the World Trade Organization (Sehr, 2022). With these insights at the fore, sociological analyses of technology have shown that the relationship between society, science, and technology defies straightforward deductions and simple explanations (Mackenzie and Wajcman, 1999:5-7). Latour (1990:21) comments on understanding science as a Marxian "superstructure" (Sohn-Rethel, 1978), noting that the "net result of this strategy is that nothing is empirically verifiable since there is a yawning gap between general economic trends and the fine details of cognitive innovations", thus omitting a large portion of science and technology. In other instances, technology production is performed under rubrics of open science and open-source software.

After the 2000s, digital sociology became a recognized area of study, harboring themes of digital inequality and online communities (DiMaggio et al. 2004). This subfield is wide and resists reduction, it is nevertheless primarily focused on digital communities, subjectivities, and forms of digitally mediated sociality. Most sociological scholarship on technology, like digital sociology, studies how different actor groups use the technology once it is up and running, a concern also invoked by STS scholars (see Van den Scott, Sanders, and Puddephatt, 2017, for a review). However, digital sociology involved, in my reading, a relatively distinct approach compared with the research programs developed in social studies of science and technology. In STS, social analysts begin to conduct research on the politics of infrastructure (e.g., Star, 1999) in parallel with computerized, datafied, and digital infrastructure growing prevalent in society.

Meanwhile, a cohesive research strand adopted the rubric of the sociology of algorithms during the 2010s (Ziewitz, 2016). Some of these studies interpret the social impacts of computational technology by pointing to the sorting and eligibility-granting procedures that code and software enact as a mechanism of power (e.g., Beer, 2009), and related to classification and discrimination (e.g., Burrell, 2016). Burrell (2016) suggests that the reasoning for particular decisions is "opaque" in practice, because of the complex mathematical optimization techniques that underlie their data processing. Concepts such as “algorithmic culture” centering technological ‘impacts’ on society were introduced (Striphas, 2015). Scholars then began tracing these computational orderings in different areas of society, from endemic to digital settings, such as social media (Gillespie, 2014); to settings mediated by computing and algorithms, including, but not limited to, finance (Fourcade and Healy, 2013). Widely cited works found pervasive discrimination in corporate search algorithms (Noble, 2018) and ambiguous processes and discriminatory results in areas such as marketing, hiring, pedagogy, and law enforcement (O’Neil, 2016). Other studies explored technological impacts on various aspects of daily life, such as learning algorithms fed by user data (e.g., Airoldi, 2022). Compare these results with Latour's (1990:103) "Technology is society made durable", in which he writes "to understand domination we have to turn away from an exclusive concern with social relations and weave them into a fabric that includes non-human actants, actants that offer the possibility of holding society together as a durable whole". The sociology of algorithms ties into the waves of social studies of artificial intelligence, but as I show, it does not exhaust it.

Alongside these developments, the interdisciplinary fields of critical data and algorithm studies, and more recently, critical AI studies, became popular, which is more thoroughly reviewed in this dissertation's articles. The subject matter of technology and computation touches many disciplines and subfields. A consequence of this is that several existing strands of research co-exist with partial overlaps.

The level of interdisciplinarity appears to fuel the forking dynamic, sometimes hastily grouping studies as 'STS'. Adjacent clusters include media studies literature and software studies (e.g., Manovich, 2001). Other works in media studies include McCosker and Wilken's (2020); Gaboury's (2021)*;* Zylinska's (2023), and historical research like Dobson (2023). I have decided to not engage as thoroughly with this strand of media scholarship because of my analytical concentration on contemporary practice, I have nevertheless appreciated these studies. As an example, Paglen's (2016:1) accurate realization helps in understanding the present significance of images, not unlike how a sociologist might proceed in deciphering computer-mediated areas of social life; "[i]mages have begun to intervene in everyday life, their functions changing from representation and mediation to activations, operations, and enforcement".

What is more, the social sciences and computer science have shared some exchanges. In early computer science and engineering communities, 'social' predominantly became specified through 'social factors' which included ergonomics and flaws in human judgment (Cardon, 1998). The contemporary notion of 'human-centered AI' guiding research directions at some recognized artificial intelligence laboratories promotes emplacing computation in its human context. The contemporary AI ethics literature centers ethical principles such as data privacy, accountability, safety, security, transparency, explainability, and fairness in AI systems (Fjeld et al. 2020; Floridi et al. 2018). A position that actors tend to share, however, is that AI systems should be designed and properly geared to be inclusive of all groups in society. Efforts range from policy regulations; ethics frameworks; design procedures; data diagnostic tools; and sociotechnical solutions, such as general citizen panels or involving users directly in design practices, all the way to abolitionist critiques calling for the elimination of certain technologies (e.g., Brown, Davidovic and Hasan, 2021). AI ethics literature has a broad sociotechnical audience and has earlier counterparts in literature such as Computer Supported Cooperative Work (CSCW) and Human–Machine Interaction (HCI), which exist in the purview of computer science and artificial intelligence. The formation of CSCW, HCI, and AI ethics reveals an integrative thrust between the territory held by social and technical disciplines.

This is important to mention, in my view, because it shows that sociology is implicated in the practices of various computer science fields, even if the implication may be slight. Ethnography and ethnomethodology have, for instance, been enrolled in observing users and the accomplishments of their interactions during the development of technical systems. Institutional locations examples include Ethnographic Praxis in Industry Community, and the early corporate laboratory of Xerox. Contemporary equivalents occur across many corporate laboratories; e.g., Microsoft Research. Concepts such as data colonialism, critical race theory, and various methodologies flow into technological communities. This traffic of knowledge is significant for how we as social analysts position ourselves toward our objects of research, and the imagined distance between disciplines we bring into play.

As reflected in the interdisciplinary field of critical data and algorithm studies which formed in the late 2010s, the relentless and disorderly march of specialization continues into the present, shaping our understanding of the phenomenon of computational technology. What stands out in the literature is the entrenchment of increasingly narrow lines of research that tend to run parallel to one another. As expanded upon below, social studies of artificial intelligence relate to parts of these strands of literature, particularly lab studies and the sociology of algorithms, but more focused on the object of study.

3.3 First Wave of Social Studies of Artificial Intelligence

I next turn to one of these forked ends and distinguish two waves. The first generation is distinguished by its object of study—symbolic artificial intelligence—commonly called 'expert' or knowledge systems'. This wave consists of research tied to engineering practices of manually encoding computational capabilities. This predates the current data-centric approach of machine learning in artificial intelligence, substantialized in the second wave. Some research directions that I place in this wave include the study of computing, including earlier forms of artificial intelligence system, as a social movement (Kling, 1991) not unlike activist organizations. Another research direction is the controversy studies approach common in STS; Olazaran (1996) studied the cleavage between connectionism (neural networks) and symbolic artificial intelligence. Another direction aiming at reconciling some interdisciplinary differences was attempted, e.g., integrating symbolic interactionism and symbolic artificial intelligence (Strübinger, 1988), and thinking about ways to formalize emergent action (Star, 1995; Suchman, 1987). Key concerns include demarcation criteria relevant to the human–computer distinction, and the culture of the artificial intelligence community. I review the main positions within the distinction, and the main findings regarding cultural beliefs and practices in the community. As I show in the articles' literature reviews, some findings from lab studies have analytical value on recent artificial intelligence although their topics of interest differ.

An influential concern in the early literature is the philosophical discussion of demarcation criteria between humans and computers. Dreyfus's (1979) phenomenological critique of the computer science discourse of artificial intelligence argues that the particularities of human intelligence cannot be elicited or designed in computers without limitations, stressing embodied mind, context sensitivity, and tacit knowledge. Collin's (1990, 2018) work on artificial intelligence largely centers on the same problem, arriving at similar demarcation criteria. Woolgar (1985) realized that these different territories tend to be fluid and defined in relation to each other. As computers increase their capabilities, Woolgar argues that the 'essence' of humanity tends to be redefined in corresponding ways. In his view, establishing and maintaining the distinction between humans and technology is a social act, rather than a naturally occurring difference. In my reading of this perspective, parceling out what belongs in social or technical categories is best understood as performed within "purification processes" (Latour, 1999), and may be implicated in interactional power plays, and, in less controversial instances, become neatly and categorically ordered in socially shared stocks of knowledge.

Demarcation criteria became the topic of an evocative debate that dealt with the relationship between computers and the social, whose relevancy to the enduring notion of autonomy is pertinent. An especially contentious article was published in *Social Studies of Science*, which prompted the editor to invite a set of authors to respond. The article authored by Slezak (1989) posits that the emergence of an artificial intelligence system known as BACON, which purportedly generated scientific discoveries independently or 'autonomously', challenges the notion that scientific and technological knowledge is linked to and causally shaped by societal influences. This contention thus directly challenged the foundational theoretical premises of the strong program within the sociology of scientific knowledge (SSK), according to the author, as evident in its title, “Scientific Discovery by Computer as Empirical Refutation of the Strong Programme”. SSK was considered a diverse category at the time, not only in EPOR, but also in new literary forms, discourse analysis, and actor–network theory (see Woolgar, 1989; Collins, 1983; Barnes and Bloor, 1982). Slezak's interpretation asserts that the system existed as an autonomous entity, devoid of cultural influence, thereby producing inferences that were deemed "socially uncontaminated".

The article was met with various responses that ultimately countered the argument. In this debate, a clear position crystallization that considered the computer as instantiating an “encapsulated collectivity”; that is, it is simply equal to the human programmers that instructed it to do what it did and observed it doing as much (Collins, 1989). The computer’s inductive processing, then, is not autonomous, but a construct delegated to the computer. Another commentator, Brannigan (1989:611) writes, "[t]he imported design brings human agency in through the back door". The programmer must not only set up the program, contribute some data and code, but interpret what results amount to a scientific discovery along social standards. In Woolgar’s (1989:662) position, scientific discoveries must pass through "retrospective (and sometimes prospective) accounting practices which take place in a social context”. A similar argument from another context and other theoretical assumptions may be found in Pickering’s (1995) concept of the “dance of agency”, in which scientific processes in natural science involve humans temporarily retreating to observe nonhuman performativities at play only to then continue to intervene.

Furthermore, Woolgar (1989:661) describes a concern related to using the terms 'factors' or 'causes' when discussing ‘the social’: it is insufficient because it implies that ‘the social’ is a separate variable that may or may not influence scientific activity or may or may not be present in any action. This implication could mislead by suggesting the possibility of non-social actions. Woolgar states, following Wittgenstein, that all actions, including those related to science and technology, are performed in interpersonally reproduced “language games”.

The takeaway from this debate, in my view, is that it is worthwhile to interpret the different positions as guided by disciplinary conventions and commitments. What is clear is that it is a valid position to consider ‘autonomous’ computers not as free-standing vis-à-vis the social, but deeply wedded to it. Forsythe found herself in a stark debate which is re-told at length in (2001) on privileging the technical or the social in these settings.

Turning to the cultural and practice-oriented dimension, Forsythe's lab study focuses on an artificial intelligence lab working on so-called "expert systems" (2001). She studied the cultural beliefs that undergird practices in artificial intelligence as an emergent scientific discipline, combining experimental science and software engineering. She found that engineers' practice aimed at the formalization of behavior through mathematical systems (see also Agre, 1994, on 'capturing'). Similarly, Suchman and Trigg (1993) analyze engineers' work in developing a meeting scheduler system within a corporate "basic research" symbolic artificial intelligence lab, focusing on the significant role of the whiteboard as a workspace alongside computers. Through detailed observation, particularly of engineers' interaction, the study analyzes how engineers negotiate between the constructed reality of scheduling and computational logic. They find that the whiteboard serves as a mediating entity for shared understanding, acting as a bridge between abstract models of practical reasoning and the computer's affordances. Their work centers on knowledge representation formalisms in symbolic artificial intelligence, aimed at encoding real-world scenarios into computational symbols for logical and representational problem-solving.

Suchman and Trigg (1993:174-177) note that both artificial intelligence and social sciences take an interest in explicating practice. Attempts to model practical reasoning computationally are contrasted with social science's emphasis on the detailed analysis of specific instances of practice. They question the capability to fully capture the emergent and situated dimensions of human reasoning and interaction through computational means. Referencing Agre (1990), they critique artificial intelligence methodology for its reliance on "pseudo-narratives"; narratives of imagined practice that, while seemingly empirically developed, are retroactively engineered from technical categories to typify associated practices, a methodology distinctly divergent from the methodologies employed in the social sciences to study action in context.

3.4 Second Wave of Social Studies of Artificial Intelligence

The second generation is distinguished by its focus on machine learning. Roberge and Castelle's (2021) call, "Toward an End-to-End Sociology of 21st-Century

Machine Learning", is indicative of this concentration. Epistemic and political stakes have been at the center of this line of research (e.g., Amoore, 2023, 2020). It has been argued that machine learning contributes to structuring diverse situations in society—sorting, classifying, and enacting decisions—through data-centric probabilistic models (Fourcade and Farrell, 2023). Due to the probabilistic nature of the practice, the sociology of quantification (e.g., Espeland and Stevens, 2008), including the use of statistics in society (Hacking, 1975, 1990; Desrosières, 1988; Porter, 1995), and the strong kinship with studies of big data under the rubric of critical data studies (boyd and Crawford, 2012), and methodological tools for studying algorithms (e.g., Lee and Björklund Larsen, 2019), all gain relevancy compared to the first wave characterized by research on so-called symbolic, 'expert systems'.

Some of the first wave's focus areas have been brought to the second wave. An example includes the moving boundaries of the autonomous attribution and demarcation criteria between humans and computers. Bechmann and Bowker (2019) employ this line of reasoning to examine human involvement in the classification tasks associated with the implementation of unsupervised learning, a machine learning approach, by analyzing their work process of model development on Facebook data. They found that numerous interventions were required to administer the process of moving the emerging object along, mainly, a lot of classificatory work. Dahlin's (2023) ethnographic study of 'autonomous' technology in a medical context shows that actors structure their work activities based on modes that differently define autonomy encapsulated in situational occurrences. Another recurring example is the "encapsulated collectivity" theme, now substantialized in variations such as "machine habitus" (Airoldi, 2022) modeled on social media platforms and hiring algorithms, drawing attention to computational patterns emerging at existing lines of stratification. Another example is Olazaran's (1996) controversy study to Marres and colleagues (2024) 'super-controversy' approach.

Other recent themes include the commercialization of so-called cloud computing—i.e., computer servers placed remotely—which is central to the current industrial artificial intelligence setting (van der Vlist, Helmond and Ferrari, 2024). In a cultural register, De Seta (2023) traces the various usage of the volumetric term, 'depth', increasingly common to characterize computation. De Seta (2023:248) contrasts this with the earlier use of 'cyber' as a "domain of military activity beyond the existing arenas of land, air, sea and space" (see also

Bowker, 1993). De Seta considers the 'hidden layers' in a neural network within the regime of 'deep' learning, as well as 'deepfakes', as metaphorical; whereas other instances refer to an infrastructural usage of the term, such as 'deep-sea' cables. This extensive use shows that the linguistic structure of common cultural expressions appears, in fact, to have been deep-fried.

The second wave is distinguished by its close attention to data and computational techniques. These clusters of attention are useful in a move away from focusing on the impacts of algorithmic outputs in societal contexts to investigate the technical work that engenders them. This includes studies focused on machine learning frameworks (Mackenzie, 2017), datasets (Thylstrup, 2022), laboratory activities (Jaton, 2021), and application-centric studies on, for instance, development practices (e.g., Seaver, 2022).

Studies close to the laboratory and engineering settings relevant to this dissertation include Jaton's (2021) laboratory ethnography in which he analyzes the activities of an academic group specializing in image processing, focusing on developing algorithms for detecting prominent features in images. He shows how numerous human inputs help materials become quantifiable data that algorithms can process. He identifies narrative scenarios within which programmers operate, and how these shape the boundaries of programming episodes. He also analyzes the process of formulating, which involves converting or flattening material into datasets that align with mathematical grammar, allowing for extensive processing and for establishing a connection with certified mathematical knowledge. An important finding is 'ground-truthing', which refers to the practice of generating labels that signify and categorize data, thus creating directions for algorithms to 'learn' numerical patterns associated with certain labels. Moreover, Seaver's (2021) ethnography of machine learning and data scientists developing recommender systems shows that actors make sense of information through the framework of the concept of "data spaces". Ribes and colleagues (2019), building on Ribes' ethnographic fieldwork with academic data scientists involved in knowledge production, show how actors make sense of their reality regarding other areas of expertise primarily through the lens of "domains", which actors seek to influence and reconfigure (see also Jung, Steinberger and So, 2023)—a phenomenon that bleeds into artificial intelligence research more broadly. I tie into this line of research more in the subsequent texts.

4 Doing Computer Vision Together

This chapter develops the theoretical framework that guides this dissertation's analysis of computer vision activities. Grounded in the traditions of pragmatist and interactionist sociology, its purpose is to construct a precise analytical lens for examining scientific and engineering activity as a form of cooperative, materially-embedded work. It begins with the broad sociological concept of social worlds before moving to the more granular process of interaction and work. It then confronts the challenge of accounting for materiality within an interactionist paradigm. Finally, it proposes a Mead-informed approach to synthesize these elements into a workable, sensitizing foundation. The chapter concludes by outlining what this framework offers both for understanding computer vision research activities and for the methodological task of studying them empirically.

4.1 Science and Technology as Collective Action

This section argues for the benefits of an agnostic relationship to the concept of social worlds while hooking into a more specific concept outlined by Strauss and others which centers clearer slices of collective action.

In this theoretical lineage, a central point of departure is Hughes' notion of institutions. Countering narrow and idealized definitions of institutions, Hughes (1971) uses the term "going concerns" to describe institutions as produced by ongoing and mundane activities or what actor groups do together in practice, including science and engineering. Meanwhile, the concept of social worlds draws attention to the ways groups organize around shared activities, commitments, and "universes of discourse" (Shibutani, 1955). The concept includes considering people who at first sight seem peripheral to the core activity as contributing to the main processes of a social world (Becker, 1976:703, cited in Unruh, 1980:279). 'Going concerns' point to the persistence of institutions as forms of ordinary collective activity, and social worlds build on this by situating those activities in broader constellations of interaction, “boundary objects”, and negotiation that sustain sociotechnical and technoscientific enterprises (Clarke and Star, 2008; Star, 2010).

However, the idea of a consistent world with sharp edges, whether a social world or a computing subworld, is challenging to propose as a meaningful analytical unit. Inhabitants of a social world may bear multiple memberships, shifting in strength, and have various overlaps with other worlds, and subworlds splinter (Kling, 1978; Unruh, 1980). The flexibility of the social worlds concept, while a strength, therefore presents an analytical liability for this study. Its fluid and potentially limitless boundaries make it difficult to define a crisp unit of analysis. As Unruh (1980) asked, how can participation in such a loosely organized phenomenon be considered a form of social organization? Attempting to map the full and sprawling contours of the contemporary social world of artificial intelligence is a task that would need to account for a vast ecosystem of academic labs, corporate research groups, open-source communities, and the often-invisible "ghost workers" who annotate data (Gray and Suri, 2019; Sambasivan et al. 2021), which risks analytical diffusion. The social worlds of computer vision are therefore better understood as multiple. Therefore, rather than attempting a comprehensive cartography (e.g., Clarke, Washburn, and Friese, 2022:103), this dissertation adopts a Straussian focus on collective action (Strauss, 1993). For Strauss (1993:223), a social world is not a static social structure or container, but is more productively understood as "a recognizable form of collective action" (see Becker, Geer, Hughes, and Strauss, 1961). This shifts the analytical focus from defining membership and demarcating boundaries across multiple worlds to more specific, coordinated activities and processes that constitute clearer slices of empirical settings, viable for research activities.

4.2 Interaction and Work as Core Processes

With collective action established as the central concern, interaction becomes the primary analytical process. This strand of pragmatist sociology of science and technology shares territory with research programs like the construction and shaping of technology. My approach follows the lineage of interactionism, understood here as a tradition that has developed along several theoretical strands. Atkinson and Housley (2003:37) distinguish between approaches focusing on face-to-face interaction (e.g., Goffman, 1959) and those examining the organized activities of actors within particular social worlds (e.g., Strauss, 1993). In the pragmatist tradition that informs this study, interaction refers to “people doing things together or with respect to one another in regard to a phenomenon” (Becker, 1986). It thus includes the “action, talk, and thought processes that accompany the doing of those things” (Strauss and Corbin, 1990:162; Strauss, 1993). Positioned in this lineage, my analysis centers on computer-science-oriented activities. These activities mobilize multiple forms of knowing that cut across scientific and engineering lines of reasoning. To understand how such interactional processes take shape in these settings, I extend this interactionist orientation through science and technology studies. This move enables closer attention to the material and epistemic arrangements that organize interaction in computational processes.

To clarify why this dissertation does not proceed by reconstructing logical structures or treating science as an exclusively cognitive enterprise, it is necessary to revisit the dominant conceptual framework often referred to as the “received” view (Woolgar, 1988). The received view defines science and technology as enterprises grounded in logical structures, and the material applications of such enterprises. Sourced from the philosophy of science, it draws on experimentation as part of justifying beliefs and positions objectivity as the main mode of engagement (cf. Daston and Galison, 2007). As Pickering (1992:689) notes, this conception faced sustained criticism beginning in the 1950s by Kuhn, Feyerabend, Quine, and others, who showed that “considerations of the underdetermination of theory by evidence, the theory-ladenness of observation and incommensurability…all contrived to suggest that, in one way or another, the search for a logic that could satisfactorily explicate important instances of theory-choice was futile”. These elaborations left the received view in an epistemic deficit: there is thus no justification in sight to approach science and technology as a purely cognitive or logical enterprise, in which rationality reigns in isolation (Woolgar, 1988; Bloor, 1984).

In parallel, Zwart (2022), building on Vincenti’s (1990) thesis, rejects the claim that engineering is the application of science. Instead, he describes a heterogeneous engineering epistemology that includes scientific principles, empirical experimentation, functional descriptions, computational models, know-how, external collaboration, and problem-solving strategies. For my purposes, the significance of Zwart’s perspective is not to catalogue complexity but to underline that the boundaries between computer science and engineering are convoluted, and that approaches focused on reconstructing logical structures are inadequate. Lynch (1991:74) makes a related argument, urging attention be paid to “contextures of practice and equipment” rather than abstract “knowledge”. Scientists, he notes, “construct and use instruments, modify specimen materials, write articles and make pictures, and build organizations” (ibid.). These practices unfold within and reshape material environments. Knowledge, then, is not only what scientists and engineers know but what they do; how they enact their knowledge in situated settings.

Even when the practices at stake involve highly complex and specialized forms of knowing, a knowledge-as-action perspective foregrounds interaction as the primary site of inquiry. The central task is therefore to understand how scientists work, not only the logics they invoke. This web of ongoing interaction, such as talk, gesture, thought, and physical manipulation, is where conventions are forged, instruments configured, meanings negotiated, and knowledge produced (cf. Strauss and Corbin, 1990:162; Strauss, 1993). This leads to the foundational premise: science is "just another kind of work" (Clarke and Star, 2008), a ‘going concern’ that, like many others, involves assembling resources, negotiating divisions of labor, and stabilizing practices through ongoing effort. Clarke and Gerson (1990:180) describe this lineage accordingly:

They focus on science as work rather than science as “knowledge,” refusing to divorce knowledge from interaction and social organization. They have not been concerned with selves or individuals, but instead with all other scales of work organization from research projects to laboratories to disciplines to political and economic relations in the wider society.

Knowledge production therefore takes form through action, such as designing, optimizing, coordinating, and sustaining computational processes. To provide further precision, this framework is oriented to the specific form of interaction. For instance, work encompasses the interactional processes of "working things out" which includes the crucial efforts of articulation and coordination necessary for different lines of activity to mesh (Strauss, 1993:52, 87; Strauss, 1985). This includes the often "invisible work", meaning the interstitial activities that resolve ambiguities and maintain continuity in shared projects (Strauss, 1985; Star, 1989).

Scientific facts and technological artifacts are not entities awaiting discovery (Woolgar, 1988) nor are they governed by a determinist force in which the technology with the strongest technical performance by default wins the race (Mackenzie and Wajcman, 1999:33); they are the emergent and provisional outcomes of interaction. Recognizing that these phenomena arise from collective action does not negate their constructed character, which is sometimes conflated. Star (1985:202) acknowledges the seemingly solidified nature—the "sturdy complexity"—of science and technology, thus countering an idealist stance sometimes associated with social constructionism, where an individual can supposedly reverse their order. As Star (1985:201) puts it, echoing Hughes, "To say... that 'it could have been otherwise' is not to say that it is [easily undone]".

It requires practical, local action to develop, sustain, and spread the materials of science and technology (Latour 1988:227): effectively renegotiating local orders (Strauss, 1993). Once materially settled in place, these appear as if they were solid from the beginning (cf. Latour, 1987).

An exchange between Mol and Clarke and Star helps elaborate this tension between malleability and solidity. Mol's (2002) work argues for ontological multiplicity: objects such as diseases are enacted differently in various practices—through diagnostic tests, clinical consultations, or surgical interventions. Clarke and Star (2008) argue that the interactionist tradition already attend to difference and heterogeneity and show how actors manage to cooperate without fully resolving divergence: stressing the point that cooperation can and does occur without consensus (Clarke and Star, 2008:122–123, n.8). They contend that Mol misreads "perspective" as merely seeing the same thing differently, whereas interactionists recognize that perspectives generate different things in practice (ibid.). In this view, scientific and technological objects become solid and robust because actors coordinate and move forward, despite differences following their varied local doings.

4.3 The Lasting Problem of Materiality

A framework focused solely on interpersonal interaction remains incomplete. This argument has been powerfully introduced in posthumanist scholarship, particularly actor–network theory, which insists that nonhuman and constructed objects—from scientific instruments and laboratory reagents to algorithms and computer code—are active participants that fundamentally shape events (Latour, 2005; Cerulo, 2009; Bennett, 2010). Such posthuman arguments are a central concern in social theory, which in their genealogical variety aim to decenter humans to include objects in sociological analyses. To ignore the agential capacity of material objects is to miss an important dimension of how science and engineering are practically accomplished and how such products are sustained or stabilized: the infrastructure behind why certain facts and artifacts defy being otherwise.

The concern is that defining interaction only in interpersonal terms reinforces a constructionist idealism (Puddephatt, 2005:360). This idealism reproduces a longstanding sociological bias wherein the material world is treated as a passive backdrop, a mere prop or screen onto which humans project meaning and social structure. Latour (1999:176) compares this stance to the National Rifle Association's slogan, "Guns don't kill people, people do". This neglect of materiality narrows sociology to tracing meanings—all the way down. As Puddephatt (2005:360) writes, this leaves "nature without a voice" and downplays material interactions analytically; for instance, a problem in how the Social Construction of Technology, or SCOT, primarily traces human meanings in its approach to studying technology formation.

A central theoretical challenge, then, is to incorporate the active role of objects into an interactionist framework without dissolving a robust social psychology and approach to social processes. The goal is not to attribute intentionality to machines in recognizing them as active, but to develop an account for a hybrid world where humans and nonhumans are co-constitutive in action. This requires moving beyond the limitations of a strictly human-centric interactionism to find resources for theorizing a more-than-human sociality.

Fine and Tavory (2019:458) argue that "interactionism must develop some of its core tenets" to better attend to "a world that questions human exceptionalism, bringing nonhuman actors into the shaping of meaning". They (Fine and Tavory, 2019:460) suggest that "the growing literature on nonhuman interactants and materiality makes this point persuasively". Cefaï (2016:176) too, suggests that "Mead's ecology helps us rethink the place of objects, a task that was neglected by symbolic interactionism". In Joas and Huebner (2016), Brewster and Puddephatt (2016) articulate a strong claim: "recognizing Mead's originally intended concept of the social makes Bruno Latour's (2005) mission of 'reconstructing the social' to include nonhuman relations far less revolutionary". Joas and Huebner (2016:6) read their chapter as "an important contribution to the rethinking of the place of objects in social theory". Furthermore, Joas and Huebner (2016:7) state, "They clearly find anticipated in Mead what is presently debated as a new view of the social—that is, a view that includes non-humans".

Meanwhile, similar progress on this issue is found in interactionist science studies. Strübing (1998:451-452) notes that "interactionists and advocates of actor–network theory established a productive discourse resulting in some degree of approximation of both viewpoints (Star and Griesemer, 1989; Fujimura 1992, 1991; Callon 1994; Bowker and Star 1996)". Strübing (1998:451-452) states that "[t]he recent claim by some interactionists that technical artifacts are capable of action, far from being eccentric, turns out to be both obvious and productive (Clarke and Gerson 1992)". Henderson (1991) argues along similar lines, drawing an explicit connection to the interactionist conception of 'situation'. Clarke and Gerson (1990:198) state that Latour and colleagues "challenged interactionist thinking in significant ways",

> they insist, correctly we believe, that we view all participants in a setting as actors, not just humans. For example, microbes were major actors in the rise of the germ theory of disease and the Pasteur Institute. Door-closers are actors in both scientific and non-scientific contexts. This point is an important extension to basic interactionist principles and ties to issues of meaning and action which Mead (1932) explored philosophically.

They, however, abstain from specifying how it relates to Meadian concerns of meaning and action, which is why I take on this task below. It would be misleading to suggest that Mead anticipated the role of non-humans or that his theses align fully with later theories like actor–network theory; for instance, their insistence on troubling scientific categories that Mead accepts. Differences exist in the rhizomatic approach to agencies in which objects achieve stability only through their connections (Latour, 1999) compared with Star's (1995, 2016) ecological approach rooted in interactionist social worlds theory, but the difference is primarily a matter of analytical style, rather than cosmology.

4.4 Cooperative Action with Objects

A reading of Mead from this perspective provides a useful, and perhaps surprising, resource for theorizing such hybrid worlds. While often associated with a symbolic, human-centric sociology that draws out shared understandings and meaning (e.g., Blumer, 1969), Mead's (1938) philosophical project, especially in *Philosophy of the Act*, articulates a processual and non-dualist ontology and a clearer stance on material organization. Commentators suggest that Mead outlines his “cosmological views” more thoroughly in this work (Morris, xlv, in Mead, 1938). Mead engages at length with the dynamics that make up the organism's environment, and introduces several theses on its significance: dynamics that tend to be called non-human agency in contemporary scholarship. The use of the term ‘agency’ is contested in many accounts, meanwhile, posthumanist scholars often operate with a minimal definition: anything that affects a change, pressure, etc., which as we will see is quite similar to Mead’s assertions, like how ‘emergence’ draws attention to shifting constellations of materiality that in turn engender ontological novelties.

Puddephatt (2005) shows how Mead accounts for meso-level material organization shaping parts of social life. For Mead, consciousness is inseparable from the social and material environment, making objects central to meaning production (Puddephatt, 2005:359). In this view, the world is composed not of inert matter, but of resistant objects that are in a dynamic, ongoing, and cooperative relationship with human actors (cf. Bennett, 2010). Together with Joas (1985), Puddephatt thereby frames meaning as emerging within a “trifold matrix” of human and material interaction. This leads Puddephatt to state that “Mead was never modern”, commenting back on Latour’s (1993) slogan, “We have never been modern”. Mead can thus be interpreted as companion along with posthuman insights, even if this includes centering a “weird” Mead (cf. Kohn, 2013; see also Fine and Kleinman, 1986).

A complex technical artifact like a software program can be understood as an "encapsulated collectivity" (Collins, 1989). It is the crystallized result of former collective action, embedding the intentions, logic, and symbolic structures of its human creators. Therefore, to interact with a computer is to engage in a cooperative, materially-mediated relationship with a sedimented history of previous action (cf. Strauss, 1993). Qualifying this reasoning, Agre (1994:745) addresses the informational content held in computer’s states and data structures in a software program, which in his example is attributed to mirroring some aspect in an organization: “This assumption does not, strictly speaking, derive from any inherent property of computers. It is, rather, a theory of representation that is embedded in the way that computers have customarily been used”. The issue, then, is that when following this orientation, it is easy to stick with tracing those human processes at the cost of accounting for materiality's shaping of action. Brannigan (1989:611) writes, "The imported design brings human agency in through the back door". This creates a slippage: treating materiality, and its "affordances", solely as past human action. As McDonnell (2023:204) puts it, "the easiest-to-swallow argument for sociologists is that objects have agency by standing in for humans". Puddephatt (2005:375) raises a similar concern and stresses the socio-material origins of computer-generated symbols to overcome such idealist reduction.

As a remedy to this slippage, Mead's concept of the act provides an understanding for how objects participate in action. Within the "manipulation phase" of the act, an object provides a "contact response", meaning its resistance, pressure, texture, and responsiveness that actively shape the course of action (Mead, 1938:195, 218; Lewis, 1981:130). Mead's distinctive contribution becomes apparent in the manipulation phase, where objects enter into the act as active entities. The concept of 'contact response' is central here, as the completion of an act depends on the interaction between the actor and the object (Mead, 1938:195, 218). He asserts that "[t]he object itself is the locus of pressure" which is 'transferred' to the human as the act is performed (ibid.). It is not a one-sided manipulation, but a process in which the actor and object is mutually active, shaping the course of action. Chang (2004:409) argues for a similar interpretation, "'The organism (an animal, a human individual, a social group, etc.) determines the environment as fully as the environment determines the organ[ism],' said Mead (1934:129)". In a striking formulation, Mead (1938:109) argues, "the physical object must literally do as much as we do if we do anything", an insight that predates and enriches similar arguments on symmetry in contemporary posthumanist and practice theory (see Turner, 1994). This human–object relationship is asymmetrical; an object does not "take the role of the other" in a self-reflective, symbolic manner (Mead, 1938:108-109; Strübing, 1998:451-452). However, through this engagement, the object presents a "gesture"—a non-symbolic but agential response—that the human actor must in turn address, generating meaning through a mutual, material, and interactive process (Mead, 1938:310-311).

Mead (1962:185) also illustrates this with the example of an engineer building a bridge who is "talking to nature", meeting its stresses and strains with new actions, to which nature "comes back with other responses". This object gesture is conceptually similar to the idea of non-linguistic, non-discursive material signification by sociologists within the actor–network tradition (Callon and Law, 1995:502-503). Even if Mead's use of metaphor was rhetorical, it points to material changes shaping action; otherwise, he could have remained on the anthropocentric carousel, where meaning is reduced to interpersonal and symbolic negotiation alone. Mead's (1938) ideas of the manipulation area, gestures, and emergence provide support for understanding interaction with dynamic objects grounded in a robust social psychology (Mead, 1962; Puddephatt, 2017). It is accurate that actors interact with materiality and encounter emergent resistance, occasionally blocking action and prompting subsequent ones in variable ways (Strauss, 1993; see also Pickering, 1995). The exclusive focus on meaning in classical symbolic interactionism seems more an analytical choice than a theoretical or ontological deficit.

The Meadian foundation in this interactionist orientation thus harmonizes with later developments, e.g., object 'performativity' (Pickering, 1995) or 'affordances' (e.g., Gibson, 1979; Latour, 1995a; Davies, 2020; Davies, 2023). This retains a sensibility analytically committed to materiality and to knowledge infrastructures (Star, 1995; Bowker, 2016) within an action-centric perspective (cf. Strübing, 2017). Accordingly, I situate computational tools, human–computer interfaces, and artificial neural networks as analyzable elements in interactions occurring within broader research apparatuses and attend to how they come to matter in situated settings.

4.5 Analytical and Methodological Directions

By following this path from social worlds to materially-embedded interaction, we arrive at a sensitizing framework (Blumer, 1954) for the empirical study of contemporary science and technology that I bring to bear on the empirical setting. Conceptualizing science and engineering through interactionist and pragmatist theory means eliciting attention to the collective, materially embedded, and ongoing processes of 'working things out' within institutionally located social and technical arrangements. It provides the foundation for understanding science and engineering as ongoing work. This offers two primary contributions to the dissertation.

In short, the dissertation provides a coherent ontological understanding of the research setting *as a hybrid world* and *orients the analysis to interactional processes*. I approach computer vision research activities as a hybrid in which outcomes arise from cooperation between researchers and active objects, computational infrastructures, and research apparatuses. These are material arrangements that shape action in the concrete work of training models and preparing datasets and so on. The focus is on the situated work—the "working things out"—through which this human–material collective builds, sustains, and troubles computer vision components and systems.

Directing my attention to situated work, the analysis entails elaborating on concepts in close connection to the empirical materials. These concepts, developed close to the empirical material, are particularly important for the dissertation, which merits a brief review. I specifically use Strauss (1993) as an analytical toolbox, and engage with extant research findings.

I use the concept of 'standards' (Lampland and Star, 2008) as devices evoking temporary islands of stability: interpersonal agreements on how to proceed with subsequent action, stabilized by protocols like annotator guidelines and interactive documents and user interfaces—in this case, promoting specific action possibilities—and the alignment work required in their role as middle acts, enabling such work to move along (Strauss, 1993).

Key conceptual developments (Engdahl 2024a, 2024b, 2025) include assembling computational models in what I term pipeline welding, setting conditions for model learning in delegated construction, interpreting model behavior in awareness contexts, coordinating and supporting standards in alignment work, adapting models in practice in trajectories, and mobilizing credibility and networks to handle uncertainty in social license.

Specifically, pipeline welding describes the process of assembling and integrating computational components within complex technical infrastructures. This involves specifying architectural modules, selecting and preprocessing data, determining optimization routines, and calibrating parameters: work that requires explicating, dividing, and coordinating tasks across computational workflows. Delegated construction characterizes the distinctive form of knowledge production in deep learning, where researchers configure conditions for models to learn rather than directly programming their behavior (cf. Jaton, 2021). The primary outputs of this process are 'model weights' or internal representations that encapsulate learned knowledge while remaining opaque to their creators.

In connection to this, I draw on the concept of "awareness contexts" coined by Glaser and Strauss (1965) and reinterpret it accordingly. This concept draws attention to how differential access to information (who knows what, when, and how) structures interaction. It is worth noting the concept's origins: Strauss and colleagues dealt with this concept in a wholly different setting compared to the present one; namely, to conceive how differently located persons access information that is known to some but collectively undisclosed during the terminal state of life, performed in efforts to protect a person's insight into how dire their medical condition is. Originating in studies of medical work, reinterpreting this concept from terminal states to computer terminals, the commonality, I argue, is how people relate to opaque conditions, whether a disease or a model. It draws attention to interactionally evoked contexts in which actors gain knowledge about opaque processes. As we will see, I develop awareness contexts and use it to refer to the shared processes through which researchers monitor and make sense of model behavior, including actions such as monitoring loss curves, visualizing outputs, and performing qualitative inspection. These awareness contexts are relationally evoked and sustained, and enable scientists to assess when models are functioning within expected bounds and to identify points of divergence.

Alignment work describes the ongoing efforts to forge agreements and ensure that protocols and standards are consistently applied during dataset construction and model evaluation. This work extends from high-level protocol development to moment-by-moment decisions about data capture and annotation. Implementation trajectories characterize the non-linear, negotiated processes through which pre-trained models are adapted for specific local applications. These trajectories are contingent on technical, infrastructural, and regulatory conditions that require ongoing navigation and adjustment. Social license describes how actors leverage their expertise, credibility, and networks to navigate unforeseen contingencies during implementation trajectories. This concept directs attention to the social resources that enable technical work to proceed despite uncertainty and complexity. Together, these concepts show how research and development outcomes emerge through situated, negotiated, and materially-active processes.

Drawing on these assumptions in my empirical engagement, I also engage with second-order concepts, from emic (e.g., "foundation model") to etic (e.g., "sociotechnical") primarily as translation devices for situating my research for interdisciplinary audiences.

Methodologically, while retaining human primacy in analysis, I relax the boundary between humans and objects and use interaction as the root concept in the socio-material area where such boundaries are difficult to sustain: a category that Latour's (1999; 2005) calls the "middle kingdom". Analysts often lean on expert definitions (Collins and Yearley, 1992), hybridity (Callon and Latour, 1992), juxtaposition (Mol, 2024), or speculative modes of reasoning. I follow how actors define situations (Thomas and Thomas, 1928:571–572) and place their definitions within the material arrangements that enable them: infrastructures such as datasets, sensors, and tooling. I direct attention to how those definitions function in practice, what material support makes them possible, and how they become active through work. This does not imply humans are the only active force. Instead, I draw attention to the socio-material aspects involved in computer vision research practices and trace how outcomes are worked out in situated settings, rather than through single-cause explanations, such as a specific set of privileged meanings.

5 Ethnographic Approach

This chapter explains how my research was constructed and performed. My methodological approach is shaped by several orientations that occasionally overlap. The ethnographic tradition provides a central foundation, particularly lab studies' (e.g., Latour and Woolgar, 1979; Lynch, 1985) immersion in scientific and technological sites, and 'multi-sited' research approaches (Marcus, 1995). Pink and Morgan (2013) describe ethnography as relying on sometimes intrusive forms of engagement, and point to the importance of researchers making their presence and intentions clear from the outset. Beaulieu et al. (2007:679) suggest that lab ethnographers rethink traditional assumptions about physical presence, instead considering how research sites can take on variable scales and mediated traces of action. Recent methodological developments also inform my method, including techniques discussed by Emerson, Fretz, and Shaw (2011) and approaches associated with digital ethnography (Postill and Pink, 2012), 'trace' (Geiger and Ribes, 2011) ethnography, and strategies for locating patchy phenomena (Seaver, 2021). Moreover, I draw on grounded theory, particularly the constructivist strand developed by Charmaz (2014), which extends the foundational approach (Glaser and Strauss, 1967; Strauss and Corbin, 1994), by renewing tools for addressing epistemological challenges associated with these earlier approaches. This analytical tradition, with its strong ties to fieldwork and participant observation (e.g., Blumer, 1969), maintains a pragmatist and reflexive orientation. These methodological influences provided guidance for my method in ways that remain sensitive to both ethnographic principles and contemporary developments in studying digital and multi-sited phenomena. Key concerns that run through the discussions in this chapter is how the analyst is implicated in how fieldsites are conceptualized, how data are produced, and how partial and empirically-grounded analyses are constructed.

5.1 Overview of Data

Over the past four years and with varying levels of intensity, I have conducted ethnographic fieldwork within the computer vision community. My multi-method approach combines direct observation with ongoing engagement in digital spaces. As part of this research, I have conducted 36 interviews with some follow-up interviews, in addition to many chats, traces of which ended up in my fieldnotes. I have observed work in academic laboratories, attended multiple conferences, and observed both online and in-person meetings. I have participated in online communication platforms and collaborative work platforms. The primary prism through which my observations focused was on technological work. Beyond observation, I have analyzed archival and digital materials as key artifacts of the community. These include technical manuals, scientific papers, and informal documents, such as documentation. Additionally, I have examined computational artifacts or components central to their practice, including models and datasets. To remain informed about developments in the field, I have followed newsletters, podcasts, and social media accounts popular in the field.

5.2 Field-making

In this section, I outline my research journey and how I assembled a workable field. I was hired as a doctoral candidate in the multi-staffed and multi-sited ethnographic research project, "Show and Tell", funded by the European Research Commission via the Department of Sociology at Lund University. This position strengthened my institutional legitimacy and economic resources, helped render my research legible to scientists, and my project collaborators helped me navigate the research process. Project-shared concerns, alongside theoretical discussions, shaped assumptions during my field-making.

The project's central focus is the use and application of three-dimensional reconstruction techniques (like photogrammetry, LiDAR, and Structure from Motion (SfM) in different actor groups, particularly archaeologists and forensic scientists. Considering the level of accuracy required in these sensitive sites, compared with something like a digital twin for animation or e-commerce purposes, a central concern surfacing in the project was the reliability of capturing meaning in instances of data collection by practitioners on the ground, manual malleability in their reconstruction work, and their data archiving practices. My inquiries into these techniques led me into their histories: photogrammetry's pre-digital past, LiDAR's introduction in the 1960s, and the formulation of the SfM algorithm at MIT's AI Lab in 1979. Instead of pursuing an archive-oriented historical reconstruction, I chose to investigate their contemporary context. Building on the collective effort of "Show and Tell", I aimed to extend our ethnographic sites to computer vision research, which is the techniques' disciplinary home, to understand how they emerge and gain legitimacy in the present. This involved expanding the project's field sites beyond the established sites of archaeology and forensics to include the labs in which computer vision techniques are actively designed, debated, and tested. I aimed to contribute ethnographic accounts detailing the daily work and decision-making involved in this shaping and stabilization, providing links between the computer vision research observed and its application in the archaeology and forensics contexts researched more directly by my project colleagues. I wanted to provide ethnographic entry points that could deepen our understanding of how contemporary computer vision researchers shape and stabilize these techniques, knowing from my project collaborators that their influence extends far beyond initial application areas. This pivot towards contemporary research and development practice appeared significant from my perspective, given the major changes the field had undergone with the recent success of deep learning, allowing analysis of current processes.

This ultimately led to specific analyses, such as examining the evaluation mechanism of benchmarks internal to the community and the work involved in that practice. As I developed these sites, I also grew an interest in analyzing other facets of the research and development work I observed, although it took time to feel confident pursuing this direction that was more attuned to my interests. This positioning allowed me to bring insights back to the project team regarding the field's assumptions, shifts (like towards deep learning and generative models), and specific technical vernacular (around concepts like 'models', 'algorithms', 'priors', 'ground-truth', and 'benchmarking'). Concurrently, my approach benefited from the collective expertise within the project, particularly during discussions on organizing, sharing, and analyzing the rich qualitative data emerging from the computer vision labs alongside data from the other fieldsites as well as regular collective analysis sessions on data from different theoretical standpoints.

In practice, this meant that I spent significant time searching for potential participants through online networking platforms like LinkedIn and attended various academic and industry events relevant to computer vision. I used a purposive sampling strategy, reaching out to active scientists with job titles like "Machine Learning Scientist" at "Vision Lab X", "Computer Vision Scientist" at "Artificial Intelligence Lab Y" and searched for professionals with degrees in image processing and analysis, and computer vision in both academic and industrial organizations, identifying possibly informative, typical actors.

I began subscribing to newsletters relevant to the community and searched for events and ongoing work projects. The inclusion criteria began with a more principled and broad-sweeping curiosity. The early stage of my research process included following all things 'neural' in computer vision. While this criterion entertained my interests, it became unwieldy for concentrated research. Over time, I partook in computer vision and artificial intelligence conferences, as well as other gatherings, like networking events. These efforts proved productive in developing ethnographic relationships, as it provided me with opportunities to recruit more participants face-to-face. I made some demarcations in terms of inclusion criteria for participation in the study, focusing particularly on active scientists working with computer vision research. As my material and understanding developed, I turned to a more targeted sampling procedure, seeking out scientists involved in specific types of work, like model work and data work, respectively. My workflow coalesced into inviting participants to an interview, where I asked whether I could visit their workplace, or partake in any of their upcoming meetings, developing my ethnographic relationships accordingly.

As my fieldwork took shape, that which we broadly call 'artificial intelligence' grew horizontally across new areas and vertically by intensifying research efforts in society. This involved managing a reality rush of sorts, where my object of study appeared to free itself from my hands, presenting an overwhelming number of possible entry points and making it difficult to decide where to begin and what to follow. Many paths promised insight, forcing me to make choices about what to prioritize and what to leave aside, in mitigating risking of a loss of coherence. Beyond a pure methodological challenge, complying with norms in my research field, this expansion seemed characteristic to artificial intelligence as it is practiced, drawing and trading upon the existing infrastructures and institutions already sustained in datafied, digital, and algorithmic societies. To move beyond the abstraction associated with the ever-shifting concept of 'AI', I focused on specific and tangible research and development projects, examining how artificial intelligence takes form through concrete computer vision initiatives within institutional settings—a more manageable and analytically fruitful approach. From a reflexive perspective, this entails foregrounding academic research in my field-making, which in my view, is a plausible approach in line with the research aim.

Learning from the extant research designs of lab ethnographies, and the 'patchwork' approach recently espoused by ethnographers (e.g., Seaver, 2021), I decided to be less oriented to maintaining organizations as the core anchor, instead traversing many sites and research projects while learning about their work practices. Rather than only focusing on a single lab group, creating the conditions for comparison between multiple sites helped contextualize the practices in several. My fieldwork experience also helped me develop new connections and sites, enriching my access to the field. A drawback was the extensive exposure to unfamiliar research projects and terminology. Relatedly, ethnographers stress caution regarding the "fine line between site hopping and site staying" (Prasad and Shadnam, 2023:850), which is why I foreground some sites analytically, detailing my engagement to these sites more extensively, while recognizing the influence of the other sites on my research process and understanding, not least in shaping my understanding of the diversity of work in the field.

5.3 Sites and Materials

I foreground the following sites. This selection is based on two factors: their centrality to computer vision processes and the richness of the empirical material.

Table 1. Foregrounded Sites

Research Setting	Locations	Team Structure	Computer Vision Focus	Primary Research Materials
Academic Research Lab	Academic institution A (USA)	Single research team	Three-dimensional reconstruction methods; transformers; generative models	Interviews; on-site observations of daily work; work meetings; work communication platform; computational artifacts; scientific publications; technical documentation
Consortium	Academic institutions in multiple countries, including Academic institution A (USA) and Big Tech Corporation	Distributed across various research teams	Benchmark dataset of first-person recordings of everyday activities	Interviews; work meetings; trouble-shooting sessions; computational artifacts; scientific publications; technical documentation (e.g., data annotation guidelines); online forum
Three Academic Research Labs	Academic institution C and E (Sweden), Academic institution B (USA)	Three separate research teams	Object detection systems for bird analysis (C), car detection (B), and radiation therapy (E)	Interviews, conference presentations; work meetings; computational artifacts; scientific publications; technical documentation

A central part of my fieldwork occurred in a specific academic laboratory, where I spent a month observing daily work activities on-site. At this lab, I attended project meetings, observing researchers collaborating on multiple computer vision projects. These meetings gave me insight into how they structured their work, probed problems, and made design choices. In addition to observing, I conducted interviews to learn more about their work, often asking participants to compare different research approaches within the lab. I took on a passive role when observing their interactions, but engaged more actively by prompting discussions during interviews. Prior to the on-site visit, I had engaged with the research conducted at the lab over several months, as well as after the visit. As in the other sites, I conducted digital observations by engaging with computational artifacts, like rendered images, blueprints of model architectures, and model outputs (e.g., bounding boxes, masks, labels, depth maps), for instance, visualized as bounding boxes around detected objects in an image. This engagement included reading technical papers of their methods, observing presentations, interviewing people working with similar methods, and developing my understanding of computer vision as a discipline, in efforts to strengthen my readiness of understanding their work.

Another part traced the development over 14 months of a consortium's construction of a dataset project, conducting interviews and analyzing digital traces. I encountered this dataset project in one of the many news lists on artificial intelligence that I had begun subscribing to early on in my research process. The entry communicated the dataset project conducted by a large group consisting of scientists located at a large tech corporation and several universities across the world. This dataset project and its distributed and temporary organization became the first fieldsite on which my ethnographic efforts became concentrated in a temporal perspective.

To construct a coherent fieldsite related to this, I pieced together digital entry points (Postill and Pink, 2012). The project's main website provided an overview of the dataset, access procedures, and a discussion forum where researchers exchanged insights and addressed challenges. Its GitHub repository documented code updates, worklogs, and issue reports, revealing ongoing modifications and problem-solving strategies. By tracing these interactions ethnographically, I gained insight into how practitioners negotiate data quality, annotation standards, and benchmarking processes. This aspect of my research aligns with laboratory ethnography but extends beyond physically bounded spaces to include online forums, GitHub repositories, recorded workshops, publicly available documentation, and computational artifacts. I introduced myself in the forum associated with the dataset project, reached out to key contributors to seek their consent, and secured formal access by signing a data license agreement. For more detailed description of the fieldsite, see Article 2. What is worth adding to the existing description is that this digitally mediated and distributed organization involved sourcing key participants for interviews through digital observations, and retrieving informed consent from those participants in particular—through the group as a whole through their shared project email, although likely not including every actor participating in the project.

The third site grew out of my meeting observations. My contact with site C originated with one person who had several affiliations: an academic computer vision group, and a national artificial intelligence center. In this sense, he acted not only as a gatekeeper in an organization, but more of a door opener to a looser and more personally held network. This site thus initially hinged on this person's work and network. My contact with him led to exposure to other scientists and practitioners. His group also arranged a national conference in the field which I participated in. My engagements at here involved spending time with participants under informal terms, such as lunch, after work, and dinner. I observed numerous presentations of work at his workplace and the conference. Some of the scientists that I met through this person became incorporated into my study after I acquired their consent. My contact with site B and E developed similarly from bigger meetings: I met them in person and built my engagement with them and their work from there, like interviewing and taking part of their work artifacts and documentation.

Taken together, the other sites involved fieldwork that deviate from the norms of thematic cohesiveness, and the conciseness required in scientific articles. Meanwhile, they are testifying to the process of artificial intelligence becoming ubiquitous and commercialized. For instance, I have traces spanning vivid scenes in my fieldnotes, like when I had closed up shop at an STS conference boasting more AI-focused sessions than I could attend, and sought safe refuge by taking a boat tour in Amsterdam, only to see ads from companies promoting their recent "AI-powered" invention on other boats in their busy canals. Another snippet chronicles my journey by car to an event in San Francisco on a Saturday morning as my fieldwork began earlier than I expected. "I talked to family back home on the phone for a bit, but as we got closer to the city, I became more and more distracted. I noticed the large billboards that accompany the highway into and partly through the city. I couldn't help but notice that each and every ad contained 'AI'. Out of the 6 big billboards I noticed along the highway, every billboard somehow made claims related to 'AI'". Or, when I would attend social gatherings and receive questions from curious acquaintances, sometimes defining me as part of the computational collective driving transformations they had encountered in the news.

While these segments of my material evoke several sociological themes, I decided to direct my attention to some of the sites that required the most labor and ethnographic relationship-building to access. In addition to the norms of thickness and cohesiveness I have been navigating, my motivation was to make use of my positionality and work to the benefit of it.

To give a sense of these engagements, I provide examples next. I developed an ethnographic relationship with a person that I found on a digital work platform and I eventually interviewed him at his work site several times and participated in their seminars. We had meetings which included my sitting along with him as he worked on his computer. Sometimes he would be quiet for a bit and sum up what had happened on the screen from his point of view every now and then. Sometimes he engaged in 'live reporting' where he tried to describe what he was doing while doing it. His group had an ongoing seminar series in which external scientists would present their work. I routinely contacted incoming scientists, acquired their consent for my observations of their talks, and incorporated my observations of these episodes into my dataset. These were crucial moments in my ethnographic journey, like instances of other fieldwork visits to companies, and it is unfortunate these did not find their way more explicitly into one of the publications. They were highly informative, particularly the mundane dimension of programming, exposure to common terminology and ways of working, and the richness of the field, and provided me with coordinates to navigate other sites.

I have, moreover, participated in leading conferences in the field, and attended many networking meetings, a hackathon, attended talks by tech moguls, and went on guided tours at computing museums. Similarly, I have participated in academic workshops and courses in python programming focusing on computer vision techniques, and completed non-academic online courses on machine learning. Other cases include my daily monitoring of research artifacts, papers, code repositories, technical reports, niche forums and online channels. Moreover, I have engaged various computer vision techniques accessible online, analyzing images and generating videos (see de Seta, 2024, for a discussion on making "latent space" into ethnographic material; de Seta, Pohjonen, and Knuutila, 2024).

5.4 Producing Data

I produced research materials such as interview transcripts through qualitative interviews (see Table 2) and fieldnotes through ethnographic observations (see Table 3) with varying degrees of participation (Strauss and Corbin, 1990). I conducted the interviews by preparing themes before each session, including open topics such as 'current work project', 'data annotation', 'experiments', 'evaluation', 'resources', 'the field', and so forth. These themes were subject to change over time, reflecting my analytical interest at the time, and the person I interviewed. Some of the interviews, about 8, were conforming to a norm of 30-minute-long sessions common in environments that cherish efficiency. This involved meeting the actors' expectations, without disrupting their schedule and damaging the relationship, thus ethnographically adjusting to the field. The majority of the interviews, however, took between one and two hours: a time scale more attuned to sociological norms. I was more direct in the shorter ones, and more probing with the longer ones, and I feel resistant regarding concluding that the longer interviews were more generative.

In a strict skeptical sense, any interview account is merely the enactment of talk, which is detached from practice or the object to which the interview report makes claims (Silverman, 2013): the infamous gap between talking and doing. I have nonetheless cautiously relied upon some claims and assumed these contents to have a meaningful relationship to the experiences and earlier dealings in which the person has been involved: as an actor's 'definition of the situation', which itself may be subject to change over time. Following Holstein and Gubrium's (1995) approach to "active interviewing" shifts the focus away from identifying a singular, 'authentic' answer, and instead to inciting the production of meanings that specifically address the research concerns at hand through various ways of knowing. I have moreover used strategies such as asking participants to show digital work components to ground the accounts in tangible work (Tavory, 2020). I have moreover used various materials such as project documents and work logs to add additional grounding to some of these accounts.

My observations and fieldnotes primarily attended to forms of talk and doings related to technological work. Generally, my observations tried to represent "what those in the setting experience and react to as 'significant' or 'important'" regarding technological work as I interpreted it in their gestures and speech, including online meetings (Emerson, Fretz and Shaw, 2011:26). I also observed collaborative work platforms, such as GitHub, noting traces of former action (cf. Geiger and Ribes, 2011; Turner, 2019), documenting digital materials through screenshots, reasoning that someone must consider what I encountered digitally 'significant' or 'important' at some point in time.

My observations included programming episodes. However, the primary objects of observation were units of action and interactions in which actors engaged in various displays of digitally-mediated work: talking and showing technical objects present in proxy form; for instance, through a set of slides. These were performed in settings such as various work meetings (e.g., meetings, seminars, conferences). These displays often included actors articulating the blueprint (or 'architectures'), segments of code or equations, and technical performance tests. Performance tests include both 'qualitative' ones such as retrieving an output image that shows some object assigned to a class or 'bounding boxes' drawn over an object, and 'quantitative' ones, such as statistical calculations and scores on different metrics used in the community. Other instances included meetings in which actors displayed technology by collectively trouble-shooting systems. Per the interactional nature of these displays, they have been varied, and at times, various unexpected analytical themes have emerged.

The connotation of the term 'display' warrants a brief comment. It may resemble but is different from what the wider tech community may call 'demo', an abbreviation of 'demonstration', which is a highly specific cultural performance common in technology worlds that draws on an ordered and stylized demonstration, which is typically performed to seduce potential investors or customers. In my experience of demos, they differ from displays in academic settings where open-endedness appears to be valued more highly, amongst other things.

5.5 Analytical Strategies

I adopt a broad definition of analytical strategy. As engaged in STS, I draw on perspectives that challenge the sharp division between observation and theory, as well as between description and analysis. The written traces I produce are theory-laden, arising from observations that are shaped by conceptual frameworks. In other words, they are 'cooked' from the outset. A key implication of this view is recognizing that my representations of the field, whether in the form of ethnographic vignettes or stories of events, are partial and constrained (Charmaz, 2014). Describing what I observe is an analytical act shaped by my perspective and theoretical lens, and the selection of specific research materials to represent processes deemed significant is shaped by a set of conventions (cf. Timmermans and Tavory, 2014). Consequently, I expand the analysis to encompass not only the analysis of the data itself but also the invocation of research material, and focus its partial and constructed character.

In my analysis, I have aimed at understanding the terminology and situations I encountered. A key strategy involved re-reading my research materials repeatedly. I have analyzed the materials, fieldnotes, and interview transcripts for relationships between elements in the research materials and used the principle of comparison by grouping data according to my recognition of these phenomena being similar. As an example, this involved concentrating my attention on instances that dealt with work on the same object, such as a specific model, or accounts of a shared practice, such as data annotation. As I have progressed in my understanding of the specialized terminology and work practices, my indexing of interviews through 5–10 keywords, for instance, have served to open up new interpretations of interview and fieldwork segments over time (cf. Charmaz, 2014). Lingering in the contact zone between concepts and sections of research materials, compared to a line-by-line coding approach, may be expressed as doing analysis by focusing on chunks rather than bits; on slices rather than crumbs. Moreover, I have consistently used the technique of memoing: writing up half-formed ideas based on fieldwork impressions (Emerson, Fretz, and Shaw, 2011). Often, these ideas made me return to my dataset to analyze it anew through these lenses, sometimes leading to my abandoning them, or to writing them up more comprehensively.

The principled approach of 'abductive reasoning' involves formulating plausible commentary based on observed data based on a recurrent back and forth between concepts and empirical data (Timmermans and Tavory, 2014). My approach to this has primarily been through using imagined audiences (see Timmermans and Tavory, 2012) and concepts as sensitizing devices (cf. Blumer, 1954), asking concept-laden questions to the research materials (e.g., what would Strauss see, what led participants' research process forward, how do they achieve alignment?). Moreover, as noted by Timmermans and Tavory (2022) peer-reviewing during publication may evoke moments for analysis anew, which tends to be overlooked. In addition to this specific moment noted by these authors, conducting textual revision after receiving comments from colleagues reviewing drafts has been a recurring engine through which my analysis developed, sometimes leading to invoking even more empirical materials, and sometimes to rethinking my conceptual approach ("this is unclear", "is this concept suitable?"). Adding to this, another principle of comparison that has been implicated in my analysis is comparing the object of study with research practices as they are organized and performed in sociology. This has been a situational tactic to make sense of fieldwork impressions, rather than a systematic analytical strategy.

My ability to interpreting participants' activities is not without its limits. When I encountered confusing moments during my research, I made a note of key phrases and terms. Later, if these phrases came up again elsewhere, it indicated they were important, and I would research them further. I have abandoned sets of these as my focus became concentrated on a limited set of observations.

6 Overview of Articles

The presentation of the studies follows how they build on each other, not the date of publication.

6.1 Learning by Doing: An Ethnographic Account of Computer Vision Laboratory Work

Abstract

This article presents an ethnographic account of computer vision research in a leading academic artificial intelligence laboratory, focusing on model development in action. The structural opacity of deep learning models presents an epistemic challenge for the researchers who build them. This paper argues that rather than representing a unique crisis, this condition marks a return to the ordinary scientific practice of working with partially accessible phenomena. This article redirects the analysis from narratives of exceptionalism and crisis to the practical organization of work. Based on ethnographic fieldwork in an academic computer vision laboratory, this analysis advances three concepts that describe how researchers make opaque models workable. ‘Pipeline welding’ refers to the assembly of modular computational components into functioning systems, a process requiring articulation work between humans and computers. ‘Awareness contexts’ are the layered interpretive environments, from dashboards to lab meetings, where researchers develop a functional, rather than fully causal, understanding of model behavior. Informal vocabulary, such as 'tricks' and 'hacks,' instantiate enactments of ‘shop talk’ or ‘model talk’ used to adjudicate and communicate technical choices. The article contributes to our understanding of computational knowledge production, focusing on how scientists build, probe, and make sense of computational models.

Keywords: lab ethnography, computer vision, deep learning, knowledge production, model development, opacity

6.1.1 Introduction

Computational systems are integrated into almost every nook and corner, an expansionist force that Ribes and Lee (2025) describe as "computational universalism", now concentrated in the techniques and practices of artificial intelligence development. In parallel to the recent build-out and rearrangement, Marres et al. (2024) point to the many matters of concern compounding in artificial intelligence, which they refer to as a "super-controversy". This feeds into Ziewitz's (2016) framing of the 'algorithmic drama': a tendency in social studies of computing to depict algorithms as simultaneously powerful and opaque. This drama tied into a phase of bias busting, in which critical attention turned toward uncovering the discriminatory, exclusionary, or representational effects of algorithmic systems (e.g., Noble, 2018). Today, social studies of artificial intelligence appear to proceed within a similar dynamic moving from bias busting to hype deflation: a mode of critique that position inquiry as a counterweight to inflated claims of capability or risk (Sloane, Danks and Moss, 2024; LaGrandeur, 2024; Narayanan and Kapoor, 2024; Bender and Hanna, 2025; Bohner and Vertesi, 2025; Varoquaux, Luccioni and Whittaker, 2025). With generative artificial intelligence, I propose we are, in fact, seeing a spin on the algorithmic drama, which now manifests as a tension between its structural impact and its perceived value: on one hand, it reshapes the social order by reorganizing workflows and knowledge work; on the other hand, its outputs are seen as 'empty' in substantive or epistemic terms, lacking the human judgment, contextual knowledge, or authorial credibility that conventionally confer value (e.g., Bender et al. 2021).

This article responds to these trends by locating analysis in laboratory sites where artificial intelligence is produced in practice, shifting attention from discursive inflation to the practical arrangements that stabilize its capabilities. Following Ziewitz and Lynch (2018), I "return to the lab" to analyze how knowledge is performed, and how work is organized (Strauss, 1993), drawing attention to the "warm" (Latour, 1987) zone and toward the practical activities through which researchers develop working relationships with computational objects as they stabilize them. Rather than seeking to puncture the promises attributed to artificial intelligence, this article turns to the laboratory, examining how practitioners organize, interpret, and sustain their work.

The question of how we calibrate our analytical lenses toward computational processes gains renewed significance as the recent boom in artificial intelligence raises new problems for analysis. This article argues it is important to follow the actors (Latour, 1987) and their activities in their local settings (Blumer, 1969) in their efforts to make analytical categories grounded (Charmaz, 2014) and to absorb shifts and continuities in the social study of computing, particularly after the consequential turn to deep learning.

This article analyzes how computer vision scientists at a leading artificial intelligence lab perform activities that make computational objects workable. Studying these activities helps in understanding how models become durable objects of engineering and knowledge. Those activities aggregate and propagate across meeting rooms and digital sites, which is why they are useful in explaining how artificial intelligence stabilizes in society (see Asdal, 2008, on 'little tools'; Callon and Latour, 1981).

This article advances three related concepts based on ethnographic observations of model development work: pipeline welding, awareness contexts, and model talk. The concepts respond to how scientists put together and negotiate their understanding of complex and opaque models, interpreting the signals and visualizations, experimental results, and engineering infrastructure, developing their awareness of model behavior. Pipeline welding characterizes the work of assembling modular components into operational systems. It captures the work of transforming research questions, like predicting human dance moves, into a working piece of machinery, often by welding pre-existing components together. During welding, "articulation work" (Strauss, 1993) occurs across humans and computers: specifying components, mapping tasks to computational units, and aligning representations. The output is the result of a chain of transformations that passes various information along, with each step shaping what the next can do. Once a pipeline is welded together and running, the work shifts. Scientists cannot open the hood and see the logic ticking away. Instead, they learn to interpret the model's behavior through its signals. Model development creates opacity, and awareness practices resolve it enough to continue development.

Awareness contexts are interactional settings that scientists invoke, often drawing on computational outputs, to understand model behavior. Awareness contexts shift attention away from a model as a binary black box to a model being negotiated in practice. They arise in everyday work within organizations, evoked in work meetings, where courses of action are negotiated. The actors' mobilization of computational outputs means these extend beyond face-to-face interaction into the often-layered, computer-mediated environments: terminals, code, dashboards, and visualizations. Awareness contexts draw attention to the layered environments and interpretive spaces where researchers make sense of what a model is doing.

Awareness contexts are formed at multiple levels. On a foundational level, the emic orientation to the 'data manifold' is an underlying assumption that data has some learnable structure possible to model. On an engineering level, scientists rely on being aware of the code, the model weights—the "boring" infrastructure that makes the work possible (Star, 1999). On an interface level, dashboards, visualizations, and printouts represent and translate the model's inscrutable internal state into something researchers can react to. Then, there is the relational level. It is the conversations in lab meetings where someone holds up a weird-looking generated image and speculates on the origins of the deviation. In such interactions, scientists evoke awareness contexts and negotiate computational model behavior. Such awareness is not necessarily about achieving causal understanding. The goal is more immediate: to develop a functional grasp that allows them to act further: to tweak the code, adjust a parameter, and rerun the experiment.

Within these settings, researchers occasionally use language to evaluate technical choices in situ. This is where the informal vocabulary of tricks and hacks comes in. 'Model talk' (Bondo Hansen, 2021) specified as 'tricks' and 'hacks' capture conventions that adjudicate technical choices. These terms are interactional devices for expressing epistemic judgment and valuations on science and engineering. A 'hack' implies a quick-and-dirty fix that might work, but lacks robustness and theoretical elegance. It is a solution that breaks the written rules; an engineering shortcut. A 'trick', on the other hand, is a clever, non-obvious maneuver that solves a difficult problem. Tricks are referred to as the 'street knowledge' of the field, present in GitHub README files, or informal conversations that never make it into formal papers, including things like lowering the learning rate in a specific model training run or regularly saving checkpoints to avoid disaster.

The paper proceeds as follows. I first review the literature on social studies of computing and identify an epistemic shift in the production of artificial intelligence. I then describe the site that laid the foundation for the ethnographic material. After this, I develop each concept with concrete interactions from the lab drawn from different computer vision research and development processes. Close observation of welding and awareness contexts produces finer-grained categories: components, representations, transformations, involving a dance-motion pipeline, an integration between three-dimensional reconstruction and textual embeddings, and 'model-talk' around multi-view reconstruction. As an example, the final case shows how specified model talk, 'hackery', is used in a lab meeting to adjudicate a pipeline welding choice within a specific awareness context evoked by the actors. I close by drawing implications for STS.

6.1.2 The Practical Production of Artificial Intelligence

This review traces how social studies of computing have analyzed the practical production of artificial intelligence across different eras of computer science. It shows that computational arrangements contribute to reshaping what can be known, how knowledge is computationally formalized, and how work is organized in practice. It also points toward a specific epistemic trajectory: from the relative accessibility and clarity of classical algorithms, burdened via institutional black-boxing; to the epistemic opacity of deep learning; to the central role of production activities and components, such as datasets and ground-truths, and the infrastructures and collaborations that sustain computational work.

In a classical computer science perspective, Knuth's (1997) influential definition of an algorithm is "a finite set of well-defined rules or instructions that, when followed, will solve a specific problem or complete a task". Classical algorithms were deterministic: given a specific input and an explicit rule, one could expect a consistent and predictable output (cf. Simon, 1996). Because the instructions were explicit, algorithms were knowable or epistemically accessible (Wirth, 1976). If you knew the rules and the input, you could, in principle, trace the entire process from start to finish. Agre (1994) notes that symbolic artificial intelligence relied on various schemes for 'knowledge representation', or 'acquisition' (Forsythe, 1993), referred to as 'primitives' or 'minimum replicable units', such as entity–relationship diagrams and semantic networks providing grammar to the represented units. Symbolic reasoning granted researchers access to the internal logic of the determinist systems they developed. This points to a particular epistemic advantage historically associated with computer science, involving accessibility and relatively short routes of knowing the contents of its manufactured object, from a computational standpoint.

This epistemic advantage laid the foundation for social analysts to study formalization practices in computing and their limitations. Technologists interpreted and formalized mutable processes, in efforts to automate them (Forsythe, 1993; Suchman and Trigg, 1993). Scholars argued that a central predicament lies in managing complexity: formalization relied on omitting aspects of phenomena and produced abstractions through selective inclusion (Star and Gerson, 1986; Suchman, 1987; Agre, 1994; Star, 1991; Star, 1995). Adam's (1998) empirical study and feminist analysis of two symbolic artificial intelligence systems show that the engineers encoded assumptions they took to be universal into the systems, but which, as Adam argues, were in fact grounded in their own subject positions as white, middle-class men. In hindsight, while it entailed technical complexity, for a few decades, both computer scientists and analysts enjoyed a rare kind of clarity (cf. Petrick, 2019). But this was an anomaly in the history of science, and most scientific work has never been so transparent (Woolgar, 1988). Corporations gradually circumvented this accessibility and imposed various constraints as algorithms became valuable intellectual property (Pasquale, 2016; Burrell, 2016; Birch and Muniesa, 2020).

Ziewitz (2016) draws attention to the emergence of an "algorithmic drama" forming around this time, that assumes algorithms to be inscrutable and powerful. Ziewitz (2016) observes that much of the then-current literature portrayed algorithms as technically complex and socially consequential systems operating as black boxes, and that researchers proceeded to reason about algorithms' agency, inscrutability, and normativity as contested objects of governance, while methodological suggestions on how to study them proliferated (e.g., Kitchin, 2014).

One of the central contestations from this methodological debate centers analytical orientations. Seaver (2013) complicate the idea that algorithms of the symbolic era were easily knowable. Seaver argues that 'cultural' and 'social' categories as conventionally defined are significant and affect how systems are built, and risk becoming sidelined when analysts follow the distinctions central to computer science, such as those existing between algorithms and data structures, which give researchers conceptual access to the tidy objects through which systems are built, but ultimately fail to address their thicker and messier place in the world. Seaver argues that these categories are in fact central to how engineers think about and enact their work, and that their work outcomes are parts of larger sociotechnical systems that interact with users. Meanwhile, Dourish (2016; 2017) suggests treating 'algorithm' as an emic category and object of professional practice in relation to other significant computational infrastructure, and presents a materialist analysis that carefully traces how these computational objects are shaped, while acknowledging them as sociotechnical and political.

The debate resembles earlier debates in science and technology studies as to whether analysts restrain themselves in studying the social or cultural dimensions of science, or consider the full palette, as it were, including its practical, situated becoming with respect to, and for, the emic categories and activities through which these processes are enacted. However, the issue is tricky since heterogeneous categories compound in practical activity: Seaver (2013) points out 'cultural' categories bleed into technical work. This stance is supported in applied contexts where the origins of influential categories are diffuse. Neyland's (2018) study of an applied system development setting shows that failures and adjustments redefined relevant data and shaped algorithmic objects following a work process of practical experimentation. These efforts were organized by mathematical-aesthetic ideals such as 'elegance', suggesting that such judgments influence system development. Seaver (2021) argues that 'spaces' work both as technical tools (e.g., similarity and latent spaces) and as a metaphor, giving cultural data a reality effect once mapped into measurable forms. Ribes et al. (2019) wrote of how computational actors make sense of their reality through the lens of "domains", meaning areas of expertise that can be computationally integrated.

Other research has pointed to how infrastructure and collaboration condition similar practices. Suchman and Trigg (1993) showed how whiteboard diagrams translated abstract ideas into formal symbolic order. Ensmenger (2016) analyzed how flowcharts historically articulated and managed computational workflows. Vertesi (2016) notes the decentralization of labs through digital networks and shared databases, transforming otherwise centralized sites. Thomas (2019) shows that industrial computer vision often occurs in small groups embedded in wider networks of providers and open-source resources. Reliance on external components and shared infrastructure influences the tools and representational formats available, suggesting that knowledge circulates through organizational membership and open sharing alike. GitHub is a case of this transformation (Kalliamvakou et al., 2015; Turner, 2019; Burkhardt, 2023). It acts as both repository and coordination site, hosting tutorials, documentation, and discussions.

As a continuation of this debate, Castelle and Roberge (2021) suggest tracing machine learning “end-to-end”, and they center three features of contemporary machine learning: model-centric pragmatism, a cybernetic logic of adaptation, and a self-referential epistemology. Lipp and Mayer (2024) suggest focusing on "interfaces" as the analytical locus, drawing attention to the contact zones of human–computer relations. Lee and Ribes’ (2025) urge researchers to relax the use of emic computational categories in social studies of practices. When relying on the meanings actors give their objects, their argument is that computational categories smuggle in misleading stability, which obscures the labor and context that produce them. Bondo Hansen (2025) pinpoints “algo-centrism” as a burdened orientation guiding the field, and suggests that ignoring technical specifics or multiplicity on behalf of the single, over-extended category of ‘the algorithm’ steamrolls particularity, leading to shallow insight. As a corrective, Bondo Hansen proposes the notion of the “stack”, drawing on Bratton (2015) and Dourish (2017), while echoing Bowker (1994; Bowker and Star 1999) in directing attention to the activities and arrangements that make computational systems possible in practice. Jaton (2025:61) similarly draws attention to ethnographic approaches oriented to “the mundane and concrete construction of algorithms”.

In this strand, central findings on practical production are relevant for the present article. Although machine learning is often described as data-driven or autonomous, human agency remains central in arranging and managing these systems (Mackenzie, 2017; Bechmann and Bowker, 2019; Ribes and Lee, 2025). Scholars have hinted at the development of deep learning as a delegated construction of knowledge (Jaton, 2021). Rather than relying on symbolic rules, deep learning is oriented toward optimization with gradient descent being applied to minimize errors relative to a dataset (Castelle and Roberge, 2021). Jaton’s (2021) ethnography of an image processing lab shows programming as an inscription-heavy practice, where objects are translated into numerical registers that can be computed. At the core of algorithmic development is the creation and stabilization of ground truths: datasets that serve as referents against which models are trained and evaluated. This is why analysts have argued for the importance of tracing datasets (Thylstrup, 2022) and for their ground truth (Jaton, 2017; Kang, 2023). Jaton (2023) argues that some phenomena are more ambiguous and harder to pin down in such forms. Engdahl (2024) builds out this reasoning with an ethnography of a computer vision benchmark dataset—a testbed for model performance—representing “daily activities in the wild”. The study analyzes how research participants wearing cameras were aligned with research aims, a process described as “taming” the wildness of the phenomena being represented.

As we have seen, this theme connects to long-standing interests in how knowability and formalizations mediate computational practice. Questions of knowability have long been entangled with computational practice. Building on earlier studies of formalization, the review shows that recent work has shifted from symbolic rules to opaque models and the content of datasets and ground truths. Building on these insights, the current deep learning moment brings about epistemic conditions that chemists or sociologists have dealt with for long: researching phenomena that can only be partially accessed.

The rise of machine learning and, even more so, deep learning altered the epistemic landscape of computing. This shift has meant that research directions in artificial intelligence, such as interpretability and explainability, recently emerged to neutralize the emic concern of black-boxed systems. Learning from the present debates (Ribes and Lee, 2025), affirming this is important in demarcating the shift to deep learning in artificial intelligence that shapes practice on the ground without being held captive to computational categories in analyses of computational practices. I proceed by taking the computational conditions into account as a “contextual condition” (Strauss, 1993), and by understanding a deep learning model as a distanced, learned approximation of a rule.

From a computational perspective, the Universal Approximation Theorem states that neural networks promise the ability to model complex data distributions and approximate their underlying function (Cybenko, 1989). In this arrangement, the rule is implicit in being learned: it is encoded in the weights of the neural network (Goodfellow, Courville, and Bengio, 2015). This introduces structural opacity in contrast to classical and deterministic systems: deep learning models distribute ‘logic’ across many parameters whose joint effects cannot be decomposed symbolically. Once trained, the network transforms into a ‘model’ that produces inferences at test time (Szeliski, 2010). Algorithms were epistemically more open to inspection in ways that deep learning models are not. This opacity represents an epistemic limitation built into the object itself.

This epistemic shift is related to a shift in the material 'affordances' (Davies, 2020) of computational interaction: how scientists grasp computational objects. In a familiar metaphor, the "object gesture" (cf. Mead, 1938) of classical algorithms once presented itself as an open hand; deep learning now tightens around its weights, presenting the gesture of a closed fist. The question, then, is how scientists manage their work as computational materiality structures their interaction. Instead of expecting 'glass-box' transparency that rarely exists in science, a more productive approach is to analyze how researchers manage their work under conditions of partial knowledge. This article contributes to this strand of research by showing how datasets are processed matters significantly, since different model architectures embrace computational elements differently. It also contributes by providing a vocabulary with sociological sensibilities that addresses the lived techniques through which opaque models become constructed and probed, and provides fieldwork instances of the "shop talk" (Lynch, 1985) and "narrations of modeling practices" that Bondo Hansen (2021:608) refers to as "model talk".

6.1.3 The Lab

This article draws on fieldwork in a leading academic artificial intelligence laboratory in North America as part of a multi-sited ethnography of computer vision research and development sites. Prior fieldwork experiences helped with gaining access to this research lab. I visited and conducted participant observation on-site for a month, participating in daily work, attending lab meetings, and engaging in informal discussions. Before and after the visit, I followed the lab's materials that were accessible online, such as papers, documentation, visualizations, and talks, recurringly over the course of a year. I wrote fieldnotes, and documented relevant code snippets, whiteboard sketches, and screenshots when possible. I also conducted semi-structured interviews with researchers (graduate students, postdocs, and faculty). During these interviews, I used strategies such as asking participants to show work components to ground the accounts in tangible work (Tavory, 2020). I drew inspiration from Holstein and Gubrium's (1995) approach to "active interviewing" that stimulate the production of meanings that specifically address the research focus. Participants are pseudonymized.

The lab specializes in three-dimensional computer vision with some research situated at the intersection of computer vision and robotics. In the broader setting of artificial intelligence research, perception and understanding are core epistemic and technical goals at the lab. In their computer vision orientation, visual ‘understanding’ of a scene involves such things like acquiring depth information. For a visual representation, they consider width and height to comprise two dimensions, the addition of depth as the third dimension, and time as the fourth dimension. Their research pursuits include reconstruction techniques like neural radiance fields or ‘NeRFs’, understanding egocentric (first-person) visual data; exploring three- and four-dimensional (spatiotemporal) vision; developing techniques for real-time rendering of three-dimensional scenes; and working on generative models such as diffusion models and video generation; and knowledge distillation, where features from a big model are transferred into a smaller one.

In the lab, copies of books like *Foundations of Computer Vision*, *Elements of Statistical Learning* and *Multi-view Geometry in Computer Vision* are placed on the desks. As I familiarized myself with these books, a central assumption of computer vision is laid out by the authors of the first book: “One of the central metaphors of vision is that of multiple views. There is a true physical scene out there and we view it from different angles, with different sensors, and at different times. Through the collection of views we come to understand the underlying reality” (Torralba, Isola and Freeman, 2024). Moreover, whether these are single-view or multi-view creates different conditions. A key affordance (Davies, 2020) of having several frames with known viewpoints placed in Euclidean space is to retrieve the local geometry of a scene, which in turn makes it possible to create three-dimensional reconstructions. This information enables a stronger grasp of the (geometric) relationship of objects in an image, and fittingly, as we will see, semantic or language-oriented tagging and querying. And similarly, local (part of an image) and global (the whole image) features is another recurring distinction.

Instead of focusing on an image as a source of visual meaning to be interpreted, the scientists are oriented towards extracting information, which ties into the technical process of image formation. When teaching an introductory computer vision class, one of the lab members use a slide containing the following linkage to illustrate this process: "scene radiance—Lens—sensor irradiance—Shutter—sensor exposure—CCD—analog voltages—ADC—digital values—Remapping—pixel values". This overview shows how they imagine translating a physical scene into computable data. From the physical capture of light to the computational inferences drawn from digital images, distinct transformations occur that redefine the utility of the images. Thus, they focus less on interpreting image's content and more on managing and understanding the stages of its transformation into data that are processed computationally.

Three-dimensional reconstruction research, long reliant on geometric photogrammetry, has recently made strides by incorporating machine learning. One method involves overfitting a neural network, unlike a typical machine learning model aiming to generalize onto new data, to a dataset consisting of images on a local scene from different viewpoints. The method is rooted in physics, specifically the plenoptic function. This function is a mathematical representation of the intensity of light rays in a scene as a function of their position, direction, wavelength, and time, believed to capture the necessary visual information from a given observational point in reality. By sending computational rays of light through the pixel space and observing how brightness and form come together, the system learns to recreate the scene. The group and others in their field believe this allows the system, like a learner observing the land from many angles, to understand the depth, shape, and colors of what it sees. They say this happens because the errors, like whispers of numerical misalignment, are carried back through the process, guiding adjustments to the system's understanding through computing gradients of the error between the reconstructed and source images. There is consensus that this achieves close to perfect photorealism, achieving greater realism and fidelity in the representations. Subsequent research has made the core method more efficient, strengthening its geometric precision and usability.

As an insider organizing principle of their practice, some of the scientists aim to reduce constraints on model architectures, following the principle articulated in Sutton's (2019) "bitter lesson". Sutton's provocation, referenced a number of times at the lab, suggests that model generalization and scalability improve when models are designed to rely less on hand-crafted 'priors' and more on large-scale learning through exposure to data. Priors' work stems from Bayesian probability theory and 'inductive bias' from cognitive science involve embedding knowledge and structural constraints into the model architecture to guide processing capabilities. It is assumed that in minimizing constraints, developing systems along a neural-first kind of design principle helps it generalize across diverse tasks and scale up efficiently.

In the lab, scientists spend most of their time in front of their monitors displaying terminal windows in so-called integrated development environments like the program Visual Studio Code, navigating through multiple windows: code in one, documentation in another, visualizations in a third. Publications often foreground novel architectures and performance metrics, but the daily work of computer vision research is oriented towards these tools. The interaction between researcher and the terminal is through which much of the work of computer vision occurs. This resembles what Star (1999) calls "boring things": the infrastructural tools that structure the possibility of their work. These provide researchers with modules and functions, which structure the development work. These constituents shape research possibilities; for instance, through the horizon of functions contained in coding libraries and open-source tooling. Another key material resource at the lab is the computational cluster, which is a system of graphical processing units (GPUs) subdivided into multiple processing cores, sometimes informally referred to as “workers”. These are the resources that enable actors to retrieve the processing power required for high-performance computing tasks, sometimes involving distributing workloads across multiple cores.

6.1.4. Model Development as Pipeline Welding

Model development in the lab unfolds across two planes: the articulation of architectures and their computational implementation as pipelines. The work that connects these planes, what I term pipeline welding, combines conceptual coordination and technical integration. As scientists prepare commands for the processing cores or the “workers”, they confront the question of how these components are organized to perform specific tasks and brought into coordinated motion.

Lab members regularly convene in small groups to discuss their research projects. The scientists draw arrows linking different components together on whiteboards. They jot down flowcharts as a strategy for representing the makeup and informational flows in a potential architecture; others write equations after equations on whiteboards covering meeting room walls. These diagrams are attempts to model the flow of information in architectures whose computational implementation remains open (Suchman and Trigg, 1993).

A key challenge is determining how to integrate components effectively, particularly when attempting to do so in novel ways for publication. Strauss (1993) draws attention to the often-unseen coordination required to bridge different actions into a coherent flow. What Strauss (1993) refers to as 'articulation work' is useful; the only difference is that this activity is achieved not only between humans but among humans and computers: explicating and dividing tasks among computational units and components. At the lab, articulation work involves evoking components and structuring informational paths in ways that can achieve mathematical sense and become accepted as technically orderly. It involves articulation and negotiation of techniques, such as "Can we use 'RL' [reinforcement learning] here?" to which one scientist replied, "I'm not sure this is the right action space". Welding a pipeline involves mapping a high-level research question onto a set of functionally interacting modules. It constitutes a bricolage where components and computational constraints shape what questions can be meaningfully pursued. The epistemic shape of research is partly a function of what can be sensibly pipelined.

A core activity in the lab is pipeline welding. This is the process of integrating components into workable computational pipelines during model development. The metaphor of welding draws attention to both the engineered nature of this work and the labor of making disparate units coherent. Pipeline welding involves linking components, techniques and representations-as-instruments, where outputs from one stage serve as inputs for the next component in a structured pipeline (cf. Ribes and Monteiro, n.d.). These components are active, performing some operation or processing. Fitting them involves welding representational and instrumentational components together. Welding involves more than connecting components; it requires handling the incompatibilities between components that introduce friction into the pipeline.

Pipeline welding relies upon representational regimes: it structures what constitutes a legitimate representation of the research target; in the following case, dance motion. One of the group's pipelines is built to predict human dance motion. This case shows that pipeline welding brings representational units into place: an ingestion script, a lifting estimator, and discretization produce the "motion tokens" that the Transformer component treats as evidence. Those units perform operations: they create tokens and embeddings, processing the data differently. These computational units resample, lift, quantize, crop, and threshold pixels. The pipeline begins with an automated collection ('web scraping') of online videos of couples dancing. The crucial welding work emerges in what researchers call 'lifting': transforming two-dimensional pixels into three-dimensional representations. This operation reworks pixel data into a new dimensional form through targeted adjustments to the images. The process draws on prior knowledge about camera motion and human kinematics, such as joint information, to estimate three-dimensional positions, body shapes, and movements over time. In creating a digital representation of the dancers in a simulated three-dimensional space, one of the researchers refers to the dataset as a "pseudo ground truth". Building on Jaton's (2017) discussion of emic relations to ground truths, the claim here is cautious, noting the reliance on computationally mediated proxies that are imperfect but useful for the research target. With this three-dimensional motion dataset established, the next stage involves welding another module into the pipeline.

The traces of human movement are passed into a Vector Quantized Variational Autoencoder (VQ-VAE). This neural network learns to compress and discretize the motion sequences in numbers, decomposing the dance gestures and encoding them within a finite vocabulary of motion building blocks. The output is a 'codebook' similar to a dictionary, where each code or 'token', basically a number, represents a specific motion. For the researchers, a Transformer model is used to process sequences of the motion tokens derived from the encoder, thus learning the grammar of dance. The predicted token is used to match the corresponding basic three-dimensional motion piece in the learned codebook. In this way, the pipeline affords 'next token prediction': given a sequence of numbers representing the dance up to a certain point in time, the Transformer predicts the most probable token corresponding to the next dance move. This retrieved motion segment is then appended to the existing sequence. The generated three-dimensional motion data is then fed into visualization tools, where the researchers render the predicted dance to visually inspect the results, evaluate the model's performance, and communicate the outcomes.

Another case aims to deepen the capabilities of three-dimensional reconstruction models. The work involved creating a system that could connect visual scenes with written queries, so that typed instructions would return meaningful responses inside a three-dimensional reconstruction. This effort illustrates how pipeline welding involves bridging representational regimes, in this case linking image with language models. This case shows how pipeline welding occurs when the components operate across different representational regimes (images, text).

The pipeline welds neural radiance fields: a neural network which learns to generate a photorealistic three-dimensional scene based on images of the scene, with language-based representations from CLIP, a model designed to associate images with text. 'Embedding' CLIP features directly into the three-dimensional model allows users to type queries, such as "find the red chair", and receive highlighted responses within the scene.

The researchers encountered a key limitation: CLIP tends to provide a broad summary of an entire image rather than identifying specific objects within it. This sensitivity to scale posed a challenge in adapting CLIP to three-dimensional environments, where different viewing distances must be accounted for. Its produced descriptions varied significantly based on zoom level: a close-up of a coffee cup produced a different interpretation than a wider shot of a table. To address this, the researchers mobilized an older component known as 'image pyramids', which processes images at multiple levels of detail simultaneously. This enabled their pipeline to estimate distances within the scene and assign the most relevant CLIP analysis to different visual regions. This laid the ground for users to conduct semantic searches within three-dimensional reconstructions, connecting textual prompts to spatial locations within an image representation: multi-scale sampling let NeRF (image) and CLIP (text) cooperate meaningfully. The researcher described this solution as a 'brute force' approach, indicating a practical, if perhaps not the most refined, welding procedure. The metaphor of pipeline welding is useful to elicit attention to these 'brute force' efforts at connection, involving the bridging of representational regimes.

6.1.5 Awareness Contexts

Pipeline welding is only one part of the work. Researchers must make sense of what these models are doing. The work of system assembly, or fitting components together under constraints, is joined by multiple forms of interpretive work.

As a knowledge activity, artificial neural networks activate a delegated kind of construction (cf. Jaton, 2021). Trained models and their adjusted weights appear as repositories of knowledge. These weights represent the knowledge the model has acquired during training ("a diffusion model has learned something about images”). The actual contents of these values sink below the threshold of human intelligibility (“it’s just values”). The saved state of a trained model, including its adjusted weights, is retrievable via a checkpoint file abbreviated ‘CKPT’, which is possible to load into a programming environment, for purposes like inference on new data. Scientists manage their work under these conditions of partial knowledge. This necessitates different ways of knowing, requiring researchers to evoke and cultivate awareness contexts.

This leads to the second analytic: awareness contexts. Scientists regularly make ideas about model components explicit as part of their central research activity. They evoke and cultivate awareness contexts by carrying out actions to probe and make sense of the opaque models they create. Awareness contexts consist of four levels that follow the progression from root epistemic assumptions, via technical conditions, up to specific interpretations.

At the most foundational epistemic level, lab members draw attention to an emic view of a theoretical latent structure (the ‘data manifold’) assumed to underlie data. It points to an assumed underlying structure that researchers hypothesize about and attempt to identify, but which remains only partially accessible through data analysis. This is a foundational concept: the recognition that data distributions, with all their complexity, are structured and can be approximated. This approximation is not directly accessible, nor is its structure evident. The data manifold serves as a shorthand for the partiality and abstraction involved in modeling work. It anchors the problem space in which awareness contexts are formed and serves as an explanation as to why interpretive work is necessary. At the root of modeling thus lies an assumption about how data group together and allows for statistical modeling, making generalization possible.

The practical actions to probe, interpret, and make sense of the opaque models take different forms. Preparatory work occurs in initial awareness contexts. For instance, some of the members use interactive dashboards with graphs for monitoring training jobs in awareness efforts involving knowing the model is on track. A terminal display of a table of data in an ongoing process allows us to grasp the interaction in such initial stages. The terminal window is the setting within which this performance takes place. The terminal is an interface through which actors engage with and makes sense of computational performances. This act of observation and interpretation is enveloped in meaning-making, where the numbers and projected figures contribute to an understanding of the computational process in progress. The most prominent computational action is the data processing reflected in the changing numbers within the terminal. The action's progress visible through the "Step (% Done)" column acts as a marker of this progression, the speed of individual iterations through "Train Iter (time)", and the computational throughput through "Train Rays / Sec", quantifying the rate at which the unit "Rays" is being processed per second. Lab members, observing such data, direct their attention towards these metrics as they monitor the system's performance.

Similarly, developer platforms are used as an observational instrument for tracking the processes of model training, a symbolic and visual way of managing the delegated process. In the top right corner of a typical setup, the "Add Chart" button is a customizable aspect of this workspace. On the left of this dashboard, a vertical panel displays a list of "Runs" where each run is a distinct iteration of an experiment, identified by a unique name. Two sets of horizontal bar charts illustrate "Parameter Importance". These charts point to which input features or hyperparameters have the most significant impact on the model's performance. Below the parameter importance charts are line graphs depicting the model's performance over time (or over iterations of data processing). The lines vary in color and represent different metrics ('loss', 'accuracy') and provide a visual representation of the training process. In sum, during model training, color-coded heatmaps show how different hyperparameters influence convergence rates in the interface. These outputs prompt responses from human actors, creating a process where meaning emerges through human–computer interaction, as Mead describes in artifact-induced meaning production (Puddephatt, 2005).

Awareness contexts emerge in terminal windows, experiment dashboards, informal project discussions, and during model inference. Doing 'readouts' like running inference from a trained model, and extracting and creating graphs to assess convergence rates, are, moreover, meaning-making efforts with similar goals. Dashboards are an interface that actors use to build up an intuition of the model through various signals. A typical situation involves scientists running an experiment and watching a line graph update in real-time: loss curves stabilize, accuracy plateaus, or an unexpected drop occurs. Actors bring these outcomes (graphs, scores, imagery) with them to work meetings to report on the computational side of the work. These interpretive practices do not necessarily produce causal understandings but enable researchers to act further. They rerun experiments, or switch settings based on understandings gleaned from such interactions.

At the engineering level, technical conditions and artifacts such as documentation, model architectures, code, and weights, play a significant part in structuring awareness contexts. This responds to software engineering sensibilities and captures the hands-on work of implementing systems. In open-source artificial intelligence, making code accessible is as central as releasing the weights, training data, and model architecture.

One practical activity performed to build this awareness is experimentation to correlate inputs, outputs, and internal configurations. Techniques analogous to ablation studies in cognitive science, where parts of a model or input are systematically removed or altered to observe their impact on performance, are devices through which researchers advance their awareness of which components contribute to a model's behavior. These experiments help researchers map the terrain of the model's capabilities and limitations, even if the underlying mechanisms remain obscure. However, this practical, interactional approach to awareness contexts relates to, yet differs from, the goals of research focused on interpretable and explainable artificial intelligence (XAI) that seek systematic methods to make model behavior traceable in causal terms. In awareness contexts, the goal is less about achieving full model transparency of the causal relations akin to a 'white box', an ideal often unattainable with deep learning architectures, and more about establishing effective ways to work with systems that retain degrees of opacity. Awareness contexts are about the mundane and ongoing activities, necessary to navigate the delegated knowledge embedded within contemporary deep learning models: a processual view that moves beyond static 'black box' or 'white box' labels. Researchers seek a functional understanding sufficient for debugging and further development.

As one of my interlocutors observe, experiments serve to clarify the behavior of machine learning models without necessarily uncovering their internal mechanics: "some of these experiments are like helping you understand the behavior of the black box, but not necessarily why it works". While the question as to why the model behaves a certain way often remains unresolved, researchers design experiments to elicit and interpret output, gaining a functional and practical understanding of what the model is doing. This involves negotiating its behaviors rather than accessing its opaque inner processes. As knowledge about the model's opaque space is sought after, acts of querying it occur, in which the model is brought back into contact with the researchers, albeit from a distance.

Throughout model development, researchers interact with models' setup and learning in iterative, back-and-forth ways, which is why this work is reasonably understood to be hybrid human–computer achievements. Scientists interact with models, setting up training runs, monitoring progress via outputs in the terminal and specialized dashboards, interpreting metrics, and making experiments. Taking on the role of the model, they are, in a metaphorical sense, listening to how it speaks as they are querying them, and the results of specific experiments help piece together a working understanding of the model's state (cf. Mead, 1962). Scientists adjust their approaches based on observed outputs, treating the model as an interactive entity whose capabilities must be teased out.

My ethnographic fieldnotes contain numerous empirical instances where actors mobilize computational output during meetings to understand, validate, diagnose, or explain model behavior. The mobilization of these outputs serves as a mechanism for making sense of technical choices and articulating the underlying knowledge embedded within the computational systems. These outputs range from metrics, values and equations to complex three-dimensional visualizations and animations. Numerical values and output visualizations are mobilized as proof-of-concept to validate claims, compare systems, and define a model's utility.

During a one-on-one chat, three-dimensional visualizations and animations were used to explain how a model works. When discussing Gaussian Splatting, a visualization of two-dimensional Gaussians being optimized to fit an image was mobilized to explain the training process—showing how fuzzy blobs "start out as these... distinct Gaussians" and then "jitter and settle into place" over time. This visualization was mobilized to illustrate the implicit optimization behavior of the model.

In a discussion of a dynamic scene reconstruction model, visual outputs were used to diagnose errors. A member pointed out that it seemed to degrade the background, noting that "the building becomes two buildings". This qualitative output thus revealed a failure mode and spurred discussion about why the model behaved that way (e.g., whether the degradation was due to data or model architecture). In another instance, one member mobilized a video snippet of a three-dimensional model on campus to validate the successful implementation of a specific tooling device.

When demonstrating a keyframe initialization feature in a project meeting, a member used a rendered three-dimensional model of a teddy bear to explain the design decision for the visualization: the output was purposefully displayed a certain way because the presenter "figured it'd make sense if u wanna do backprop on CLIP embeddings". Here, the visual output was mobilized and directly linked to optimizing a specific deep learning component.

In a presentation on decentralized diffusion models, 'FID' scores, a common metric for generative models, were shown on a graph. This output was mobilized to demonstrate that training with 'decentralized experts' or multiple model clusters, achieved performance "in the same ballpark" as dense models, providing evidence for the proposed approach's efficiency.

Since the internal mechanism of deep learning is often opaque, actors propose and use visualizations to make the model's intermediate computations legible. In a discussion about an 'online learning model', participants proposed visualizing internal state information to understand how the model learns over time: They discussed mobilizing attention maps to see which visual patches or image regions an internal token was focusing on. They argued this visual output would show that "New parts of scene will light up more, I'd expect, because it adds to the scene", serving as an empirical test of the model's memory updating process. For them, this heatmap acts as a visual explanation for how the model is processing and encoding information.

The scientists thus relate to the visual aspects of the model outputs in their sense-making of model behavior. This shows that computational outputs are mobilized in interactions as a delegated, material, spokesperson (cf. Callon, 1986) that helps to structure the interaction. Outputs are used in negotiations regarding what part of a complex computational system is responsible for a particular effect, a process often described as trying to figure out "what's doing what". Awareness contexts thus situate the ways in which experiments are designed to probe the model's latent space and its behavior.

During a lab meeting about a generated animation of bodily locomotion that looked like a serpentine S-shape, a scientist asked, "Is this S shape in your training data, or is it something the model has learned to generate?". The question sought to clarify the causal origins of the motion, whether it was a result of the diffusion model extrapolating from patterns in the data, or an entirely new output shaped by its internal processing. The response was straightforward: "It's actually something I wrote myself. I didn't check if it was in the training data". This utterance indicates that the S-shaped motion was manually introduced by the researcher rather than emerging from the model's training, pointing to the precarious line between manual intention and automatic oversight in the crafting of opaque models. Importantly, however, the interaction shows that within the evoked awareness context, the actors' probe and relate to the state of affairs articulated by the scientist, prompting them to further negotiate their awareness of what the model is doing: how it is behaving, posing questions like "is this output from this specific component?" while negotiating model behavior.

To summarize, awareness contexts draw attention to the situated work that shapes actors' engagement with models, involving role-taking and the ongoing efforts to make sense of opaque models and ambiguous behavior. Actors evoke awareness context across multiple locations, combining interpersonal and human–computer interaction. Beyond the technical conditions of engineering accessibility are the representational and collaborative levels. At the representational level, there exist the outputs and interfaces through which models become visible: dashboards, visualizations, prediction scores, and rendered images. These human–computer interactions are where artifacts become legible and possible to interpret. Locating awareness contexts as processual means that they rely on interactional enmeshments between humans as well as between humans and computers: actions that involve negotiating knowledge of model behaviors. These processes, in which the substance of awareness contexts is negotiated and locally stabilized, are relationally contingent: addressing shifting constellations of people at invited lectures, Q&As, meetings, all the way to human–computer interaction, where computational outputs are outputted and mobilized in meaning-making processes (cf. Puddephatt, 2005)—aiming at coherence and a functional understanding of model behavior.

6.1.6 Model Talk: Tricks and Hacks

Consider next a work meeting during which a group of scientists discuss their recent progress in developing a model for predicting and reconstructing body motion from multiple cameras and images, referred to as "multi-view three-dimensional reconstruction", as they engage in "shop talk" (Lynch, 1985) or "model talk" (Bondo Hansen, 2021) involving "articulation work" (Strauss, 1993).

As the meeting starts, the project lead writes out the stages of their work on a whiteboard (cf. Suchman and Trigg, 1993), outlining the techniques currently assembled and welded in the model architecture, and their preliminary results.

The group faces the problem of getting their model's output to line up with the model components. Their goal in optimization is to reduce reprojection error, that is, the difference between where points from a three-dimensional model are predicted to appear and where they actually appear in the two-dimensional images. Only by minimizing this error could the scene align realistically with the captured data. The group agreed that if the computational alignment was off, the reconstruction would appear unrealistic. As they worked through the issue, the group discusses how small changes in scale could ripple through the entire reconstruction. They give examples of what happened when things were set incorrectly, such as figures appearing too large or too small, or the perspective becoming distorted. The work meeting turns toward strategies that might stabilize the process: some suggested setting careful starting conditions, while others proposed adjusting the proportions so that people and objects would appear in the right scale to one another. To manage this, different ideas were put forward: one was to anchor the system to an average human height, another was to draw on new tools that promised quick estimates of depth.

As the group considers whether to fold a newer tool into their ongoing work, the meeting quickly turned to questions of reliability. Some saw the tool as too untested to serve as a stable basis for further experiments. Others were more interested in what it might make possible, stressing the value of experimenting with components that stretched beyond well-worn methods. In the end, they chose to focus on strategies they trusted would work, leaving the more contested component for another time. This decision was less about closing down possibilities than about sequencing them, as more speculative approaches could be taken up later. They chose to concentrate on optimizations they felt confident would deliver results within the project's timeframe.

What stood out was how evaluative interaction shaped the moment. The interaction turned on the language people used and how participants drew that line differently depending on their stance regarding the component at hand. At one point, a member described the component as a 'hack', casting doubt on its scientific legitimacy and suggested it was shaky, while calling it a 'trick' framed it as a clever, if provisional, solution. During the discussion of how to structure their pipeline, some stated the value of long-established methods. They pointed out that certain geometric approaches had been refined over decades and offered a level of reliability that newer deep learning components had yet to prove.

As I participated in more meetings and interactions, I picked up on these recurring expressions within this register of critical evaluation. At first, I wondered whether the expressions were a particularity of this specific lab. As I attended talks at other labs, I noticed they also used these expressions. The expressions sometimes appear in scientific publications, as well.

During the daily work in the lab, the actors mobilize these terms when encountering some method component in a paper. When I brought these up as a topic during interviews, I was met by smiles and occasional laughter, and some referred to them as 'colloquial terms'. During an interview, one of the participants notes that the term 'hack' carries a derogatory connotation, implying a quick fix lacking elegance or robustness. As the person noted, "if you're trying to talk down on an idea, you call it a hack. And if you're trying to be neutral, you call it a trick". As an interactional device, the expressions enable criticism while avoiding putting the collaborative order at risk. This fluidity suggests that the expressions are not fixed but are influenced by the intended valuation of the solution. What one researcher considers a 'hack' another might view as a valuable 'trick'. The distinction is thus contextual. These vernacular expressions deal with the informal valuation of welding choices and technical knowledge claims.

The key difference between the terms lies in their perceived elegance and mathematical grounding, and, increasingly within the current machine learning era: the generalizability of the solution. The labeling of something as a 'trick' suggests a recognition of necessary aspects required to make things work effectively, even if such insights do not carry the theoretical elegance expected in academic discourse. Tricks, while seemingly ad hoc, are a form of tacit knowledge gained through practical engagement and experience.

Another scientist related the expressions of tricks to 'street knowledge', gesturing at the practical knowledge gained through experience that resists being formally documented. "It's kind of hard to think of tricks because I think for the people that write the papers they're obvious". The scientist continued, "I think there's a lot of, like, unpublished knowledge or like, things that are in the margins". The communal search for these tricks, what is described as "trying to download everyone else's tricks into your brain", points to the process of eliciting tacit knowledge through informal exchanges and peripheral documentation in the circulation of technical knowledge, sometimes crucial to getting models and components that work effectively in practice. For example, one lab member mentions a 'trick' that involves lowering the learning rate when training Gaussian splatting on a large scene. This insight, in this case, shared through a GitHub README file, exemplifies expertise beyond the formal paper. Another example is regularly storing checkpoints during training, a common 'trick' to address the issue of model divergence (where training becomes unstable and performance degrades). If the model diverges, researchers can restart from a previous checkpoint. A third example of a 'trick' concerns adjusting the learning rate based on a cosine function, leading to improved training stability and performance.

What might appear as a straightforward design decision is parsed through these terms. In one example, the group discussed whether to rely on a familiar structure for representing motion or to move toward a more flexible approach that avoided such predefined frameworks. This interaction shows how the language of 'hacks' enters into the process of building models. A solution that leaned too heavily on established shortcuts risked being labeled a 'hack', according to the scientist, as it implied reliance on a long-used, ordinary method. Adopting a more open-ended design could signal foresight, even if it involved greater uncertainty in technical performance. The debate was not only about which method would work best but also about how the choice would be perceived. Decisions about technical matters, then, were inseparable from takes on what counted as innovative according to current trends in the field or what seemed passé, which others might see as outdated, an example of technical reasoning blending with a sensitivity to what will resonate as novel and valuable in their community, anticipating how their work will be evaluated and understood.

6.1.7 Conclusion

This ethnographic study has shown how artificial intelligence researchers in computer vision navigate their research and development activities during model development. Pipeline welding describes the process of assembling and integrating computational components within arrangements that rely on representational regimes. Pipeline welding includes the configuration and implementation of a computerized workflow. This involves specifying architectural components, selecting and preprocessing data, determining optimization routines, and calibrating parameters: work that requires explicating, dividing, and coordinating tasks across computational workflows. Delegated construction characterizes the distinctive form of knowledge production found in deep learning, where researchers configure conditions for models to learn rather than directly programming their behavior. The primary outputs of this process are model weights or the internal representations that encapsulate learned knowledge while remaining opaque to their creators. Awareness contexts refer to the shared interpretive processes through which researchers monitor and make sense of model behavior, including practices of monitoring loss curves, visualizing outputs, and negotiating quantitative signals. These contexts enable researchers to assess when models are functioning within expected bounds and to identify points of divergence. One research team embedded in one institutional arrangement can have a clear understanding of a specific component of a model, while another may not (cf. Strauss, 1985). These asymmetric knowledge conditions can be said to drive the structuring of technical communication and documentation practices. This conceptual contribution brings us closer to the dynamic ways in which research and development progress.

This analysis suggests several directions for STS. The inability to achieve epistemic access to one's object of study—in this case, to know how inputs transform into outputs—is often represented as a departure from the typical conditions under which science operates in narratives of machine learning exceptionalism. The shift in analytical focus, from opacity as problem to opacity as condition, suggests that contemporary concerns about deep learning may be misplacing the source of epistemic anxiety. We should instead approach this state of affairs as a return to more typical conditions of scientific work, where knowledge emerges through sustained engagement with partially accessible phenomena: in action.

The analytical levels that I propose helps us move away from a binary understanding of black boxes: and, we could say, the black box phenomenon occurs when one of these levels are out of joint, or weakened. The evoking of awareness contexts involves interactional attunement, meaning the calibration of practice to the specific affordances and constraints of opaque models. Researchers learn how to interpret ambiguous signals from models, and which interventions are likely to be informative. Knowing the range of affordances of a specific computational component (i.e., a strong awareness context) shapes further attempts at pipeline welding. Awareness contexts thus feed into articulation work, and impact the success of pipeline welding. In lieu of a wholly rationalist project, tricks and hacks infer a more quotidian layer. One ramification of this is that tricks and hacks help legitimize the suitability of certain components in the pipeline.

Rather than focusing primarily on interpretability techniques or technical solutions to opacity, researchers may further examine how actors develop and sustain awareness contexts in mundane work settings. Strong awareness contexts appear to thrive on traceable decisions, and weak ones rely on memory and unpublished results. Where do strong awareness contexts form and where do they cluster? How are awareness contexts evoked by a specific party managed and shielded from others to maintain the competitive advantage that comes with being aware of a new successful training run with experimental results not yet published? Pursuing these questions might improve our understanding of how the broader artificial intelligence research community progresses, and becomes stratified in practice.

6.2 Agreements "in the Wild": Standards and Alignment in Machine Learning Benchmark Dataset Construction

Abstract

This article presents an ethnographic case study of a corporate-academic group constructing a benchmark dataset of daily activities for a variety of machine learning and computer vision tasks. Using a socio-technical perspective, the article conceptualizes the dataset as a knowledge object that is stabilized by both practical standards (for daily activities, datafication, annotation and benchmarks) and alignment work – that is, efforts including forging agreements to make these standards effective in practice. By attending to alignment work, the article highlights the informal, communicative and supportive efforts that underlie the success of standards and the smoothing of tensions between actors and factors. Emphasizing these efforts constitutes a contribution in several ways. This article's ethnographic mode of analysis challenges and supplements quantitative metrics on datasets. It advances the field of dataset analysis by offering a detailed empirical examination of the development of a new benchmark dataset as a collective accomplishment. By showing the importance of alignment efforts and their close ties to standards and their limitations, it adds to our understanding of how machine learning datasets are built. And, most importantly, it calls into question a key characterization of the dataset: that it captures unscripted activities occurring naturally 'in the wild', as alignment work bleeds into moments of data capture.

Keywords: alignment work, benchmark dataset, ethnography, machine learning, computer vision, in-the-wild, standards

6.2.1 Introduction

This article analyzes the construction of a dataset designed for benchmarking models in a variety of machine learning and computer vision tasks. Benchmarks are related to the Common Task Framework which is characterized by curated data, exact task specifications, standardized evaluation metrics and an evaluation dataset. This framework is known in the machine learning community for facilitating the evaluation of models based on how well they perform in comparison to one another in a constructed environment (Rodu & Baiocchi, 2023). Benchmarks tend to 'become definitions of the essential common tasks to validate the performance of any given model' (for an overview of benchmarks, see Raji et al., 2021:1).

A growing strand in critical data and algorithm studies has shown that benchmark datasets, like ImageNet, display effects that exceed simple reference points: they impose norms in the machine learning community (Denton et al., 2021; Koch et al., 2021). Moreover, scholars have argued that machine learning datasets are partial socio-technical products (Raji et al., 2021). Currently, numerous quantitative measures that aim to add epistemological handles to datasets circulate in the machine learning community (see Mitchell et al., 2023 for an overview); these might, for instance, assess harmful representations or the range of diversity in a dataset's representation.

But how do machine learning scientists construct datasets in practice? This study aims to understand how machine learning scientists create a benchmark dataset as a socio-technical knowledge object. Through an ethnographic case study of a dataset created by a large corporate-academic group, I investigate the work involved in designing, assembling and transporting the dataset within the machine learning community. In contrast to kitchen-sink approaches that integrate any kind of retrievable item (Jo and Gebru, 2020), this dataset was carefully created from the ground up.

More precisely, the footage the dataset is based on was recorded not in studios or with prepared scripts, but seemingly impromptu, with head-mounted cameras worn in data subjects' homes and other locations in their daily lives; its creators describe the dataset as capturing activities as they occur 'in the wild'. Considered as a cascade of individual scenes, this massive dataset might look like a series of first-person perspectives in which hands reach to refuel a machine, push wood through a circular saw as part of a home-improvement task, operate a sewing machine or add shading to a portrait, a desk lamp providing just the right light. Ongoing commentary accompanies the recorded footage, as data annotators have written descriptions of the events that occur and inscribed boxes around objects.

When they are seen as knowledge objects, the process of crafting benchmark datasets involves adherence to agreed-upon standards, which necessitates diverse efforts at alignment. This article details the connection between these efforts and standards and their limitations, emphasizing alignment work's crucial role in the process. At the same time, alignment work is less visible than other efforts, consisting of the informal, communicative and supportive efforts that underlie the success of standards. This regulates a number of intersections where actors and factors interact, revealing the practical work undertaken by machine learning scientists in assembling socio-technical knowledge objects.

A focus on alignment work provides perspective on the comprehensive work a dataset relies on, contributing to an understanding of how it is created and what this implies for computation efforts. For the dataset in question, practical standards were established to organize the conditions in which unscripted activities could be captured 'in the wild', and digital cameras facilitated the observation of these activities from afar. But as interviews with data subjects showed, alignment work entailed a hands-on stewarding that significantly influenced the moments of capture. The dataset creators, for instance, provided guidance on the optimal way to record activities (ensuring maximum camera coverage) and established clear agreements about their parameters – that is, what each should and should not encompass. Alignment work thus bled into – and ultimately domesticated – the activities characterized as occurring naturally 'in the wild'.

Having laid out the terms and stakes of this inquiry, in the following section, I turn to relevant studies of machine learning datasets that contributed to my analysis. Subsequently, I present the socio-technical framework including the concepts of standard-setting and alignment work. I then detail my ethnographic process, materials that informed the results, and the context of the dataset. The findings are presented step-by-step, with each section analyzing a different practical standard and the subsequent alignment work it relied on. Ultimately, I take issue with the 'in-the-wild' characteristic attributed to the dataset by its creators. The article thus presents an ethnographic story of a machine learning dataset, along with commentary relevant to the intersection of the social and computing sciences. I conclude with the suggestion that alignment work is a significant analytic for social analyses of dataset construction because it reveals the practical work that enlivens standards. Throughout the article, I refer to the recorded research participants as data subjects, and the group of machine learning scientists in the dataset project as the group or the dataset creators.

6.2.2 Studies of Machine Learning Datasets

The proliferation of deep learning in computer vision has led to a partial but significant shift in focus from algorithm development to the rigorousness of datasets, as noted by Nasser et al. (2019). The computer science-oriented literature places significant emphasis on data quality and annotation, paying particular attention to the reliability of labels. The current best practice consists of manual human annotation, in which at least two people independently annotate the same data. Consistency between the labels provided by the different annotators, known as 'inter-annotator agreement', is integral to the reliability and accuracy of annotations (e.g., Nassar et al., 2019). Evaluation methods, such as computing similarity scores, are commonly employed to measure inter-annotator agreement (e.g., Yang et al., 2023). 'Datasheets for Datasets' (Gebru et al., 2021), the gold standard for dataset documentation, includes intended uses and limitations, among others, as important criteria for datasets.

In parallel to the computer science–oriented literature, dataset analysis is a growing area of social research on computation, conducted through methods such as critical historiographies (e.g., Denton et al., 2020, 2021), archival research (Thylstrup, 2022) and ethnography (Jaton, 2021). An initial strand of inquiry found that machine learning datasets are composed solely for a computational other: models. In Jaton's (2017:827) lab ethnography, computer vision scientists created a dataset by downloading images from Flickr. These were annotated by crowdsourced workers, who placed rectangles around regions they identified as salient. Scientists then manually segmented these annotated regions into distinct 'salient features'. This process laid the foundation for the creation of a saliency detection algorithm. Trained models are thus shaped by the composition of data and the annotation labels with which they are provided, which serve as a model's 'ground-truth' (Jaton, 2017). Scheuerman, Denton and Hanna (2021:2) describe a similar machine learning convention accordingly: 'Computer vision uses a subset of data instances, known as training data, to fit the parameters of a model. Another subset of the data is known as testing data, leveraged to evaluate the model and estimate the ability of the model to generalize to images not seen during training'.

An important area of focus in social analyses of machine learning datasets concerns the power relations and exploitation in data annotation work (which is generally low-paid, crowdsourced and made invisible) and the social organization of dataset construction, especially in the harvesting of online materials (Gray and Suri, 2019). Thakkar and colleagues' (2022) interview study reveals substantial variations in the valuations of data within the machine learning data supply chain and specific facets of the social organization of workers. Model developers prioritize 'structured data' and data stewards emphasize the need for 'readable' data; while the valuations of on-the-ground data collectors are primarily linked to their performance at the worksite, despite being ancillary to their core work. According to Miceli and colleagues' (2020) qualitative study, top-down communication patterns and ad-hoc updates to datasets exclude data workers from relevant workflows in companies. The authors suggest that dataset documentation should be viewed as a boundary object derived from practice, one that allows for consistent yet flexible communication between relevant actors. And Scheuerman, Denton and Hanna (2021:24) moreover compare the textual production of multiple datasets to understand their underlying values, showing that certain value dichotomies – 'efficiency versus care, universality versus contextuality, impartiality versus positionality, and model work versus data work' – tend to be prioritized when datasets are articulated by their makers.

Another strand of research has shown that datasets contribute to the ongoing production of social and computational realities. Crawford (2021) extends this understanding by describing datasets as 'classification engines' because of their reliance on data annotation labels that influence new data instances. Crawford (2021:139) argues that these classification engines serve to 'naturalize a particular ordering of the world; which produces effects that are seen to justify their original ordering'. Jaton (2023) leverages Kang's (2023) 'ground-truth tracing' to show that the stabilization of historically situated knowledge in a benchmark dataset for cancer immunotherapy is contestable through ongoing knowledge production that revisits previous assumptions.

A related concern in social studies of datasets investigates the use of identity categories (e.g., Scheuerman et al., 2020). A key finding suggests that current practices involve 'the collapsing of human identity – which consists of both visible external and invisible internal aspects – into solely the visible. For example, visible gender expression is being used as a proxy for internal gender identity' (Scheuerman et al., 2020:21). Malevé (2021:1128) argues that in ImageNet, the context behind each and every photograph is collapsed; although photographs are shaped by a range of institutional assumptions about image characteristics such as perspective and metadata, 'The data set is not merely a collection of photographs… Flickr's amateur snaps, police portraits, submarine photographs are re-aligned as items of a training set'. In other words, data become reconfigured. Amoore's (2024) research on machine learning in border policing shows a merging of diverse data types – biometric, biographical, transactional and social media – to stack their features. Moreover, data annotation has been shown to superimpose meanings onto data (Miceli et al., 2020). A key realization from these studies is that representational claims are present in machine learning datasets, though the objects they represent become reconfigured. This is especially significant when the specific composition of datasets will form the basis of the classification and prediction architecture of subsequent machine learning models.

Analyzing the movement of machine learning knowledge has a direct kinship with the analyses of data, socio-technical systems and knowledge infrastructures that take place in science and technology studies. Some of the observations about machine learning datasets reflect earlier findings in science and technology studies. For instance, Latour (1995b) described the chain of analog transformations underlying the production and circulation of scientific referents. During his fieldwork with geologists, Latour traced the gradual transformations through which soil was decontextualized and turned into labels, numbers and graphs, producing scientific referents. In this stream of locality loss and legibility gain, Latour (1995b:170) claimed, 'Reference … is that which remains constant through a series of transformations'. Haraway (1988:583) located similar tendencies concerning the partial visions of technoscience.

Bowker's (1994) notion of 'infrastructural inversion' is another precedent approach that analyzes arrangements that contribute to socio-technical systems before fading away once they are stable. This involves administrative apparatuses (Bowker, 1994:10), classification systems and other infrastructural objects that support systems or tasks, including lists (Bowker and Star, 1999) – and in my view, datasets. Star's (1993) related call to 'study boring things' thus remains a highly compelling provocation as we seek to better understand machine learning from a socio-technical perspective.

In sum, the recent research focus in social studies of machine learning and datasets – aligning with long-standing findings in science and technology studies – has taken issue with conceptions of digital data as pristine, instead understanding data as the effect of work situated in space and time and in diverse formations of people, social contexts, power orders and geopolitical structures (Beaulieu & Leonelli, 2021; D'Ignazio and Klein, 2020). Slogans such as *Raw*

Data Is an Oxymoron (Gitelman, 2013) and *All Data Are Local* (Loukissas, 2019) reveal a common position embraced in social research on computation and data (see also boyd and Crawford, 2012).

By studying the work involved in the construction of a machine learning dataset as a socio-technical knowledge object, this article adds to the growing repertoire of social analyses of machine learning datasets that call for 'bringing people back in' (Denton et al., 2020) through ethnographic approaches (Jaton, 2021) and recognize the unequal conditions in such work (Gray and Suri, 2019). Jaton (2023:803) asserts that 'if we get the algorithms of our ground truths' (Jaton 2017), *we get the ground truths of our organizations and metrological equipment*'. Thus, I place my contribution in the existing literature that analyzes the activities of machine learning and inquire about the constitutive work that moves a socio-technical knowledge object along with a process of stabilization, deeply marked by establishing standards and aligning the relevant constituents in agreement. Moreover, it is significant that this article attends to a novel benchmark dataset, as studies on machine learning datasets tend to accumulate around ImageNet (Deng et al., 2009) – the most influential benchmark dataset in recent computer vision history – and the biased representations built into it (e.g., Shankar et al., 2017; Crawford and Paglen, 2019).

6.2.3 Socio-technical Knowledge Objects: Standards and Alignment

A socio-technical perspective unsettles the accepted distinction between social and technical processes and considers instead their continuous interaction (Bijker and Law, 1992; Latour, 1995a, Mauguin and Teil, 1992; Vertesi et al., 2019). This perspective brings into focus a heterogeneous chain involving human actors, technical elements, devices and work. A knowledge object is formed as actors work on, interact with and navigate the dataset through the various components and stages of production (cf. Latour, 1987). At the forefront is the dataset as a progressively socio-technical knowledge object, along with the work integral to its formation. This framework focuses on two work activities conducted during the construction of the dataset: standard-setting and alignment work. The progression of socio-technical knowledge objects, including machine learning datasets, hinges on the establishment of standards and, subsequently, endeavors to enforce them, which boil down to how actors create and enact agreements (Strauss, 1993).

The explicit aim of the dataset in question is to collect data from daily activities in a variety of circumstances through first-person perspectives. Thus, as a socio-technical knowledge object for computer vision and machine learning, the dataset draws on everyday knowledge. To grasp the processes at work in the dataset, I embrace an understanding of 'daily activities' as mixtures of diverse and historically situated knowledge. Enactments of everyday life are distinguished by a person's use of socially acquired knowledge, both specialized and general (Berger and Luckmann, 1969). Such variable pieces of knowledge enable enactments of something as seemingly mundane – yet socially received and contingent – as doing the dishes. In this way, the machine learning dataset as a socio-technical knowledge object gains scientific value through its link to the embodied knowledge enacted by the data subjects.

A standard serves as a communal basis for enabling the movement of knowledge objects (Timmermans and Epstein, 2010). Conventions and agreements can be made into formal standards, such as Structured Query Language (a common structure of data retrieval in databases) or established protocols for metadata (including author, file size and date created). Bowker and Star (1999:13) define such standards as 'a set of agreed-upon rules for the production of (textual and material) objects'. Similarly, Busch (2011:58, 74) contends that standards act as guiding principles for human actions and objects, and that they are used to ensure operational coherency. Standards put complexity at risk, as they aim to simplify, order and formalize work procedures (Lampland and Star, 2009). Notably, Busch (2011:151) observes variability in the definition of success across context-specific standards and asserts that some standards are intentionally crafted to regularize subsequent processes.

Standards can also be more informal (Lampland and Star, 2009). Lampland and Star (2009:16–24) postulate a 'leaky border' between standards and conventions and highlight locally negotiated efforts that 'touch very specific communities in very specific contexts'. Standards are thus connected to the conventions of professional and scientific communities. From these scholars' research, we learn that standards result from temporary agreements and are enacted in local practices to make further work consistent (Busch, 2011; Lampland and Star, 2009). What I refer to as practical standards are explicit arrangements and methods that follow from 'best practice' (established conventions that vary in their formalization) – for instance, an agreed-upon way of annotating a machine learning dataset. This definition helps identify conventions and agreements, both those that are established in the community and those emerging in local practice.

As a concept, alignment work is based on symbolic interactionist theory developed in Susan L. Star's work on infrastructures and standards, and Anselm Strauss's concept of articulation work and studies of invisible work (Kruse, 2021:5; Strauss and Star, 1999). My understanding of the concept draws inspiration from Kruse's (2021) elaborations of it, which is based on research analyzing knowledge infrastructures (see also Vertesi, 2014). Alignment work is directly tied to standards and makes up the 'work that is necessary to keep the undergirdings of movement in working order' (Kruse, 2021:11). Unlike explicit standards, it may entail fluid, emotional and communicative efforts (Kruse, 2021). The concept directs attention to the less visible, less recognized work that goes into upholding, repairing or enforcing various types of standards – including forging agreement and distributing associated tasks.

Kruse's (2021:3) research on knowledge infrastructures shows that alignment work is common in settings where different 'epistemic cultures' interact (Knorr Cetina, 1999). Alignment work is necessary to enliven agreements at the intersection of distinct knowledge practices – from daily life and from machine learning, for example – where tension inevitably occurs. Drawing on this body of work shows that alignment work is performed to smooth tensions as the knowledge object is moved between sites. I use this framework to analyze the work of defining daily activities and making them recordable, computationally retrievable and possible to disseminate.

6.2.4 Ethnographic Approach to a Distributed Object of Study

This article is produced as part of a multi-staffed and multi-sited ethnographic project. The research team studies different types of visual evidence produced by digital data. Inspired by laboratory ethnography, for the specific study described in this article, I followed the dataset project digitally over the course of 14 months. My own data was produced from multiple sources, including interviews, observations and documents. In total, the interview data consisted of 15 interviews with computer scientists and machine learning practitioners, and shorter segments of interviews with data subjects made available by the dataset creators. I also observed the dataset creators' recorded workshops, the project forum and associated traces on GitHub (Geiger and Ribes, 2011; Turner, 2019). In addition, I collected archival materials and documents, such as data annotation guidelines, benchmark challenges and a data license agreement. I also consulted a published article containing an extensive scientific account of the dataset.

I produced a fieldsite by stitching together digital entry points (Postill and Pink, 2012) for ethnographic analysis. Accordingly, it is best conceived of as, in Hine's (2015:60) words, an 'artful construction' rather than as a site I discovered. The main project website contained an overview of the dataset and related important information, including how to obtain access to the dataset; it also had a forum, with community rules and user profiles. The dataset's GitHub page included documented code sequences, associated work logs and its own issues forum. These entry points allowed for insight into several ongoing work activities. As an ethnographer, I also attempted to penetrate deeper into technological production by participating: I wrote a post in the forum, announcing my presence and seeking further interlocutors and contacted key people sourced from the project website.

After signing the license agreement protecting the dataset, I detailed my study to the group and received their informed consent. Signing this agreement enabled me to not only access the data but also establish deeper relationships with both the creators and the dataset's components, including its metadata. Since the dataset is heavily populated, I was unable to receive informed consent from everyone who contributed to it (Sveningsson Elm, 2009). However, the dataset is mostly anonymized, except for a handful of cases in which data subjects have agreed to show their faces. In line with ethical considerations, I took measures to pseudonymize my interlocutors and scrambled online materials to reduce their traceability.

6.2.5 An Overview of the Dataset

In 2021, a group of computer scientists announced the dataset in a technical publication, a promotional video and an online seminar. It was the result of a corporate-academic collaboration – the main organizing actor, a corporate team, assembled the group and was responsible for the annotation of data, while the academic institutions were responsible for recruiting research participants and collecting data in a range of continents and jurisdictions. The boards of the universities conducted ethics reviews. About 15 local groups recruited data subjects who were willing to wear cameras and contribute footage to the project; ultimately, almost a thousand people in countries across the world recorded footage. According to my interlocutors, the dataset is at the frontier of first-person computer vision, with widespread collaboration required to push the field further.

Promotional videos illustrate how the dataset's creators envisioned the technical outcomes of egocentric vision data. In one, a tidy kitchen is shown from the first-person perspective of a person wearing a pair of smart glasses. The glasses house an AI assistant, which uses a camera sensor to monitor the movements of the person's hands and other objects. During the hustle and bustle of food preparation, an avatar of an older woman appears in the corner of the image and says – to the wearer of the glasses, presumably – 'My trick is to mix two-thirds of an habanero pepper into the saucepan'. Simultaneously, a translucent box labelled 'Habanero chili pepper' appears as a minimalist overlay above the pepper on the kitchen counter.

According to the dataset creators, about 50% of the footage represents handicraft, food preparation, carpentry or domestic work such as washing clothes. The other 20% involve activities such as shopping, eating, reading and playing music. The remainder of the dataset represents activities such as sports and game-playing, watching television and other screen-based media, walking (both indoors and outdoors), commuting and interacting with others. These data play a crucial role in facilitating computer vision tasks, including action recognition, gaze understanding and prediction, hand and object interaction, and three-dimensional scene understanding[2].[1] These tasks support the functionality of systems commonly referred to as spatial computing and mixed, augmented, virtual and extended reality.

6.2.6 Setting a Daily Activities Standard

To give the knowledge object direction, the group began by establishing what daily activities meant. Deciding how everyday life should be represented required significant work in advance of the practical datafication efforts. Specifically, the group used the results of an American government agency's annual survey on domestic life – using materials from established institutions have been shown to

[2] For readers interested in an introduction to egocentric computer vision and accessible visualizations of common tasks, I recommend Plizzari et al. (2023)

add credibility to scientific projects (Shapin and Schaffer, 1985). This survey, which aggregated a flood of information from respondents in specific categories, provided the group with a point of reference and comparison, and statistical information on how everyday life is lived helped them decide which major types of activities to include in the dataset.

The empirically informed categorizations built into the survey reduced potential ambiguity and open-endedness in the group's interpretations of daily life. At the cost of maintaining productive ambiguity, the survey's categories and results were imported, becoming satisfactory and reliable standards for the datafication efforts. However, some categories of daily life are more difficult to handle and codify than others: shopping and procuring professional, domestic or personal care services fall into the category of 'buying products and services'. So does obtaining government services like food assistance, a task which blurs the line between market and welfare activities. Moreover, according to the survey's classification system, cooking breakfast is not childcare; instead, it belongs to the category of 'household activities'. As these issues show, the survey actualizes segmented valuations into the standard.

6.2.7 Aligning the Daily Activities Standard

The survey was a recurring part of the group's presentation of their work in project documents. However, an explicit standard based on these information items – my interlocutors referred to it as 'the list' – became unwieldy in practice, and it was not sufficient for developing the knowledge object in itself. The statistical findings did not seamlessly correspond to the daily lives of individual data subjects in countries far from the one in which the survey had originated. To arrive at agreement on their contributions required negotiation between individual data subjects and the dataset creators, illustrating the close kinship between alignment work and standard-setting. Subtler ways of engaging became tantamount to enlivening the group's standardization of daily activities.

'The list' was used as a first step in screening potential data subjects' eligibility for the research project. When enrolling, potential subjects identified which items on the list they could consider recording as part of their contribution. In my interview with Jane, a senior dataset expert, she recollected questions posed during further interactions with data subjects regarding the range of their hobbies.

During a ‘moment of alignment’ (Vertesi, 2014:268), Jane asked about applicants’ normal daily routines and suggested that of the activities on the list, they record those they usually engaged in. This was done to reduce the risk of prompting specific activities that the data subjects did not typically perform. The version of ‘unscriptedness’ produced here shows that the main consideration was the data subject's familiarity with the activity. Negotiating the activities for the data subject was seen as unproblematic, and it also generated further agreement on the conditions for the recordings.

Occasionally, a data subject mentioned an activity that was not on the list, such as rock climbing, and Jane would ask them to record it. In this case, Jane used ‘the list’ to make everyday life reportable (Garfinkel, 1988), and it also provided guidance on covering activities in the aggregate. Alignment work thus temporarily restored the ambiguity of daily life, fitting data subjects to the group's objectives. However, Jane emphasized the importance of not overdoing such work – allowing for many hours of cycling to be recorded, for example. It was important to accumulate coverage of select activities, not streams of dull events with little computational value. Similarly, to spur diversity in the recorded activities, in their instructions to data subjects, other members of the group set explicit time limits for how long given activities could be recorded. The computational utility of the recordings took precedence over daily life as such.

Alignment work occurred as the group negotiated agreement between the activities the survey suggested needed to be covered and those the individual data subjects actually tended to do. Through textual devices and standards – including pre-selected activities, example recordings and time limits for specific activities – the group aligned the data subjects to producing versions of everyday life that, to various extents, corresponded with items on the American survey. In the work of matching the data subjects to it, the standard as such grew, incorporating the interpersonal agreements that occurred in parallel and becoming more extensive compared to the simple activity categories that had been imported to establish a reliable list. Further alignment work stemmed from one of the costs of this import, which related to a recognized problem of surveys on domestic life: respondents generally abstain from reporting morally sensitive activities (Bureau of Labor Statistics, 2021). The group managed such activities by negotiating with data subjects which parts of their lives they were willing to record and which details would be included in the footage. For instance, data subjects were told to avoid recording visible personal traces like tattoos, which could be used to identify them, whenever feasible. In this sense, the alignment work helped forge agreement on the circumstances for the recordings, equipping the data subjects with an understanding of what was expected from them.

6.2.8 Setting a Datafication Standard

By referencing expert conventions, the dataset creators established a datafication standard. This standard involved head-mounted camera devices adhering to high-definition resolution criteria, generating the desired intensity values within the standardized range of 0–255 across red, green and blue channels, rendering the daily activities computationally legible. This type of image production stands as a long-standing convention within computer vision research. Moreover, at some sites, three-dimensional sensors were moreover prepared for supplementary capture of activities in studio settings. This was done separately from the recording of activities 'in the wild', which was performed with devices ranging from head-mounted cameras to sunglasses with built-in cameras.

The variety of devices meant differences in sensor characteristics, such as the fields of view of cameras. But the dataset creators embraced this variety, which, rather than compromising the coherence of the values produced, was seen as enhancing the quality of the data. A key concern for the dataset creators was the problem of overfitting during model development. A common problem in machine learning, overfitting is the tendency of machine learning models to become too reliant on artefacts from the training process, limiting their ability to generalize to new data instances (Alpaydin, 2020:41). The inclusion of different fields of view diversified the visibility fields of the recorded activities, mitigating the issue of a single camera structuring the dataset. The varied camera array thus served as a safeguard against this problem.

6.2.9 Aligning the Datafication Standard

The skeleton of this socio-technical object, consisting of data subjects and the list of daily activities, gained computational significance once digital values began to accumulate. As data subjects recorded their activities, pixel values – crucial for more significant machine learning accomplishments – began to be produced. In practice, however, this proved to be dependent on work that was both more and less visible. Successful preparation and collection of the recordings relied on several interventions.

One of my interlocutors, Tom, shared stories about the practical work of data collection. As part of a local hub of the project, he was in charge – together with other senior scientists located elsewhere – of transforming people into data subjects who would produce footage. In this specific locale, Tom and his colleagues set up digital meetings to introduce the data subjects to the camera equipment used for recording once the initial recruitment process was complete.

During the meeting, he and his colleagues instructed the data subjects in how to best capture their activities, even providing an example of what a successful recording might look like. Their instruction aligned the data subjects and the recording devices, with the camera as a standardizing device that output numbers to retrievable files. After all, it is one thing to cook pancakes in one's kitchen, and quite another to do so with a camera mounted on one's head, trying to ensure that it captures one's hand movements and surroundings.

When interviewed by the group about their experience of recording their activities, several data subjects alluded to the unfamiliar feeling of recording themselves. For instance, one stated that 'wearing cameras was a challenge at the beginning', but that they 'got used to it as the days passed by'. Another framed it as a 'unique experience', and a third brought up a feeling of 'performance anxiety'. These accounts highlight data subjects' reflexive consciousness of wearing head-mounted cameras for a scientific project in the context of everyday life.

Usually, Tom or his colleagues would drive to a data subject's home to deliver the camera equipment. Wearing masks to prevent the spread of Covid-19, they would knock on the door, greet the subject and help them set up the hardware. They would then leave, while the data subject would record themself performing the agreed-upon activities, playing board games or eating dinner, for instance, in order to produce data. Meanwhile, the local group representative would wait in a car outside, retrieving the equipment from the data subject once the activity had been recorded. In another locale, the equipment was mailed to subjects after the digital meetings, with procedures depending on local circumstances such as geographical proximity.

6.2.10 Setting an Annotation Standard

The dataset creation process also required adherence to a standard for data annotation. This standard determined how data would be marked and labelled, and annotation added computational indexicality to the footage. The group's data annotation standard was developed in consideration of three factors: the computer's ability to recognize and organize the data, the need for consistency across annotations and objectives related to computational tasks such as action recognition. The process involved about four million sentences and was estimated to be equivalent to almost 30 years of work. Consider this brief example of annotation for just six seconds of a data scene:

> #C places the paint roller in the container of paint
> #C reaches for the cup
> #C uses his left hand to grasp the water bottle and drilling equipment
> #C puts the drilling machine horizontally on the ground
> #C removes the water bottle
> #C reaches for the container
> #C shuts the container

Text like this, along with symbols such as bounding boxes – annotated metadata for computational indexicality – was superimposed on the footage by annotators as intermediary actors (Miceli et al., 2020). In the example above, '#C' alludes to the data subject, and like the other symbols (subjects, objects, verbs, etc.), it enables a handle that is recognizable to computers. Basic descriptions of objects and actions, such as '#C is chopping a tomato', '#C is packing a bag' or '#C is mowing the lawn', are centered in the frame.

When I asked one of my interlocutors, a senior dataset expert, why annotation matters in the first place, she replied that data couldn't be used without it. I compared annotation to a vehicle that made the data computationally workable, and she agreed, adding that it made the data indexable. It allowed the dataset creators to see in detail what they had captured, she said – for instance, how many times a cyclist turned left or right. Without annotation, they would know only that they had a video of someone cycling, not what actions were captured or how they were conducted.

As an explicit goal for their data annotation, the group had agreed on consistency, which would enable a larger basis for computational processing of the pixel sets discussed above. As my interlocutor's explanation made clear, annotation efforts are critical to rendering the footage indexable through computer-legible handles. Data annotation is thus seen as recovering the ingredients of the footage.

The interpretations, intentions, goals and affects of the data subjects were outside the scope of interest as defined by the annotation standards. Beyond this definition, we can also distinguish between categories that are locally produced and defined by actors and those that are imposed by experts (Hacking, 1999:168). With this double vision of categories from above and below, the particularities of each data scene and the local definitions of the situations they depict became streamlined into a granular, descriptive and seemingly objective vocabulary. By understanding the range in this way, annotation standards become investments in specific descriptive styles.

As I later learned, the style of the annotations was appropriated from the computer vision tasks circulating within the community. For instance, the group's interest in the computer vision task of action recognition required breaking down eventful data scenes into their smallest components (cf. Agre, 1994). For this reason, action- and object-centered annotations such as '#C folds pizza box' dominate the data scenes, making up one annotation layer out of several, with the rest geared towards other tasks. These task-specific annotation standards are spread throughout the dataset. The current convention in computer vision defines the task of action recognition as a classification problem and accuracy is usually assessed by measuring whether the predicted label matches the annotated label, or if it matches any of the five highest-scoring predictions of what it might be (Plizzari et al., 2023:18). Accordingly, the annotation standards enabled the footage to become a knowledge object integrated in machine learning practice.

6.2.11 Aligning the Annotation Standard

With a defined standard in place, further alignment work coordinated the annotators' efforts. The group commissioned two companies in the Global South to manually annotate the dataset. In this context, the dataset project exhibits parallels with prior research that asserts the role of ambiguous labor conditions as a foundation for the development of machine learning (Gray and Suri, 2019). The group instructed two workers to annotate the same 5-min chunks of footage with the aim of acquiring richer descriptions and correcting mistakes. The double annotator approach helped enforce the standard of consistency as laid out in the annotation guidelines. The annotation guidelines were a significant vehicle for alignment work vis-à-vis the data annotators, and they are key to understanding how the group's version of consistency was accomplished.

Extensive and schematic in form, these documents provided annotators with illustrative examples of how to describe the footage; one instructs annotators to 'fix the start time to the moment the person takes up the knife and apple and the end time to the moment they are done, then write '#C is slicing an apple' into the text area'. The annotation style for action recognition is translated into both a simple task described by example text ('#C is slicing an apple') and the affordances of an annotation interface ('fix the start time'). The annotators also followed open instructions, such as 'Imagine you are viewing this video while talking to a peer who is unable to see it, and you have to describe everything that happens in it'. The results of such instructions were then automatically mined to generate taxonomies that became resources for further annotation.

The annotation guidelines show how the group aligned the data annotators, through textual devices and example cases, to produce what they defined as consistent versions of the footage. The dataset shows, and through the explicit annotation guidelines, it was also made to tell in specific ways. In turn, open access to the guidelines themselves – accounting for how the annotations were constructed – helps the dataset travel to sites beyond the group and is an important resource for external scientists during model work. As such, these documents help align the community of machine learning scientists external to the group with the annotation standard.

6.2.12 Setting a Benchmark Standard

Once the dataset was considered stable enough, it was made available to the wider machine learning and computer vision community. The announcement of the dataset was tightly linked to a set of benchmark challenges promoting distinct scientific work packages. The dataset was classified as a benchmark – source of not just new models but points of comparison. The challenges consisted of computational tasks given names such as 'episodic memory', 'forecasting' and 'hand and object manipulation'. During an established conference in the field, the group organized a competition with standardized evaluative metrics to measure the best technical efficacy on these specific tasks. The substantial work required to construct 'doable problems' in science has been shown to fade into the background once these problems have been articulated; for instance, Fujimura (1987) describes the coupling of experts with relevant know-how with technical objects in situated spaces. Beyond the scientific problems, the group created a contract requiring contestants to adhere to specific licenses, rules and procedures.

Identifying the dataset as a benchmark helped it become a site for comparing the performance of different models. One of my interlocutors underlined the importance of testing machine learning models across many datasets in order to show that they can be generalized. To demonstrate that they have not been overfit to a single dataset, models must perform well on several. Benchmarking is important in this process because it offers a standardized means of assessing and comparing model performance. To date, the contest's leaderboards are populated by about a hundred contestants in various team constellations.

6.2.13 Aligning the Benchmark Standard

Beyond standardization efforts like the rules of the competition and the shared metrics in the benchmark dataset, another type of alignment work occurred. In structuring the benchmark challenges, the dataset creators managed the relationship between contestants and the dataset. To make the knowledge object circulate, they arranged a platform infrastructure on which the contestants' models would operate under the same computational conditions, making the contest fair. And while datasets have exceptionally high value in the machine learning community, the group attempted to solidify the dataset's standing even further – for instance, by offering a cash prize to the contest's winners. Moreover, external scientists circulated the knowledge object within their local community. In one of my interviews with Ken, a contestant from a Big Tech corporation, he described his uptake of the dataset accordingly: 'There was a lot of circling around it internally. I was working on a related task, and I figured that I could make use of this additional benchmark to evaluate my prior work'.

As new actors come into play, the dataset is increasingly adopted and improved, accelerating its entrenchment in the machine learning community. In addition to the competitive aspect of benchmarks, it becomes a temporary site for the work efforts of diverse actors in the wider community of machine learning practitioners. Considering the frequently distributed nature of production within machine learning, this collaboration serves to show that model evaluation first and foremost is a collective accomplishment. Moreover, measured against the scores provided by the group, the contestants produced machine learning models that performed better than the baseline results. As part of the contest arrangement, these models were looped back in to the next rendition of the benchmark contest as code and scientific papers, renewing the standard. Consistently incorporating models into the process thus contributes to strengthening their position and credibility within the community.

Interestingly, external scientists also contributed to making the dataset conform to the standards established in the community. They observed and reported numerous perceived errors in the dataset, prompting the creators to improve it, and communicate the associated updates. It has been observed that when a knowledge object is moved between sites in this way, stability is achieved not as 'a quality (of for example a knowledge object) but as a rather fleeting state that requires the work of a community and can only be temporarily attained' (Kruse, 2021:15). The ongoing troubleshooting and updates to the dataset suggest that this is also the case in machine learning.

This conduit shows that the construction process does not end when a benchmark dataset becomes a socio-technical knowledge object; rather, it provides the grounds for further work. Moreover, the movement from dataset to model performance, contributes to 'black boxing' the dataset – as its inner workings become hidden from view (Latour, 1987) among other so-called state-of-the-art datasets. As a result, once a dataset is stabilized, it becomes increasingly challenging to understand its contents and limitations.

6.2.14 Intervening 'in the Wild'

Along with the diversity of the footage, the group regularly presented unscriptedness as a key characteristic of the dataset. This seemed to imply that the recorded activities had been captured without alignment. The group also described these data as having been captured ‘in the wild’, referring to the collection of data on naturally occurring events, including heterogeneous lighting conditions, significant in computer vision. This term nonetheless suggests that data can be produced without significant interventions by the researchers (in contrast to more organized efforts like recording data subjects in studio settings), and it is sometimes used when scraping materials on digital platforms, which platforms store without researcher intervention.

As I began following the hugely peopled socio-technical knowledge object more closely, a friction developed between my outsider position and the use of this insider term, ‘in the wild’. It began to puzzle me, and my curiosity grew to encompass questions related to the work that contributed to the stabilized object. To approach these questions, my intervention began by reaffirming the setting of standards in the construction of the dataset, which in my view connects to computer science's general relationship to abstraction (Seaver, 2022:33).

Standards rely on alignment work, which my argument was developed to render visible; such work impacted not just the group's internal efforts but the ways in which data subjects proceeded during their recordings. By understanding alignment work, we come to see the comprehensive efforts that enabled the data subjects to perform their activities. This destabilizes the notion of the dataset as ‘unscripted’ or captured ‘in the wild’, suggesting instead that data subjects performed activities in accordance with their alignment.

My argument draws on the following observations: The American domestic activity survey was used to establish the initial meaning of ‘daily activities’ and provided a list of activities for the dataset. Data subjects negotiated with the group the activities they would record, further structuring their performance. The group instructed data subjects on how best to record themselves while wearing the head-mounted cameras, further aligning them to a shared agreement about the situation. Wearing their temporary visual companions, data subjects moved about in unconventional ways after pressing the record button, and they avoided recording certain aspects of their activities, their surroundings and themselves – tattoos, for example. Annotators then added symbols and text to add additional meanings to the footage, contributing to a remarkable degree to describing the data scenes. Throughout, the process of putting the dataset together relied on the group's alignment of actors and factors.

At times, these contingent efforts were performed reflexively. The instruction to avoid recording certain aspects, for instance, arose from concerns about data privacy and an attempt at de-identification. Other moments of active stewarding followed from legal concerns and expressed an ethos of care – for instance, undressed children were protected from the roving camera frame within a data subject's home. Interestingly, the group's stewarding of footage was more readily articulated when it was aligned with legal standards than it was when sourcing activities from a survey, for instance. Thus, within a more explicit moral framework, the unscripted and 'in the wild' attributes of the footage seemed to dissolve. In this sense, the dataset recomposes a clean, ethical version of daily activities.

The dataset marks the creation of not a representation of daily activities as they are performed – which it is presented as being – but, instead, of an object that must be construed through the very range of actors and factors contributing to it. The 'unscriptedness' of the dataset footage is the result of meticulous work that thoroughly configured the recorded data scenes, not only setting the stage for the recordings but structuring their contents. The argument presented here thus concerns the dataset's relationship to the 'daily activities' as being one that is first and foremost creative and additive. These are not, as they are commonly understood to be, external, natural behavior or objective phenomena that were captured from a detached point of view.

6.2.15 Conclusion

This article has told an ethnographic story of the process of constructing a dataset: the work it relied on, its composition of daily life and the work that enabled it. The story centers on the standard-setting and alignment work that was required to enact practical standards for daily activities, datafication, annotation and benchmarks during the process; in doing so, it advances previous and ongoing discussions by articulating the work necessary to establish and enact specific conventions and agreements. Crafting benchmark datasets, seen as knowledge objects, is a process that involves adherence to agreements or standards that necessitate diverse alignment efforts. Thus, the results of my study contribute to the ongoing discussion about how the machine learning datasets essential to contemporary computer vision models become constituted and how we as analysts understand them.

Throughout the process, the dataset creators performed alignment work. Using a national survey on domestic life in the United States, they created a list of 'daily activities', and in instances of alignment with the data subjects, they amended this list to screen the activities that were covered. Through practical walk-throughs on how to best record their activities while wearing a camera, they further aligned the data subjects. To achieve the goal of consistency through task-specific demands, they aligned data annotators through extensive instructions. Meanwhile, external computer scientists helped align the dataset with the standards already put in place.

Alignment work is less visible than overt standards. But once its importance to this process is registered, it becomes difficult to embrace the idea that data was captured in the wild. Comprehensive configuring, including the employment of various alignment scripts, was necessary to produce 'unscripted' footage, calling into question the significance of 'unscriptedness' as an attribute. Recognizing the persistent efforts that go into alignment work does not dethrone it – it becomes no less scientific. Instead, our understanding of scientific work practices is made more thorough, our understanding of what dataset construction entails beyond the recognized conventions of scientific practice (that standards should be perceived as rigorous work, for example) is deepened.

The mode of analysis in this article supplements quantitative measures on datasets. As new computational models are forged on and evaluated through this dataset, which activities it records is significant, as is the way those activities are performed. For instance, there is a consistent pull in the dataset towards coherent activities, which influences subsequent computational outcomes. The dataset recomposes 'unscripted' everyday activities into sanitized versions of themselves brought into existence by a collective of many, the minutiae of which are narrated by a descriptive other. It locates daily life primarily in the kitchen, not the toilet; the living room, not the bedroom; in distinct activities, and less so in the gaps between them. In this sense, the dataset commits to a situated and specified version of 'daily activities'. Put differently, the construction of the dataset culminated in a partial version of the stated data objective. This version entails that new and existing models that set out to measure their performance are under the homogenizing control of what daily activities have been made to look like. To develop or compare computational models of messier, neglected or morally ambiguous activities – inherently part and parcel of daily life as it is lived – appears out of reach in and for this dataset. This realization is critical as machine learning scientists develop models that are intended to inhabit such borders.

We have seen that to create datasets and make them valuable for machine learning, it is necessary to impose visibility, difference and similarity. Data annotation is the dominant method for 'uncovering' what takes place in recorded footage, and it contributes to the construction of the recorded content; enforcing data annotation standards makes data indexable according to certain automation objectives. I have traced the ways of performing annotation to the computational 'tasks' of interest they are contingent on, such as action recognition. This linkage opens an avenue for further research, such as genealogical studies uncovering the historical processes that have shaped specific tasks or ethnographic studies of developing such tasks, or the model work they activate.

6.3 Wuthering Weights—Localization Trajectories of Machine Learning Models for Local Ends

Abstract

This chapter analyses trajectories of contextual reconstruction of machine learning models with calibrated weights, sometimes known as pre-trained or foundation models, as they are made operational in specific settings. Through ethnographic vignettes on car detection, bird tracking, and medical technology, it underscores the importance of cooperation and negotiation among actors facing contingencies during trajectories of contextual reconstruction. These stories help us see the behind-the-scenes work of bringing machine learning systems into existence, including interactions between machine learning scientists and domain members. Trajectories are shown to be non-linear, negotiated, and contingent on technical, infrastructural, and regulatory conditions. The process of localizing such models results from the collective effort of actors, their negotiations, and resource combinations. Actors use their resources to gather expertise and pool work efforts to address unforeseen events during contextual reconstruction trajectories. The chapter thus considers the wielding of an implicit social license, whereby actors leverage their accreditations to successfully navigate contingencies during work with models that otherwise are open-source and retrievable for many actors.

Keywords: scientific work, machine learning practice, trajectory, contextual reconstruction, foundation model

6.3.1 Introduction

In the course of my ethnographic research, I have studied the work activities of academic scientists in Sweden and the United States who specialize in applied computer vision research within the machine learning community. Through following these laboratory environments located at academic institutions, I have observed a similar practice of working with a particular type of model arrangement, which is the topic of this chapter. These scientists engage in a two-step process that forms the core of localizing machine learning systems: acquiring models trained on extensive datasets and subsequently adapting them through contextual reconstruction to achieve specific objectives. In insider terms, machine learning scientists and practitioners fine-tune, re-train, or build on top of such models to meet local ends. This chapter focuses on the situated and heterogeneous interactions that form the lifeblood of such localization trajectories.

The training phase-a computational rite of passage of sorts of a neural network primarily involves the recalibration of weights. Simply, weights are akin to dials that regulate connections in a network, and help determine the influence of specific input on the network's output. The training phase helps set these dials at a good starting point. Each weight in a neural network modifies the signal passing from one node processing data to another. Through refinements in a process known as learning, these weights are iteratively adjusted, each iteration a step towards minimizing the error between the network's predictions and the dataset. Through the network's exposure to data, weights become repositories of learned experience. Sometimes termed parameters, these weights, can proliferate into the millions, occasionally billions.

However, access to the computing capabilities-collectively known as 'compute' which encompasses processing units, specialized software, and data center infrastructure-necessary for training these models, or calibrating the weights, are unevenly distributed (Vipra and Myers West, 2023:4). This is particularly pertinent considering the recent trend and success of developing large models with immense quantities of weights (Sevilla et al., 2022). A direct consequence of this distribution is that a limited number of actors can train large models. Having access to 'compute' thus appears to be a central institutional privilege. Under these dual conditions of constraint and promise, the machine learning community has begun sharing significant resources. Shared models, especially those already calibrated, become valuable resources in this context. A family of models referred to as pre-trained or foundation models is a central resource being shared[3]. The

[3] Foundation models are the current name of the game in the machine learning-centric approach to artificial intelligence vocal actors speak explicitly of them as the current paradigm (Bommasani et al., 2022:6). Foundation models are characterised by their initial training on extensive and diverse datasets, typically at scale, followed by subsequent customization for various specific applications. They can be 'adapted (e.g., fine-tuned) to a wide range of downstream tasks' just as pre-trained models (Bommasani et al., 2022:3). In an influential publication from a leading research group that defined and popularized the term, they state that 'pre-trained models' is an unfortunate term because it suggests that the important work precedes or is contained within the model (Bommasani et al., 2022). In contrast, the term foundation model is meant to highlight the work that comes after acquiring such a model. Such work is the topic of this chapter. Moreover, foundation models sometimes differ from what is called pre-trained models in scale and training regime, usually being self-supervision without the need for supervised learning with annotated

central feature of these different terms is the ability to access machine learning models with calibrated weights and localize them in new settings. Beyond the link to open-source software and open science movements, the sharing of resources is rather limited, aligns with various strategic reasons, and comes with licenses that contain various restrictions (Gray Widder et al., 2023). However, models with calibrated weights are important as they provide an advanced starting point for developing numerous specialized models for particular automation objectives. In the community, it is moreover asserted that the use of such models improves technical performance compared to creating a model from scratch. These repositories of learned experience are thus brought to bear in multiple new settings.

Models with calibrated weights are critical to understanding how machine learning or artificial intelligence comes to be placed in various contexts in society. It is recognized that significant work occurs after the acquisition of models with calibrated weights. The two-step process of pretraining and adaptation is significant in all of the three terminological variations of models with calibrated weights mentioned above, but little is currently known about such work (Bommasani et al., 2022:30). A prominent tradition in the social studies of science and technology involves precisely the analysis of the production and localization of science and technology, which serves as a valuable framework for tackling the following research questions. How do machine learning scientists and practitioners account for and use models with calibrated weights in practice? What occurs during such localization trajectories and how may we understand them? This chapter thus contributes to understanding the localization of contemporary science and technology.

The ethnographic vignettes of this chapter show how actors contextually reconstruct models with calibrated weights for local ends. The findings bring into view machine learning scientists and practitioners as situated actors partaking in trajectories of contextual reconstruction. The first vignette deals with Maria's and her group's work on automatic car detection. The group used a model called You Only Look Once (YOLO), which is trained on the Microsoft Common Objects in Context (COCO) dataset, built on a Convolutional Neural Network (CNN) architecture. The YOLO model's architecture is designed to be capable of

data. Similar struggles of competing definitions tend to reflect various institutional commitments beyond the accuracy of their meanings.

detecting objects in real-time. The model utilizes a single neural network to directly predict bounding boxes and class probabilities from full images in just one evaluation-hence the name, as the model, so to speak, only looks at an image once. The group developed a supplementary dataset to the model, while encountering unexpected challenges of participation during the trajectory. The second vignette follows a group also using the YOLO model but for bird tracking. This vignette shows interactions during technical deliberations in which data and sensors were reconstructed, sometimes referred to as 'building on top'. The third vignette deals with a trajectory in which another pre-trained Convolutional Neural Network (CNN) built for a medical application got localized in hospitals. While it required adding new data to make the model useful, it shows the work of associated actors contributing to the trajectory of contextual reconstruction in which machine learning systems become localized.

The generative actions and interactions of machine learning scientists and technical systems are critical for bringing technological futures into existence. Machine learning scientists engage with machine learning models with calibrated weights, often called foundation or pre-trained models, to craft desirable futures in localization trajectories. They are resources to accomplish context-specific machine learning automation. Actors use models with calibrated weights to manage constraints, workloads, and technical performance. The process of integrating such models to meet local ends, which machine learning scientists refer to as 'fine-tuning' involves adding new targeted datasets. Contextual reconstruction may involve adding components, changing the sensory hardware, and conforming to various policies and regulations. The vignettes reveal the contingencies that arise during the localization trajectories of science and technology. The concept of trajectory helps unpack the static idea of fine-tuning a model for a downstream task when localizing objects of science and technology. The path to get there is not straight-it twists and turns based on what is regarded to be needed for the specific context, what is available in terms of resources, and what the rules allow. Actors use their implicit social license to pool in expertise and additional work efforts to mend such contingencies during trajectories of contextual reconstruction. In the push toward a discourse of open-source artificial intelligence, unequally distributed social licenses help us understand why innovation nonetheless tends to occur in certain clusters of society.

6.3.2 Tinkering with Objects of Science and Technology

This chapter intersects with social studies of science, technology, artificial intelligence, and machine learning systems. Tinkering in scientific practice is a key concept that emerged from a set of studies conducted during the 1970s and 1980s. Initially called the cultural approach to science, these studies researching the mundane and material practices of scientists and engineers at work became known as ethnographic lab studies (Sismondo, 2011). These studies argue that mundane, material manipulations are central to scientific advancement, not mere by-products (e.g., Knorr Cetina, 1979). Moreover, Clarke and Fujimura (1992) highlighted that scientific work is situated within specific environments and times, employing a range of tools and material practices that are context-dependent and contingently performed across situations.

A persistent theme for social studies of science and technology is the transportation of knowledge objects and the process of localizing them in new contexts. Scholars have addressed the localization problem by tracing how objects of science and technology are enacted in practice. Taking inspiration from the elements, scholars have analyzed objects of science and technology as fluid and responsive to their spatial setting (de Laet and Mol, 2000) and as fire that progresses in uneven bursts through different places (Law and Singleton, 2005) in contrast to firm, durable extensions of actor networks that hold objects in place (Law and Mol, 2001). These developments were contrasted with studies such as Callon's (1987), in which individuals in engineering environments were shown to perform a complex role that blends engineering with sociology. These individuals, whom he refers to as 'engineer-sociologists', are responsible for orchestrating firm networks that include both the human participants such as economists, consumers, and technicians, and the often-recalcitrant physical components they work with, to integrate technological systems in a context. Technology development is not only a technical affair. It involves gearing humans too, Callon argued, in agreement with other sociotechnical analysts (e.g., Bijker, 1992).

Jaton (2021) advocated for the Latourian (1994) concept of delegation to describe how engineers leverage algorithmic assemblages to organize data and create models for automated analysis. Jaton's notion suggests that actors' variable use of computational architectures yields diverse outcomes that contribute to shaping realities. Other studies indicate that delegation is relatively static and at odds with the processual nature of machine learning. Broek's et al. (2021) study on the construction of a machine learning model for job hiring showed that machine learning scientists align the technical capabilities with the perspectives and needs of domain members, as the latter request changes to the system to make it more useful to their specific work practices. Researching computer and data scientists, Ribes et al. (2019) found that actors routinely parse the world and delineate the boundaries between sites through what is conceived as different areas of expertise through the vernacular expression of domains. This demarcation specifies the specific group expected to benefit from and use a constructed technology, informing the development of tools facilitating use and interaction across the minted boundaries (Ribes, 2019:5). Concerning the localization problem, a critical part is thus the interaction between machine learning scientists and domain experts with valuable knowledge of the setting undergoing automation.

Machine learning scientists (e.g., Bommasani et al., 2022) have cautioned about the potential 'homogenization' that may occur due to the proliferation of what we may describe as the social cloning of machine learning models (cf. Porter, 1995:15). In a sociological and conceptual analysis of pre-trained models, Varshney et al. (2019) argue for the immutability of such objects due to the high computational cost of modifying them, making them into Latourian (1990) immutable mobiles which are artefacts that retain their characteristics as they travel to different settings. Despite this immutability, scientists and practitioners interpret pre-trained models flexibly to redress them to meet novel aims which Varshney et al. suggest are central factors in their success. In this regard, it is tempting to invoke Star and Griesemer's (1989) concept of boundary objects, i.e., objects that have a window of adaptability that facilitates cooperation between distinct local practices. This concept was crafted as a critical response to the idea of 'immutable mobiles' in actor network theory. It suggests that as objects traverse various communities of practice, they undergo partial adaptation per the ends of the local situation, which is why their movement is successful (Bowker et al., 2015:171).

6.3.3 Trajectories of Contextual Reconstruction

The framework used in this chapter builds on these findings and adds propositions developed in pragmatist sociological studies of science and technology. A pragmatist theory of action elicits attention to the emergent and creative features of interaction, rather than succumbing to the notion of merely adjusting a model in agreement with the technical demands of a 'downstream task'. I incorporate Strauss's (1993) concept of 'trajectory' as central to our understanding of scientists, practitioners, and other actors 'doing things together' (Becker, 1986; Clarke and Gerson, 1990). As Strauss (1993:53) articulates, a trajectory not only signifies a course of action but also the contingent interplay of multiple actors that are beyond complete control. In scientific and technical work, scientists continually adjust and reconfigure their approaches as 'unanticipated and anticipated contingencies are manifest in the situation' (Clarke and Fujimura, 1992:11). This includes tinkering with recalcitrant objects of all kinds in particular situations, that is, getting objects of science and technology to do what scientists and others want them to do (Clarke and Fujimura, 1992:15). Actors within such trajectories strive to navigate and shape the emergent challenges, working within their constraints to meet local ends (Strauss, 1993:57). Chang (2004:406) explicates Mead's (1938:641) notion of emergence in pragmatist sociology, which is integral to Strauss's framework, to describe a generative process where novel entities manifest as actors and objects converge in ways not present before. This highlights the processual and emergent nature of action, where actions give rise to unintended outcomes that then inform subsequent conditions for future actions, thereby allowing for the alteration of methods, and imparting a flexible character to interactive courses (Strauss, 1993:37-38). This framework allows us to consider the trajectory of each acquired model not as a linear progression but as an emerging path shaped by various interactions that lead to the contextual reconstruction of these tools for specific and localized ends. A trajectory encompasses the diverse actions and interactional processes that machine learning scientists engage in while collaboratively developing, integrating, and contextualizing sociotechnical systems. The contextual reconstruction of a model with calibrated weights is part of such a generative trajectory, where new relationships and dynamics come into existence. This is why the vignettes pay close attention to sets of interactions.

6.3.4 A Multi-Sited Ethnographic Approach

The data for this research comes from a multi-site ethnography of machine learning practices in computer vision. That project is part of a multi-staffed team ethnography researching the use of digital methods and data for computer vision applications in a heterogeneous set of practical contexts, from computer laboratories to courtrooms. The data for this chapter stems from my fieldwork on machine learning work practices. The study spanned academic institutions, research labs, conferences, and work meetings. I purposely recruited scientists engaged in current computer vision projects, interviewed a set of them, and followed their work projects by doing follow-up interviews, and researching archival materials such as project documents and scientific publications. I briefed participants on the study's aims and their rights, obtaining their informed consent. Throughout the research, I pseudonymized relevant data. I conducted an iterative thematic analysis comparing the sites over time, to identify key patterns and themes in machine learning practices within computer vision.

6.3.5 Findings

Vignette 1. Maria is a machine learning scientist in a computer vision laboratory in one of the most prominent academic institutions in the United States. Her lab focuses on applied science, which does not require building foundation models from the ground up. A task which, in her view, is apt for foundational artificial intelligence research labs, taking up a position in the division of labour contingent on the wider community of actors. In my interview with her, Maria said, 'Object detection is domain agnostic, right? Applied AI labs are just leveraging these kinds of off-the-shelf techniques, but adjusting them to answer questions, um, that they need'. Maria explained that the process of training a foundation model from the ground up is impractical for her group. They find it unnecessary since several foundation models are 'open-source, user-friendly, and widely recognized in the scientific community'. The group's use of existing foundation models avoids the cost and complexity of training a model from scratch-a strategic choice that makes use of existing work. Machine learning models thus gain specific locations in a hierarchy of credibility that allows scientists to trade on their status (cf. Clarke and Fujimura, 1992:16). Training a model from scratch is an extensive and time-consuming act. However, after reflecting on the work of making models suitable to a specific setting, Maria said, 'the fine-tuning piece still requires a large amount of GPU'. Maria detailed how she and her lab members routinely struggle to get access to their departmental computing possibilities. In Maria's lab, they run a timetable protocol to manage coordination. Computing capabilities are thus a condition that intersects with local work activities (Strauss, 1993). The coordination of access to 'compute' among members within individual laboratories also reflects a strategic allocation of computational resources. It represents a form of commitment to use Becker's (1960) term-to the potential success of particular training regimes of models. Such commitments to specific computational strategies under operational constraints have implications for the scientist's credibility within the group.

Maria's account of their localization of a model with calibrated weights in practice begins with the aim of the work: a 'use case' focused on enabling dash cams in vehicles to detect passing cars, identifying their make, model, and speed, to gauge real-time carbon emissions. This illustrates how the projected trajectory of technical work is delineated according to the 'use case' of car detection to determine carbon emissions (Strauss, 1993). The organizing principle of the 'use case' is crucial: it aligns the machine learning model with the system's anticipated needs within its anticipated setting. It defines the automation objective, and as a principle, it directs the subsequent adaptation of the model according to this specified objective. Her group acquired a version of the YOLO object detection model trained on the Microsoft Common Objects in Context (COCO) dataset. Maria underscored the significance of ensuring that the model is paired with additional relevant data to optimize its performance. Maria articulated this process as 'fine-tuning', where some of the layers of the YOLO model are adjusted to enhance its relevance to the specified use case. The fine-tuning involved extensive data engineering work, particularly the collection and labelling of diverse images of car make and model. That was a hard piece of work', Maria stated, which required 'scraping' relevant repositories and annotating the dataset. Initially working with 'emissions profiles for different cars that are all open source' and the Stanford car model dataset, which contained 152 classes, the group expanded this to 4,900 classes, incorporating granular details such as the release year of each car model due to frequent updates by manufacturers.

The following shows that contextual reconstruction also implied other technical work beyond adding relevant data to 'fine-tune' the model. A central concern for Maria and her group was to balance model accuracy and inference speed, given the operational constraints of the setting. Their focus on deploying the model on a dashcam for real-time inference highlights the practical challenges in machine learning localizations outside a controlled laboratory environment. The group used an 'autoencoder' added 'on the end of the YOLO model' to enhance the model's performance for vehicle make and model classification. Their decision to use a 'quantized version' of the encoder reflected a compromise between maintaining model performance and ensuring feasibility in scenarios called edge computing, where neural networks are integrated on smaller, data-generating devices that have much less 'compute' than a laboratory server. They used additional algorithms like 'Deep SORT' and 'Strong SORT' for tracking and ranking objects in the images in real-time. They decided to exclude objects further than 15 meters away from the camera relatively arbitrarily. Maria motivated the decision by referencing the organizing principle once more, 'It's kind of use-case driven, right?'

Maria and her group encountered several unexpected hurdles along the trajectory. She described that her team faced difficulties in persuading people to install this equipment in their vehicles, indicating a participation issue. Additionally, she described a lack of knowledge within the group regarding the deployment of their algorithm on dashcam hardware. Despite their intention to implement the algorithm directly on the hardware, their model made this type of deployment unfeasible for them.

Vignette 2. "To really train a network from scratch is something that almost no one does anymore", Robert said during one of our more extended conversations on the issue of foundation models. Robert is a senior member both at an academic computer vision laboratory and a national artificial intelligence research center in Scandinavia. Robert described the technical work required for localization, "You're doing so-called 'transfer learning'. You take your data, reweight some of the layers in the networks, and essentially tell the networks, 'these are the real, important combinations", thus fitting the model with the needs of another setting. Referring to machine learning scientists and practitioners with limited but sufficient resources, he continued by saying, "These models have the advantage that it is reasonably economically feasible for people to do things"

Robert and his academic group consisting of computer vision and machine learning scientists worked together with John, a biologist, who used computer vision in his research. He had been working on a data habitat of sorts for several years. He had carefully set up camera sensors, and lately, used computer vision techniques to record a specific bird species and their behaviour. John wanted to track individual birds to reconstruct their movements. He explained that they hoped to calculate the average time it took for this bird to capture fish by tracking individual birds to understand how long they left their chicks unattended. Additionally, they aimed to track pairs of birds further down the project.

John also incorporated the YOLO model in his work. Despite that COCO-the dataset on which the YOLO model is trained-had no coverage on this specific bird, John and his team nonetheless retrained the YOLO model to classify the species into three classes: eggs, chicks, and adult birds. As in Maria's case, this necessitated a comprehensive data engineering effort as retraining was dependent on collecting numerous images of the bird species and annotating them along the classes of interest to John, to create the grounds for the model to statistically comprehend the pixels in images of this specific bird and their varying makeup. For this data engineering project, John and Robert assembled a temporary workgroup that made external scientists contribute to their work during a 'hackathon' the work required more hands than their own.

For the contextual reconstruction, John wanted to track rings placed on the birds' feet to understand individual bird paths, in addition to the classification task. The group, including myself, listened as John explained the integration of the YOLO model with the Google Cloud Vision Optical Character Recognition Engine (OCR). He highlighted the engine's ability to accurately recognize the bird's foot rings and extract textual information based on the rings, enabling the model to identify individual birds more effectively. After evaluating various options, he said the Google engine emerged as 'the best performer for interpreting ID numbers based on the image data.

However, John was dissatisfied with the results. During the meeting with Robert's group, John showed a video of the tracking algorithm he used, which produced bounding boxes over individual birds in the recorded footage. John narrated the video clip as it went on for us to inspect, detailing a recurring problem. The footage depicted birds perched on a shelf with bounding boxes hovering around them. Each bounding box also displayed a unique bird identification number (ID), intended to be used for tracking individual bird trajectories. When several birds grouped closely, however, the automated bounding boxes would occasionally shift over to a bird encroaching on the position of a previously ID'd bird. The model, in short, allocated birds with the wrong ID.

John explained that this confusion in the algorithm happened when the birds clustered together and that this was especially troubling as another bird species frequently entered the habitat and attacked the group of birds, prompting them to huddle together. The IDs changed multiple times throughout the data or footage, consisting of 5-minute-long chunks. The configuration of birds changed as they flew away and returned to the data habitat at different times, making it difficult to track their routes. This unforeseen problem of shifting IDs meant continuous work for John. John asked for a specific solution: 'It would be great to have a metric on the predicted strength of an individual bird path'.

Referring to this task as 'robust tracking', John initiated an in-depth discussion on the issue of automatically tracking individual birds as they fly in and out of the camera frame. The interaction occurred between John, the domain expert, and the technical specialists. The members began to engage in a discussion on the model's limitations and areas of improvement. One member wondered if the model's performance would be affected by birds of different sizes or species, while another inquired about the impact of occlusion when birds overlapped in the footage. A third member suggested exploring other OCR engines to further improve the model's performance. Another member proposed training a neural network exclusively on challenging cases. Yet another suggested using a technique commonly used in human pose estimation. This technique involves measuring and predicting key points' of the body, such as joints, which could potentially help identify individual birds by supplementing their measurements. Neither of these suggestions seemed to resonate with the problem.

As the discussion began to wander, John acted reflexively as he pondered on whether the problem definition might limit the possible answers. A decisive moment followed as he demonstrated how the metadata could be utilized, which deviated from the previous interaction. The data frame had been set up to extract temperature readings. He mentioned that the infrared (IR) camera showed that the birds' feet heat up due to sweating to manage body heat when they return to the data habitat, the shelf, after catching fish by diving into the sea. Upon hearing this, a member excitedly pointed out that the intruding bird, which causes the mix-up of bounding boxes and IDs, would have hot feet. The biologist's face lit up, and he agreed, saying, 'True! That's really cool!'. Making this knowledge object-temperature readings-present in this contingent way thus became essential in this trajectory. Distinguishing between the hot and cold feet became a resource for keeping track of which bird was which. The conversation shifted to a more practical register as John mentioned that the results might vary on sunny days. A third member, the lab manager, remarked that thermal cameras were expensive and that this approach would rely on such sensors. The group eventually reached a consensus that not all of the 40 cameras in the current construction should have heat detection capabilities, but one station could be set up for that purpose. They could then train a model using the data collected from this station, information which could then be used to inform the other cameras. The atmosphere in the room felt content, having produced a practical solution to the problem, and a sense of quiet satisfaction settled among the members.

This new arrival-the concept of heat as a tracking parameter-prompted a significant reconfiguration of the sensor setup. John's recourse was to continue rearranging the model and hardware setup. The unpredictability of bird behaviour, and the limitations of computer vision in tracking individual birds illustrate the dynamic characteristics of machine learning and the importance of collective problem-solving. Once very incidental data, the temperature readings became central to John's tracking system. This circulation of knowledge helped the actors regain a sense of control during the trajectory. As the model's bird confusion compromised the original intent of tracking the individual birds, this shows the importance of ongoing interaction in the face of circumstances emerging in the localization trajectory.

Vignette 3. The following trajectory draws on a story about Tony, a machine learning scientist at an academic institution who is also the CEO of a company in the medical technology sector. Tony and his group developed what he called the MRI Assistant for radiotherapy in the medical domain. The idea was to aid the decision-making of clinicians in determining radiation therapy dosages. He described it as an 'intermediate layer between Magnetic Resonance Imaging (MRI) and linear accelerators' the latter delivering radiation therapy based on the information provided by the MRIs. The MRI Assistant analyses MRIs as input images, and creates synthetic representations containing 'segmentations' that highlight the organs and tumors, which in turn provide a basis for radiation dose calculations. They used a Convolutional Neural Network (CNN) architecture that was described as a network passing images along layers that act as filters, resulting in a limited set of output values ultimately used for object detection in images. Tony and his team had then built on top' of this model by pairing it with a segmentation model, which helped to automatically highlight organs in images, creating a proprietary product.

As in the other two vignettes, the data for training the model required substantial work. With his double affiliation, Tony had been able to leverage his collaboration with medical experts to add additional data to the model curated and annotated by the domain experts themselves. As medical experts had annotated the images, academic collaboration was crucial in developing this training data. Tony and his group used various computational techniques to augment those datasets-generating synthetic data. Various agencies had vetted their pre-trained model, but only after his group had received additional consultancy support on how to pass the evaluations, and public funding to hire the consultants. The system passed the typical technical evaluation tests, which made the developers ready to 'bring it to market'. The interaction of actors in the trajectory would ultimately reshape the technical system, not once but twice, despite its operational performance. His account of other actors' involvement - effectively nudging the system design based on another construction of the situation-shows the emergence inherent in a trajectory (Strauss, 1993).

Tony recounted different challenges encountered during the distribution of the pre-trained model-turned-system. Ahead of the first 'distribution round', he stressed their team's efforts in creating encryption layers for the patient data, de-identifying 'clinical data and patient information' before sending it to 'the cloud'. At another site with stronger computational capabilities, the system would work its heavy operations and then return the model's output in a lighter format to the local hospital. At this point, Tony said they got 'painfully aware of hospitals' unpreparedness of cloud solutions'. Clinicians could not send their patients' data outside the hospital's systems regardless of the encryption. From Tony's perspective, stringent data privacy regulations, such as the GDPR in the EU, complicated legal and logistical aspects. This added nuisance to 'onboarding' hospitals to the system. 'So much for cloud-based solutions in the medical sector... he shrugged, and after a brief pause, he said, 'So we went back to the drawing board'.

To counter these issues, he said, his team introduced the Heaven server, a high-performance system with 'the best CPUs and GPUs to process neural networks on-site', or much 'compute'. Moreover, communication layers had been implemented to facilitate their control over the local systems from their computer laboratory. The heavy machinery, accessible online, was now to be brought in and placed in hospitals worldwide. For Tony, 'complications' now emerged in Asian markets. Because of their policy, he said, hospitals' online access is restricted and this implied that his group could not update the systems remotely. This made them prone to reduced functionality, for instance, if they discovered a flaw in the network. These conditions appear emergent as surprises brought to Tony's perspective, but are effectively the productive outcome of other actors' work in the trajectory. Once more, Tony repeated with additional emphasis, 'Back to the drawing board.... In response to these conditions, he stated that the team implemented yet another solution relying this time on 'external disks' to carry the information from afar. In yet another effort of contextual reconstruction prompted by the productive nudge of associated actors, this was needed to perform updates to the computational powerhouses in the hospitals.

6.3.6 Conclusion

These ethnographic vignettes show how actors contextually reconstruct models with calibrated weights during localization trajectories. This involves fine-tuning models by producing, collecting, or pairing data to meet these defined objectives, starting from an advanced baseline to manage constraints and performance. Actors engage in complex work organized by the specific 'use cases' that determine the relevancy of such efforts. But beyond the relatively static understanding of fine-tuning a model according to a use case, these trajectories are continuously renegotiated by the interaction of various actors within their constraints. This makes the paths to clear endpoints challenging, as the trajectories are replete with emergent problems requiring collaborative resolution. I therefore borrow 'wuthering' a term from the Yorkshire dialect-describing the windy weather around the main farmhouse of Emily Brontë's (1847) story to denote the whirling turbulence of the work accumulating in and around these objects. This parallel between the farmhouse's exposure to relentless winds and the models' exposure to vast datasets for weight calibration and the ensuing contextual reconstruction helps emphasize the observed whirlwind of activity. Whether referred to as foundation models or pre-trained models, the vignettes illustrate the subsequent and diverse work following their acquisition.

A successful localization of models with calibrated weights is contingent upon the mobilization of diverse expertise and the pooling of cooperative work efforts.

The contextual reconstruction of such models depends on diverse expertise to mend the contingencies seemingly inherent to localization trajectories. These cooperative actions extend beyond technical tasks, encompassing the assembly of temporary workforces, solicitation of scientific feedback, navigation of regulations, and leveraging various social networks for data acquisition. This underscores that localization trajectories bridge disparate worlds, facilitating the collective enterprise of contemporary automation, under the condition of expansive actor cooperation. Such cooperation does not by default necessitate symmetrical consensus, recognition, or compensation.

While the YOLO model was fine-tuned in the first vignette, the effort of contextual reconstruction was bolstered by combining it with an 'encoder' and other algorithmic components. Yet, beyond these integrations, the trajectory depended on a cooperative effort to enrich the model with added and annotated data. Cooperation was, however, not fully realized in user engagement for necessary hardware installations. In the second vignette, the enhancement of the YOLO and OCR engines necessitated pooling external practitioners for data annotation. The cooperation moreover extended to soliciting insights from machine learning scientists to devise a solution to the tracking problem as another effort of contextual reconstruction. The third vignette illustrates that beyond the contextual reconstruction of CNNs, the trajectory relied on the cooperative involvement of medical experts. These experts not only contributed data but also annotated these themselves, which could later be augmented. Further cooperation was sought with consultancy services to navigate regulations, and with local hospitals to meet policy requirements. In these trajectories, cooperation is paramount.

When mining these vignettes for analytical insights, the idea of a social license thus seems to play a crucial role in the success of these cooperative efforts. This term refers to the informal accreditation that actors receive based on their institutional credibility and the leverage of their organization. A social license enables individuals or groups to facilitate cooperation with a range of experts-be they technical, regulatory, or domain-specific by leveraging their status, influence, or funds. Accreditation of this kind is necessary to establish and sustain cooperative relationships, which are necessary for the successful handling of anticipated and unanticipated events. A social license is indispensable for cooperation: it enables actors to mobilize work efforts.

The capacity to engage in effective cooperation-to wield a social license is as important to the advancement of machine learning as the more technical recalibration of models themselves. Such a social license may implicate the very trajectory of contextual reconstruction. The significance of a social license precedes the technical aspects of collaborative work. Its influence can set the stage for successful collaborations and subsequent technical achievements. For example, well-regarded laboratories or scientists may be granted access to corporate data as part of reputation-building strategies, or receive preferential access to computational resources within their institutions. As critical as the technical licenses, it is important to note that just as with computing capabilities, social licenses are not evenly distributed among actors. This is especially relevant to the topic and debate of open-source technologies, often imagined to level the playing field. Social licenses help us understand why machine learning innovations tend to occur in specific clusters, even if many actors may access and retrieve powerful models. Future research should investigate whether the reliance on social licenses is an attribute specific to the resourceful participants of this study or a more widespread practice within the machine learning community.

7 Conclusion

This dissertation has traced how computer vision capabilities stabilize through the situated work of research and development. It contributes to the social study of computing by demystifying the activities through which artificial intelligence models that 'see' are constructed. Understanding how this work is organized and performed is significant because it shapes the operation of consequential computer vision systems. Doing so clarifies how computational knowledge is produced at the frontier of contemporary science and technology.

A central conclusion of this study is that what makes computer vision components function is the situated work through which they are assembled, ordered, and sustained. A deep neural network may operate as a computational assembly line, transforming inputs across layers through additions and multiplications. But actors must construct and maintain this: specifying architectures, coordinating data flows, and evaluating outcomes. Suitable data must be carved out, contingencies must be managed during implementation, and outputs must be interpreted in light of the conditions in which they are produced. We have seen how the material production of pixel values and recalibrated model weights, and the calculations performed by hardware like processing units, shape how actors perform these processes.

The three analytical texts presented here represent a partial but integrative account of computer vision research and development processes. The first examines day-to-day work in the lab, focusing on the complexities involved in model development; the second analyzes how datasets used for evaluation are constructed and the work that holds them together; and the third traces how models are adapted and implemented in different settings. They show how model development, dataset construction, and model implementation rely upon interactional processes.

The concepts developed throughout this research are attempts to articulate the interactional processes of researching, building, and managing these model-centric systems, to understand how scientists think and navigate the challenges they encounter in practice. The necessity of pipeline welding, sometimes using 'tricks', and avoiding 'hacks', and evoking relationally bound awareness contexts to interpret opaque model behavior, shows how research and development rely on interactional processes. Building datasets through alignment work, adapting models in contextual reconstruction trajectories, and gaining access to resources by social licenses depend on cooperation between people, and the interactional device of 'use case' is a way for actors to tether models to new settings.

The relationships between these conceptual propositions point to the dynamics of progress in science and technology. Pipeline welding refers to the assembly work through which representational and instrumentational components are joined into workable pipelines. In this process, researchers evoke and cultivate awareness contexts, meaning, layered interpretive environments that make model behavior actionable in practice. These layers span epistemic assumptions, engineering arrangements, interface representations, and the relational work of interpreting results together. Awareness contexts are comprised of local conditions of intelligibility, including the technical configurations that render model behavior observable, the interpretive frameworks that make such observations meaningful, and the relational arrangements through which actors share, negotiate, and sustain these understandings in practice. Rather than a fixed condition or cognitive achievement, intelligibility appears as a practical accomplishment maintained through mundane work with computational artifacts. Such contexts do not necessarily remove uncertainty or supply causal explanation, but they provide enough understanding to continue development.

The findings suggest that while awareness contexts tend to remain within research labs and immediate collaborations, social license operates at the interface between the lab and its wider institutional surroundings. Social license may be seen as an informal accreditation that enables cooperation, access to data, and coordination across sites. It helps sustain collaborations and, at times, provides the resources that pipeline welding and awareness contexts depend on. In this sense, the infrastructures that support pipeline welding (e.g., component documentation) and awareness contexts (traceable decisions, experimental readouts) and the conditions that sustain social license (transparent agreements, partnerships) form an important foundation for how research and development progresses in practice: how actors keep work moving, negotiate obstacles, and sustain momentum across technical and organizational boundaries.

When analyzing how actors make sense of their work, several tensions surface as they navigate complexity and opacity daily practice. One concerns the development of neural-first models designed for scalability versus models that rely on specific constraints or priors. Another involves adopting novel deep learning techniques while still valuing the robustness of traditional computer vision methods. A further tension arises between individual creativity and the limits of available resources, particularly academic computing in contrast to the capacities of industry powerhouses. Researchers also negotiate between the assumption that camera sensors and data capture a pristine reality and the recognition that representations, including image datasets, are constructed and partial. They face the opacity of deep learning models alongside the demand for understanding and control. And, they move between idealized workflows of research and development and the local realities of situated work where contingencies arise.

In addition to social studies of science and technology, this work hopefully speaks to scholars in interdisciplinary fields such as critical data studies, and critical algorithm studies, offering an empirical account of work within computer vision research and development. Human–computer interaction (HCI) and computer-supported cooperative work (CSCW) scholars may also find this research valuable for the analysis of how scientists interact and collaborate computationally.

A common way of thinking in science and technology studies (STS), following Hughes, is that things could be otherwise (cf. Woolgar and Lezaun, 2013). I apply that reasoning to the question of the representational capability of the analyzed benchmark dataset. In this concluding chapter, I want to address another dimension related to this, which deals with research priorities. Understanding how technical performance, infrastructure, and collaboration are negotiated in practice, rather than being stable, or simply governed by technological determinism is important. There is indeed an ongoing debate about the politics of research and development priorities in artificial intelligence, and it is not only social scientists who raise these questions. Computer vision scientists themselves reflect on the priorities that shape their field.

Within research groups, scientists themselves pause to ask what kinds of knowledge are being pursued and to what end. For instance, at one division meeting, this took the form of a discussion on the pressures of chasing state-of-the-art results. What emerged was a collective negotiation over what kinds of knowledge should be pursued. Benchmarks and leaderboards were described as both structuring progress and narrowing what could count as valuable work. Participants asked whether PhDs were trained as visionaries or as paper producers. Some argued for slower science and for taking intellectual risks. In these exchanges, the field itself became the topic of reflection, as researchers negotiated what directions might be possible beyond the paper treadmill with reports of incremental benchmarking results. The push to scale up systems and chase state-of-the-art results appears necessary to advance technical performance. These priorities can also be understood as being shaped by career-building strategies in competitive academic and industry settings. Publishing papers that break new ground on benchmarks, creating larger and larger datasets or models, are ways, it appears, to gain recognition and secure positions, funding, and promotions. This calls attention to the competitive forces felt by scientists, as well as the question as to what kinds of work are neglected in this concentration of certain priorities. The focus on bigger outputs can crowd out other considerations, such as maintaining systems, better understanding them, or addressing broader impacts. Promoting sensibilities of care (Lindén and Lydahl, 2021) seems useful in this discussion, drawing attention to the values and judgments behind what researchers choose to care for, improve, or ignore.

Some consider artificial intelligence to be the central technological advancement of our time. Computer vision work shapes technologies that, in turn, shape action elsewhere. The study may be therefore interesting to readers outside of these specialized fields of inquiry. I would be happy if this dissertation could serve as an act of translation between different groups, making artificial intelligence research and development slightly more understandable to others outside the community. By providing this kind of translated understanding, the hope is to contribute towards more productive dialogue about the governance and use of the technologies that shape our worlds.

The texts presented here have shown how STS sensibilities rub up against insider terms in artificial intelligence, some of which are: autonomy, in-the-wild, and building on top of pre-trained or foundation models for downstream tasks. For autonomy, I draw on concepts such as articulation work, delegated construction, and awareness contexts juxtaposed with the technical pursuit of systems capable of acting independently: 'autonomy' is hence continuously 'worked out' among people and computers (cf. Strauss, 1993). The notion of 'in-the-wild' used in artificial intelligence to describe data or experiments conducted outside controlled environments resonates with STS attention to how the lab extends into the world. Here, the figure of the detached observer gives way to a more situated understanding of scientific practice, where boundaries between experimental and “natural” settings blur. The practice of building on top of pre-trained or foundation models for downstream tasks aligns with STS interests in how reuse, adaptation, and legitimacy in and around science and technology are managed as models circulate and are repurposed across contexts. Accordingly, I have approached these terms not as fixed terms but as part of ongoing accomplishments through a processual lens.

This has involved straddling social scientific and computer science-based terminologies and perspectives. The analytical move performed here, staying close to insider talk and emic terms, can be interpreted as reproducing and accepting certain realities rooted in dominating information–managerialist views. This reasoning, I believe, leads to forfeiting empirical scholarship on certain actor groups, which is troubling. I hope readers of this study will parse the work through their perspectives, instead of withdrawing their interpretive agency, that they will assume that the stories that are told here are not the only possible stories to tell. Understanding these actor groups, I maintain, benefits our understanding of what we may broadly refer to as how power operates in society.

Moreover, if we briefly take ‘AI’ as a category, it becomes evident that it fragments into a wide range of procedures, techniques, and methods, and has recently concentrated narrowly around generative models. To describe this condition, we might draw on terms such as multiplicity, pluralism, or splintering.

Another benefit of staying close to insider terminology is that my research more easily travels to the field and the practitioners with whom I have engaged. But movement between the emic and the etic, technology and the social, involves troubled passages, even at the level of terminology. An example of this is my use of the word ‘emergence’, rooted in a Meadian conceptual universe. Meanwhile, ‘emergence’ became a contested term in the artificial intelligence community after the publication by Microsoft researchers (2023) “Sparks of Artificial General Intelligence: Early experiments with GPT-4”. They argue that a specific AI system appeared to exceed its training regime; generalizing in such a way that it bypassed its immediate constrains, a phenomenon connected to the imaginary of advanced artificial general intelligence. Similarly, when I spoke of trajectories in the field, my interlocutors understood it in the mathematical sense, like the path of a projectile, rather than the sociological sense of an emergent line of collective action. I believe these moments of misalignment point to the strains that accompany interdisciplinary engagement, which require a willingness to work within these frictions. Making sense across worlds or ‘epistemic cultures’ (Knorr Cetina, 1999) demands continual effort.

List of References

ACM A.M. Turing Award (2018, 2024) 'Turing Award Recipients by Year', *ACM A.M. Turing Award*. (Accessed: 2 October 2025)

Adam, A. (1998) *Artificial Knowing: Gender and the Thinking Machine*. London: Routledge. doi:10.4324/9780203005057.

Agre, P. (1990) 'Portents of planning: A critical reading of the first paragraph of Miller, Galanter and Pribram's Plans and the Structure of Behavior'. Paper presented at the Conference on Narrative in the Human Sciences, Iowa City.

Agre, P.E. (1994) 'Surveillance and capture: Two models of privacy', *The Information Society*, 10(2), pp. 101–127. doi:10.1080/01972243.1994.9960162.

AIAAIC (2025) AIAAIC Repository. (Accessed: 9 February 2025)

Airoldi, M. (2022) *Machine Habitus: Toward a Sociology of Algorithms*. Cambridge, UK: Polity Press.

Akrich, M. (1992) 'The de-scription of technical objects', in Bijker, W.E. and Law, J. (eds.) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, pp. 205–224.

Alpaydin, E. (2020) *Introduction to Machine Learning*. Cambridge, MA: MIT Press.

Amann, K. and Knorr-Cetina, K. (1988) 'The fixation of (visual) evidence', *Human Studies*, 11(2–3), pp. 133–169. doi:10.1007/BF00177302.

American Sociological Association (1988) *Footnotes (ASA Newsletter)*, 16(5).

Amoore, L. (2024) 'The deep border', *Political Geography*, 109, 102547. doi:10.1016/j.polgeo.2021.102547.

Amoore L. (2023) 'Machine learning political orders'. *Review of International Studies*. 49(1):20-36. doi:10.1017/S0260210522000031

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.

Asdal, K. (2008) 'On Politics and the Little Tools of Democracy: A Down-to-Earth Approach'. *Distinktion: Journal of Social Theory*, *9*(1), 11–26.

Atkinson, P. and Housley, W. (2003) *Interactionism*. London: SAGE Publications Ltd.

Bachelard, G. (1953) *Le Matérialisme Rationnel*. Paris: Presses Universitaires de France.

Barnes, B. and Bloor, D. (1982) 'Relativism, rationalism and the sociology of knowledge', in Hollis, M. and Lukes, S. (eds.) *Rationality and Relativism*. Oxford: Blackwell, pp. 21–47.

Beaulieu, A. and Leonelli, S. (2021) *Data and Society: A Critical Introduction*. London: SAGE Publications.

Beaulieu, A., Scharnhorst, A., and Wouters, P. (2007) Not Another Case Study: A Middle-Range Interrogation of Ethnographic Case Studies in the Exploration of E-science. *Science, Technology, & Human Values*, 32(6), 672-692. https://doi-org.ludwig.lub.lu.se/10.1177/0162243907306188

Bechmann, A. and Bowker, G.C. (2019) 'Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media', *Big Data & Society*, 6(1), pp. 1–11. doi:10.1177/2053951718819569.

Beck, U. (1992) *Risk Society: Towards a New Modernity*. London: SAGE.

Becker, H. S., Hughes, E. C., Geer, B., and Strauss, A. L. (1961) *Boys in White: Student Culture in Medical School*. Chicago: University of Chicago Press.

Becker, H.S. (1986) *Doing Things Together: Selected Papers*. Evanston, IL: Northwestern University Press.

Becker, H.S. (1982) *Art Worlds*. Berkeley, CA: University of California Press.

Becker, H.S. (1976) 'Art worlds and social types', *American Behavioral Scientist*, 19(6), pp. 703–718.

Becker, H.S. (1960) 'Notes on the concept of commitment', *American Journal of Sociology*, 66(1), pp. 32–40.

Beer, D. (2009) 'Power through the algorithm? Participatory web cultures and the technological unconscious', *New Media & Society*, 11(6), pp. 985–1002.

Bell, D. (1964a) *The Coming of Post-Industrial Society*. New York: Basic Books.

Bell, D. (1964b) *The End of Ideology: On the Exhaustion of Political Ideas in the Fifties*. New York: Free Press.

Bender, E. M., Gebru, T., McMillan-Major, A. and Mitchell, M. (2021) 'On the dangers of stochastic parrots: Can language models be too big?', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21), pp. 610–623. doi:10.1145/3442188.3445922.

Bender, E.M. and Hanna, A. (2025) *The AI con: How to fight big tech's hype and create the future we want*. New York, NY: Harper Books.

Bengio, Y., LeCun, Y. and Hinton, G. (2021) ‘Deep learning for AI’, *Communications of the ACM*, 64(7), pp. 58–65. doi:10.1145/3448250.

Bennett, J. (2010) *Vibrant Matter: A Political Ecology of Things*. Durham, NC: Duke University Press.

Berger, P.L. and Luckmann, T. (1966) *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Doubleday.

Bijker, W.E. (1992) ‘The social construction of fluorescent lighting, or how an artifact was invented in its diffusion stage’, in Bijker, W.E. and Law, J. (eds.) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, pp. 75–102.

Bijker, W.E. and Law, J. (eds.) (1992) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.

Bimber, B. (1990) ‘Karl Marx and the three faces of technological determinism’, *Social Studies of Science*, 20(2), pp. 333–351.

Birch, K. (2013) ‘The political economy of technoscience: An emerging research agenda’, *Spontaneous Generations*, 7(1), pp. 72–85.

Birch, K. and Muniesa, F. (2020) *Assetization: Turning Things into Assets in Technoscientific Capitalism*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/12075

Bloor, D. (1984) ‘The sociology of reasons: Or why ‘epistemic factors’ are really ‘social factors’’, in Brown, J.R. (ed.) *Scientific Rationality: The Sociological Turn*. Dordrecht: Springer, pp. 295–324.

Blumer, H. (1969) *Symbolic Interactionism: Perspective and Method*. Englewood Cliffs, NJ: Prentice-Hall.

Blumer, H. (1954) ‘What is wrong with social theory?’, *American Sociological Review*, 19(1), pp. 3–10.

Bohner, J. and Vertesi, J. (2025) ‘Towards a socioeconomics of hype: Hype dynamics and symbolic boundary work within the speculative AI bubble’, *Social Science Computer Review,* 0(0), pp. 1–19. doi:10.1177/08944393251361935

Bommasani, R., et al. (2022) ‘On the opportunities and risks of foundation models’, *arXiv preprint,* arXiv:2108.07258.

Bowker Geoffrey C. (2016) ‘How Knowledge Infrastructures Learn.’ In *Infrastructures and Social Complexity*, edited by Harvey P., Jensen C., Morita A. Milton Park, UK: Routledge, 391–403.

Bowker, G.C. (1994) *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920–1940*. Cambridge, MA: MIT Press.

Bowker, G.C. (1993) 'How to be universal: Some cybernetic strategies, 1943–1970', *Social Studies of Science*, 23(1), pp. 107–127.

Bowker, G.C. and Star, S.L. (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bowker, G.C. and Star, S.L. (1996) 'How things (actor-net) work: Classification, magic and the ubiquity of standards', *Philosophia*, 25(1), pp. 195–220.

Bowker, G.C., Star, S.L., Turner, W. and Gasser, L. (eds.) (2015) *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Mahwah, NJ: Lawrence Erlbaum.

Bowker, G.C., Timmermans, S., Clarke, A.E. and Balka, E. (eds.) (2015) *Boundary Objects and Beyond: Working with Leigh Star*. Cambridge, MA: MIT Press.

boyd, d. and Crawford, K. (2012) 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, 15(5), pp. 662–679.

Brannigan, A. (1989) 'Artificial Intelligence and the Attributional Model of Scientific Discovery'. *Social Studies of Science*, *19*(4), 601–613.

Brewster, Z.W. and Puddephatt, A.J. (2016) 'George Herbert Mead as a socio-environmental thinker', in Joas, H. and Huebner, D.R. (eds.) *The Timeliness of George Herbert Mead*. Chicago, IL: University of Chicago Press, pp. 147–164.

Brontë, E. (1847) *Wuthering Heights*. London: Thomas Cautley Newby.

Brown, A.L., Davidovic, A. and Hasan, A. (2021) '*Abolitionist AI*. AI For the People'. arXiv:2101.02011. https://doi.org/10.48550/arXiv.2101.02011

Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional accuracy disparities in commercial gender classification', in *Proceedings of the Conference on Fairness, Accountability and Transparency (FAT* 2018)*. New York: ACM, pp. 77–91.

Bureau of Labor Statistics (2021a) *American Time Use Survey*. Washington, DC: U.S. Department of Labor.

Bureau of Labor Statistics (2021b) *American Time Use Survey—Technical Note*.

Burrell, J. (2016) 'How the machine 'thinks': Understanding opacity in machine learning algorithms', *Big Data & Society*, 3(1), pp. 1–12.

Busch, L. (2011) *Standards: Recipes for Reality*. Cambridge, MA: MIT Press.

Button, G. (ed.) (1993) *Technology in Working Order: Studies of Work, Interaction, and Technology*. London: Routledge.

Callon, M. (1994) 'Four models for the dynamics of science', in Jasanoff, S., Markle, G., Petersen, J. and Pinch, T. (eds.) *Handbook of Science and Technology Studies*. Thousand Oaks, CA: SAGE, pp. 29–63.

Callon, M. (1987) 'Society in the making: The study of technology as a tool for sociological analysis', in Hughes, T.P. and Pinch, T.J. (eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press, pp. 83–103.

Callon, M. (1986) 'Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay', in Law, J. (ed.) *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge, pp. 196–223.

Callon, M. and Latour, B. (1992) 'Don't throw the baby out with the Bath School! A reply to Collins and Yearley', in Pickering, A. (ed.) *Science as Practice and Culture*. Chicago, IL: University of Chicago Press, pp. 343–368.

Callon, M. and Latour, B. (1981) 'Unscrewing the Big Leviathan: How actors macro-structure reality and how sociologists help them to do so', in Knorr-Cetina, K. and Cicourel, A.V. (eds.) *Advances in Social Theory and Methodology*. London: Routledge & Kegan Paul, pp. 277–303.

Callon, M. and Law, J. (1995) 'Agency and the hybrid collectif', *South Atlantic Quarterly*, 94(2), pp. 481–507.

Callon, M., Law, J. and Rip, A. (eds.) (1986) *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London: Macmillan.

Cardon, D. (1998) 'Social science and co-operating machines: a bibliographic approach to CSCW', Réseaux. Communication – Technologie – Société, 16(92), pp. 129–166. doi:10.3406/reso.1998.2136.

Carusi, A., Hoel, A.S., Webmoor, T. and Woolgar, S. (eds.) (2015) *Visualization in the Age of Computerization*. New York: Routledge.

Castells, M. (1996) *The Rise of the Network Society*. Malden, MA: Blackwell.

Cefaï, D. (2016) 'Mead's moral and political philosophy', in Joas, H. and Huebner, D.R. (eds.) *The Timeliness of George Herbert Mead*. Chicago, IL: University of Chicago Press, pp. 165–183.

Cerulo, K.A. (2009) 'Nonhumans in social interaction', *Annual Review of Sociology*, 35, pp. 531–552.

Chang, J.Y. (2004) 'Mead's theory of emergence as a framework for multi-level sociological inquiry', *Symbolic Interaction*, 27(3), pp. 405–427. doi:10.1525/si.2004.27.3.405.

Charmaz, K. (2014) *Constructing Grounded Theory*. 2nd ed. London: SAGE Publications.

Clarke, A.E. and Fujimura, J.H. (1992) 'What Tools? Which Jobs? Why Right?', in Clarke, A.E. and Fujimura, J.H. (eds.) *The Right Tools for the Job: At Work in Twentieth-Century Life Sciences*. Princeton, NJ: Princeton University Press, pp. 3–44.

Clarke, A.E. and Fujimura, J.H. (eds.) (1992) *The Right Tools for the Job: At Work in Twentieth-Century Life Sciences*. Princeton, NJ: Princeton University Press.

Clarke, A.E. and Gerson, E.M. (1990) 'Symbolic interactionism in social studies of science', in Becker, H.S. and McCall, M.M. (eds.) *Symbolic Interaction and Cultural Studies*. Chicago, IL: University of Chicago Press, pp. 179–214.

Clarke, A.E. and Star, S.L. (2008) 'The social worlds framework: A theory/methods package', in Hackett, E.J., et al. (eds.) *The Handbook of Science and Technology Studies*. 3rd ed. Cambridge, MA: MIT Press, pp. 113–137.

Clarke, A.E., Friese, C. and Washburn, R.S. (2022) *Situational Analysis in Practice: Mapping Research with Grounded Theory*. 2nd ed. London: SAGE.

Collins, H. M. (1989) `Computers and the Sociology of Scientific Knowledge', *Social Studies of Science*, 19(4), 613-24.

Collins, H. M., and Evans, R. (2002) The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science*, *32*(2), 235-296.

Collins, H.M. (2021) 'The science of artificial intelligence and its critics', *Interdisciplinary Science Reviews*, 46(1–2), pp. 53–70. doi:10.1080/03080188.2020.1840821.

Collins, H.M. (2018) *Artifictional Intelligence: Against Humanity's Surrender to Computers*. Cambridge: Polity.

Collins, H.M. (1990) *Artificial experts: Social knowledge and intelligent machines*. Cambridge, MA and London: MIT Press.

Collins, H.M. (1983) 'The sociology of scientific knowledge: Studies of contemporary science', *Annual Review of Sociology*, 9, pp. 265–285.

Collins, H.M. and Pinch, T.J. (1982) *Frames of Meaning: The Social Construction of Extraordinary Science*. London: Routledge & Kegan Paul.

Collins, H.M. and Yearley, S. (1992) 'Epistemological chicken', in Pickering, A. (ed.) *Science as Practice and Culture*. Chicago, IL: University of Chicago Press, pp. 301–326.

Coopmans, C., Vertesi, J., Lynch, M. and Woolgar, S. (eds.) (2014) *Representation in Scientific Practice Revisited.* Cambridge, MA: MIT Press.

Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Crawford, K. and Joler, V. (2023) Calculating Empires: A Genealogy of Technology and Power Since 1500. (Accessed: 2 October 2025)

Crawford, K. and Paglen, T. (2019) 'Excavating AI: The politics of images in machine learning training sets', *Critical Inquiry*, 46(1), pp. 315–346.

Cybenko, G. (1989) 'Approximation by superpositions of a sigmoidal function', *Mathematics of Control, Signals and Systems*, 2(4), pp. 303–314.

Dahlin, E. (2023) 'And say the AI responded? Dancing around 'autonomy' in AI/human encounters'. *Social Studies of Science*, *54*(1), 59-77.

Daston, L. and Galison, P. (2007) *Objectivity*. New York: Zone Books.

de Laet, M. and Mol, A. (2000) 'The Zimbabwe Bush Pump: Mechanics of a fluid technology', *Social Studies of Science*, 30(2), pp. 225–263.

de Seta, G. (2024). 'Synthetic probes: A qualitative experiment in latent space exploration'. *Sociologica*, 18(2), 9–23.

de Seta, G. (2023) 'Digital depth: A volumetric speculation', *magazén: International Journal for Digital and Public Humanities*, 4(2), pp. 245–270. doi:10.30687/mag/2724-3923/2023/02/004

de Seta, G., Pohjonen, M., and Knuutila, A. (2024) 'Synthetic ethnography: Field devices for the qualitative study of generative models'. *Big Data & Society*, *11*(4). https://doi-org.ludwig.lub.lu.se/10.1177/20539517241303126

Dean, J. (2022) 'A golden decade of deep learning: Computing systems & applications', *Daedalus*, 151(2), pp. 58–74. Cambridge, MA: MIT Press.

Deng, J., Dong, W., Socher, R., Li, L-J., Li, K. and Li, F-F. (2009) 'ImageNet: A large-scale hierarchical image database', in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*. Miami, FL: IEEE, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

Denton, E., Hanna, A., Amironesei, R., et al. (2021) 'On the genealogy of machine learning datasets: A critical history of ImageNet', *Big Data & Society*, 8(2), pp. 1–15.

Denton, E., Hanna, A., Amironesei, R., et al. (2020) 'Bringing the people back in: Contesting benchmark datasets in machine learning', *arXiv preprint,* arXiv:2007.07399.

Desrosières, A. (1998) *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

DiMaggio, P., Hargittai, E., Neuman, W.R. and Robinson, J.P. (2001) 'Social implications of the Internet', *Annual Review of Sociology*, 27, pp. 307–336.

Dobson, J.E. (2023) *The Birth of Computer Vision*. Minneapolis: University of Minnesota Press.

Donoho, D. (2017) '50 years of data science', *Journal of Computational and Graphical Statistics*, 26(4), pp. 745–766. doi:10.1080/10618600.2017.1389735.

Dourish, P. (2001) *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press.

Dreyfus, H.L. (1979) *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row.

Drucker, P.F. (1993) *Post-Capitalist Society*. New York: Harper Business.

D'Ignazio, C. and Klein, L.F. (2020) *Data Feminism*. Cambridge, MA: MIT Press.

Emerson, R.M., Fretz, R.I. and Shaw, L.L. (2011) *Writing Ethnographic Fieldnotes*. 2nd ed. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226206868.001.0001.

Engdahl, I. (2025) 'Learning by Doing: An Ethnographic Account of Computer Vision Laboratory Work', Unpublished manuscript.

Engdahl, I. (2024) 'Agreements 'in the wild': Standards and alignment in machine learning benchmark dataset construction', *Big Data & Society*, 11(2). doi:10.1177/20539517241242457.

Engdahl, I. (2024) 'Wuthering Weights—Localisation Trajectories of Machine Learning Models for Local Ends', in Fors, V., Berg, M. and Brodersen, M. (eds.) *The De Gruyter Handbook of Automated Futures: Imaginaries, Interactions and Impact*. Berlin/Boston: De Gruyter, pp. 321–336. doi:10.1515/9783110792256-020.

Engelhardt, H.T. Jr. and Caplan, A.L. (1987) 'Introduction: Patterns of Controversy and Closure – The Interplay of Knowledge, Values, and Political Forces', in Engelhardt, H.T. Jr. and Caplan, A.L. (eds.) *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*. Cambridge: Cambridge University Press, pp. 1–24. doi:10.1017/CBO9780511628719.002.

Ensmenger, N. (2016) 'The Multiple Meanings of a Flowchart: The Visual Culture of Software', *Information & Culture*, 51(3), pp. 321–351. doi:10.7560/IC51302.

Espeland, W.N. and Stevens, M.L. (2008) 'A Sociology of Quantification', *European Journal of Sociology*, 49(3), pp. 401–436. doi:10.1017/S0003975609000150.

European Union (2024) 'Article 3: Definitions', Artificial Intelligence Act. (Accessed: 2 October 2025)

Farrell, H., and Fourcade, M. (2023) The Moral Economy of High-Tech Modernism. *Daedalus*, 152(1), 225–235.

Fine, G.A. and Kleinman, S. (1986) 'Interpreting the sociological classics: Can there be a 'true' meaning of Mead?', *Symbolic Interaction*, 9(1), pp. 129–146.

Fine, G.A. and Tavory, I. (2019) 'Interactionism at the crossroads: Then and now', *Journal for the Theory of Social Behaviour*, 49(4), pp. 458–471.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020) *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research Publication (2020-1).

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018) 'AI4People — an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations', *Minds and Machines*, 28(4), pp. 689–707.

Forman, L. (2023) 'AI Has Its 'iPhone Moment'', *Wall Street Journal*, March 2. (Accessed 2025-09-01)

Forsythe, D.E. (2001) *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence*. Stanford, CA: Stanford University Press.

Forsythe, D. E. (1993) 'Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence'. *Social Studies of Science*, 23(3), 445-477. Doi:10.1177/0306312793023003002

Fourcade, M. and Burrell, J. (2021) 'The sociology of algorithms', *Annual Review of Sociology*, 47, pp. 39–59.

Fourcade, M. and Healy, K. (2013) 'Classification situations: Life-chances in the neoliberal era', *Accounting, Organizations and Society*, 38(8), pp. 559–572.

Friedland, A. (2023) 'Meta's powerful language models leak online, the U.S. Commerce Department opens up applications for CHIPS funds, and a PRC reorganization will impact R&D and data', *Georgetown University Center for Security and Emerging Technology Newsletter*, March 9. (Accessed: September 3, 2025)

Fujimura, J.H. (1992) 'Crafting science: Standardized packages, boundary objects, and 'translation'', in Pickering, A. (ed.) *Science as Practice and Culture*. Chicago, IL: University of Chicago Press, pp. 168–211.

Fujimura, J.H. (1996) *Crafting Science. A Sociohistory of the Quest for the Genetics of Cancer*. Cambridge: Harvard University Press.

Fujimura, J.H. (1987) 'Constructing 'Do-able' problems in cancer research: Articulating alignment', *Social Studies of Science*, 17(2), pp. 257–293.

Fukushima, K. (1969) 'Visual feature extraction by a multilayered network of analog threshold elements', *IEEE Transactions on Systems Science and Cybernetics*, 5(4), pp. 322–333.

Gaboury, J. (2021) *Image Objects: An Archaeology of Computer Graphics*. Cambridge, MA: MIT Press.

Garfinkel, H. (1988) 'Evidence for locally produced, naturally accountable phenomena of order, logic, reason, meaning, method, etc. in and as of the essential quiddity of immortal ordinary society (I of IV): An announcement of studies', *Sociological Theory*, 6(1), pp. 103–109.

Gebru, T., Morgenstern, J., Vecchione, B., et al. (2021) 'Datasheets for datasets', *Communications of the ACM*, 64(12), pp. 86–92.

Geiger, R.S. and Ribes, D. (2011) 'Trace ethnography: Following coordination through documentary practice', in *Proceedings of the 44th Hawaii International Conference on System Sciences*. Honolulu: IEEE, pp. 1–10.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilbert, G.N. and Mulkay, M. (1984) *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. Cambridge: Cambridge University Press.

Gillespie, T. (2014) 'The relevance of algorithms', in Gillespie, T., Boczkowski, P.J. and Foot, K.A. (eds.) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167–194.

Gitelman, L. (ed.) (2013) *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press.

Glaser, B.G. and Strauss, A.L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.

Glaser, B.G., and Strauss, A.L. (1965) *Awareness of Dying*. Routledge.

Goffman, E. (1959) *The Presentation of Self in Everyday Life*. New York, NY: Anchor Books.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al. (2014) 'Generative adversarial nets', in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pp. 2672–2680.

Gottschalk, S. (2018) *The Terminal Self: Everyday Life in Hypermodern Times*. New York: Routledge.

Gray Widder, D., West, S. and Whitaker, M. (2023) 'Open (for business): Big tech, concentrated power, and the political economy of open AI', *SSRN Electronic Journal*.

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

Grint, K. and Woolgar, S. (1997) *The Machine at Work: Technology, Work and Organization*. Cambridge: Polity Press.

Hacking, I. (1999) *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.

Hacking, I. (1975) *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hansen, K. B. (2020) 'Model Talk: Calculative Cultures in Quantitative Finance'. *Science, Technology, & Human Values*, *46*(3), 600-627.

Haraway, D. (1989) *Primate Visions: Gender, Race, and Nature in the World of Modern Science*. New York: Routledge.

Haraway, D. (1988) 'Situated knowledges: The science question in feminism and the privilege of partial perspective', *Feminist Studies*, 14(3), pp. 575–599. doi:10.2307/3178066.

Heath, C., Knoblauch, H. and Luff, P. (2000) 'Technology and social interaction: The emergence of 'workplace studies'', *The British Journal of Sociology*, 51(2), pp. 299–320.

Hecht, G. (2002) 'Rupture-talk in the nuclear age: A review of the social studies of technology', *Social Studies of Science*, 32(5–6), pp. 691–727.

Heikkilä, M. (2023) 'Data-poisoning: Artists fight generative AI', *MIT Technology Review*, 23 October, 2023. (Accessed: 2 October 2025)

Henderson, K. (1991) 'Flexible sketches and inflexible data bases: Visual communication, conscription devices, and boundary objects in design engineering', *Science, Technology, & Human Values*, 16(4), pp. 448–473.

Hess, D.J. (2001) 'Ethnography and the development of science and technology studies', in Atkinson, P., et al. (eds.) *Handbook of Ethnography*. London: SAGE, pp. 234–245.

Hine, C. (2015) *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury Publishing.

Holstein, J.A. and Gubrium, J.F. (1995) *The Active Interview*. Thousand Oaks, CA: Sage.

Hughes, E. (1971) *The Sociological Eye*. Chicago: Aldine-Atherton.

Hughes, T.P. (1987) 'The evolution of large technological systems', in Bijker, W.E., Hughes, T.P. and Pinch, T.J. (eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press, pp. 51–82.

Hughes, T.P. (1983) *Networks of Power: Electrification in Western Society, 1880–1930*. Baltimore: Johns Hopkins University Press.

Inglehart, R.F. (2018) *Cultural Evolution: People's Motivations Are Changing, and Reshaping the World*. Cambridge: Cambridge University Press.

Jasanoff, Sheila and Sang-Hyun Kim. (2009) 'Containing the Atom: Sociotechnical Imaginaries and Nuclear Regulation in the U.S. and South Korea.' *Minerva* 47(2):119-146.

Jaton, F. (2023) 'Groundwork for AI: Enforcing a benchmark for neoantigen prediction in personalized cancer immunotherapy', *Social Studies of Science*, 53(5), pp. 787–810. doi:10.1177/03063127231192857.

Jaton, F. (2021) *The Constitution of Algorithms: Ground-Truthing, Programming, Formulating*. Cambridge, MA: MIT Press.

Jaton, F. (2017) 'We get the algorithms of our ground truths: Designing referential databases in digital image processing', *Social Studies of Science*, 47(6), pp. 811–840. doi:10.1177/0306312717730428.

Jo, E.S. and Gebru, T. (2020) 'Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning', in *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAT*)*. doi:10.48550/arXiv.1912.10389.

Joas, H. (1992) *The Creativity of Action*. Cambridge: Polity Press.

Joas, H. and Huebner, D.R. (eds.) (2016) *The Timeliness of George Herbert Mead*. Chicago, IL: University of Chicago Press.

Jung, J.Y., Steinberger, T. and So, C. (2023) 'Towards actionable data science: domain experts as end-users of data science systems'. *Computer Supported Cooperative Work (CSCW)*, pp.1–45.

Kalliamvakou, E., Singer, L., Gousios, G., German, D.M., Blincoe, K. and Damian, D. (2014) 'The promises and perils of mining GitHub', in *Proceedings of the 11th Working Conference on Mining Software Repositories* (MSR 2014). New York: ACM, pp. 92–101.

Kalluri, P. R., Agnew, W., Cheng, M., Owens, K., Soldaini, L. and Birhane, A. (2023) 'The Surveillance AI Pipeline', *arXiv preprint,* arXiv:2309.15084.

Kang, E.B. (2023) 'Ground truth tracings (GTT): On the epistemic limits of machine learning', *Big Data & Society*, 10(1).

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. and Amodei, D. (2020) 'Scaling Laws for Neural Language Models', *arXiv preprint,* arXiv:2001.08361.

Kitchin, R. (2014) 'Big Data, new epistemologies and paradigm shifts'. *Big Data & Society*, *1*(1).

Kling, R. (1991) ‘Computerization and Social Transformations‘. *Science, Technology, & Human Values*, *16*(3), 342-367. doi:10.1177/016224399101600304

Kling, R. (1978) ‘Patterns of segmentation and intersection in the computing world’, *Symbolic Interaction*, 1(2), pp. 24–43.

Kling, R., & Scacchi, W. (1982) ‘The Web of Computing: Computer Technology as Social Organization’. *Advances in Computers*, 21, 1-90.

Knorr, K.D. (1979) ‘Tinkering toward success: Prelude to a theory of scientific practice’, *Theory and Society*, 8(3), pp. 347–376. doi:10.1007/BF00167894.

Knorr-Cetina, K. (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Knorr-Cetina, K. (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Knuth, D.E. (1997) *The Art of Computer Programming, Volume 1: Fundamental Algorithms*. Reading, MA: Addison-Wesley.

Koch, B., Denton, E., Hanna, A. and Foster, J.G. (2021) ‘Reduced, reused and recycled: The life of a dataset in machine learning research’, *arXiv preprint,* arXiv:2112.01716. doi:10.48550/arXiv.2112.01716.

Kohn, E. (2013) *How Forests Think: Toward an Anthropology Beyond the Human*. Berkeley, CA: University of California Press.

Koselleck, R. (1982) ‘Begriffsgeschichte and social history’, *Economy and Society*, 11(4), pp. 409–427. doi:10.1080/03085148200000015.

Kreibich, R. (1986) *Die Wissenschaftsgesellschaft: Von Galilei zur High-Tech-Revolution*. Frankfurt am Main: Suhrkamp.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ‘ImageNet classification with deep convolutional neural networks’, in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105.

Kruse, C. (2021) ‘Containing the stable movement of knowledge objects through the Swedish criminal justice system: Thinking with infrastructure’, *Science & Technology Studies*, 34(1), pp. 2–18.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kunda, G. (1992) *Engineering Culture: Control and Commitment in a High-Tech Corporation*. Philadelphia, PA: Temple University Press.

Kwet, M. (2021) ‘Digital colonialism: The evolution of US empire’, in Guss, D.M. and Tu, S.H.-C. (eds.) *The Palgrave Handbook of the Anthropology of Technology*. Singapore: Palgrave Macmillan, pp. 143–162.

LaGrandeur, K. (2024) 'The consequences of AI hype', *AI and Ethics*, 4(3), pp. 653–656. doi:10.1007/s43681-023-00352-y

Lampland, M. and Star, S.L. (eds.) (2009) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca, NY: Cornell University Press.

Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, B. (1999) 'On recalling ANT', in Law, J. and Hassard, J. (eds.) Actor-network theory and after. Oxford: Blackwell, pp. 15–25.

Latour, B. (1995b) 'The "Pedofil" of Boa Vista: A Photo-Philosophical Montage', *Common Knowledge*, 4(1), pp. 145–187.

Latour, B. (1995a) 'Social Theory and the Study of Computerized Work Sites', in Orlikowski, W. J. and Walsham, G. (eds.) *Information Technology and Changes in Organizational Work*. London: Chapman and Hall, pp. 295–307.

Latour, B. (1994) 'On Technical Mediation'. *Common Knowledge* 3(2): 29-64

Latour, B. (1993) *We Have Never Been Modern*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press.

Latour, B. (1992) 'Where are the missing masses? The sociology of a few mundane artifacts'. In W. E. Bijker and J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 225–258). MIT Press.

Latour, B. (1990) 'Drawing things together', in Lynch, M. and Woolgar, S. (eds.) *Representation in Scientific Practice*. Cambridge, MA: MIT Press, pp. 19–68.

Latour, B. (1988) *The Pasteurization of France*. Cambridge, MA: Harvard University Press.

Latour, B. (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Latour, B. and Woolgar, S. (1979) *Laboratory Life: The Construction of Scientific Facts*. Beverly Hills, CA: Sage.

Latour, B., Mauguin, P. and Teil, G. (1992) 'A note on socio-technical graphs', *Social Studies of Science*, 22(1), pp. 33–57.

Law, J. (1987) 'Technology and heterogeneous engineering: The case of Portuguese expansion', in Bijker, W.E., Hughes, T.P. and Pinch, T.J. (eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press, pp. 111–134.

Law, J. (1987) 'The Structure of Sociotechnical Engineering — A Review of the New Sociology of Technology'. *The Sociological Review*, *35*(2), 404-425.

Law, J. and Mol, A. (2001) 'Situating technoscience: An inquiry into the spatialities of technoscientific practice', *Environment and Planning D: Society and Space*, 19(5), pp. 609–621.

Law, J. and Singleton, V. (2005) 'Object lessons', *Organization*, 12(3), pp. 331–355.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', Nature, 521(7553), pp. 436–444. doi:10.1038/nature14539.

LeCun, Y., Bottou, L., Orr, G.B. and Müller, K-R. (1995) 'Convolutional networks for images, speech, and time series', in Arbib, M.A. (ed.) *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press, pp. 255–258.

Lee, F., and Björklund Larsen, L. (2019) 'How should we theorize algorithms? Five ideal types in analyzing algorithmic normativities'. *Big Data & Society*, 6(2). https://doi-org.ludwig.lub.lu.se/10.1177/2053951719867349

Lee, F., Bier, J., Christensen, J., Engelmann, L., Helgesson, C.-F., and Williams, R. (2019) 'Algorithms as folding: Reframing the analytical focus'. *Big Data & Society*, *6*(2).

Lewis, J.D. (1981) 'G.H. Mead's Contact Theory of Reality: The Manipulatory Phase of the Act in the Constitution of Mundane, Scientific, Aesthetic, and Evaluative Objects'. *Symbolic Interaction*, 4: 129-141.

Li, F.-F. (2023) *The Worlds I See: Curiosity, Exploration, and Discovery at the Dawn of AI*. New York: Flatiron Books.

Li, F.-F. and Krishna, R. (2022) 'Searching for Computer Vision North Stars', *Daedalus*, 151(2), pp. 85–99. doi:10.1162/daed_a_01902.

Lindén, L. and Lydahl, D. (2021) 'Editorial: Care in STS', *Nordic Journal of Science and Technology Studies*, 9(1). doi: 10.5324/njsts.v9i1.4000.

Loukissas, Y.A. (2019) *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

Lynch, M. (2024) 'Laboratory studies and scientific practice', in Felt, U. and Irwin, A. (eds.) *Elgar Encyclopedia of Science and Technology Studies*. Cheltenham: Edward Elgar, pp. 273–282.

Lynch, M. (1991) 'Laboratory space and the technological complex: An investigation of topical contextures', *Science in Context*, 4(1), pp. 51–78.

Lynch, M. (1985) *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. London: Routledge & Kegan Paul.

Lynch, M. and Woolgar, S. (eds.) (1990) *Representation in Scientific Practice*. Cambridge, MA: MIT Press.

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

MacKenzie, D. and Wajcman, J. (eds.) (1999) *The Social Shaping of Technology*. 2nd ed. Buckingham: Open University Press.

Malik, J., Arbeláez, P., Carreira, J., Fragkiadaki, K., Girshick, R., Gkioxari, G., Gupta, S., Hariharan, B., Kar, A. and Tulsiani, S. (2016) 'The Three R's of Computer Vision: Recognition, Reconstruction and Reorganization', *Pattern Recognition Letters*, 72, pp. 4-14

Manovich, L. (2001) *The Language of New Media*. Cambridge, MA: MIT Press.

Marcus, G.E. (1995) 'Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography', *Annual Review of Anthropology*, 24, pp. 95–117. doi:10.1146/annurev.an.24.100195.000523.

Marres, N., Castelle, M., Gobbo, B., Poletti, C. and Tripp, J. (2024) 'AI as super-controversy: Eliciting AI and society controversies with an extended expert community in the UK', *Big Data & Society*, 9(4).

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Betts Lotufo, J., Rome, A., Shi, A. & Oak, S. (2025) 'Artificial Intelligence Index Report 2025'. *arXiv preprint,* arXiv:2504.07139.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.

McCosker, A., and Wilken, R. (2020) *Automating vision: The social impact of the new camera consciousness*. Routledge.

McDermott. D. (1976) 'Artificial intelligence meets natural stupidity'. ACM SIGART Bulletin, (57):4–9

McDonnell, T.E. (2023) 'Cultural objects, material culture, and materiality', *Annual Review of Sociology*, 49, pp. 195–220. doi:10.1146/annurev-soc-031021-041439.

Mead, G.H. (1962) *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Chicago, IL: University of Chicago Press.

Mead, G.H. (1938) *The Philosophy of the Act*. Edited by C.W. Morris. Chicago, IL: University of Chicago Press.

Mead, G.H. (1932) *The Philosophy of the Present*. Chicago, IL: University of Chicago Press.

Merton, R.K. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.

Mitchell, M., Luccioni, A.S., Lambert, N., et al. (2023) 'Measuring Data', *arXiv preprint,* arXiv:2212.05129. doi:10.48550/arXiv.2212.05129.

Mok, C. (2025) 'Taking Stock of the DeepSeek Shock', *Freeman Spogli Institute – Cyber Policy Center (Stanford University)*, February 5.

Mol, A. (2024) 'Bernal Lecture keynote', talk given at EASST/4S 2024, Amsterdam, 17 July.

Mol, A. (2001) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.

Morris, C.W. (1938) 'Introduction', in Mead, G.H. *The Philosophy of the Act*. Chicago, IL: University of Chicago Press, pp. i–lxxv.

Mulkay, M. (1979) *Science and the Sociology of Knowledge*. London: Allen & Unwin.

Narayanan, A. and Kapoor, S. (2025) 'What if artificial intelligence is just a "normal" technology?', The Economist, 4 September. (Accessed: 2025-09-06).

Narayanan, A. and Kapoor, S. (2024) *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton, NJ: Princeton University Press.

Nassar, J., Pavon-Harr, V., Bosch, M. and McCulloh, I. (2019) 'Assessing Data Quality of Annotations with Krippendorff Alpha for Applications in Computer Vision', *arXiv preprint,* arXiv:1912.10107. doi:10.48550/arXiv.1912.10107.

Neyland, D. (2019) *The Everyday Life of an Algorithm*. Cham: Palgrave Pivot.

Noble, D.F. (1979) *America by Design: Science, Technology, and the Rise of Corporate Capitalism*. Oxford: Oxford University Press.

Noble, S.U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Nora, S. and Minc, A. (1980) *The Computerization of Society: A Report to the President of France*. Cambridge, MA: MIT Press.

Olazaran, M. (1996) 'A sociological study of the official history of the perceptrons controversy', *Social Studies of Science*, 26(3), pp. 611–659.

Orr, J.E. (1996) *Talking About Machines: An Ethnography of a Modern Job*. Ithaca, NY: Cornell University Press.

O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Paglen, T. (2016) 'Invisible Images (Your Pictures Are Looking at You)', *The New Inquiry*, 8 December.

Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Patel, D. (2025) *The Scaling Era: An Oral History of AI, 2019–2025*. San Francisco: Stripe Matter Inc.

Petrick, E.R. (2019) 'Building the Black Box: Cyberneticians and Complex Systems', *Science, Technology & Human Values*, 45(4), pp. 575–595.

Pfotenhauer, S. M., Juhl, J. and Aarden, E. (2019) 'Challenging the "deficit model" of innovation: Framing policy issues under the innovation imperative', Research Policy, 48(4), pp. 895–904. doi:10.1016/j.respol.2018.10.015.

Pickering, A. (1995) *The Mangle of Practice: Time, Agency, and Science*. Chicago, IL: University of Chicago Press.

Pickering, A. (ed.) (1992) *Science as Practice and Culture*. Chicago, IL: University of Chicago Press.

Pinch, T.J. (1993) '"Testing—One, Two, Three… Testing!': Toward a sociology of testing', *Science, Technology & Human Values*, 18(1), pp. 25–41.

Pinch, T.J. and Bijker, W.E. (1987) 'The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other', in Bijker, W.E., Hughes, T.P. and Pinch, T.J. (eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press, pp. 17–50.

Pinch, T. J., and Bijker, W. E. (1984) 'The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other'. *Social Studies of Science*, 14(3), 399-441.

Pink, S. and Morgan, J. (2013) 'Short-term ethnography: Intense routes to knowing', *Symbolic Interaction*, 36(3), pp. 351–361.

Plizzari, C., Goletto, G., Furnari, A., et al. (2023) 'An outlook into the future of egocentric vision', *arXiv preprint,* arXiv:2308.07123. doi:10.48550/arXiv.2308.07123.

Porter, T.M. (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Postill, J. and Pink, S. (2012) 'Social media ethnography: The digital researcher in a messy web', *Media International Australia*, 145(1), pp. 123–134.

Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Prasad, A. and Shadnam, M. (2023) 'Balancing breadth and depth in qualitative research: Conceptualizing performativity through multi-sited ethnography', *Organization Studies*, 44(5), pp. 739–761. doi:10.1177/01708406221145655.

Puddephatt, A. (2017) ‘George Herbert Mead: The evolution of mind, self, and society through interaction’, in Jacobsen, M. H. (ed.) *The interactionist imagination*. Basingstoke: Palgrave Macmillan, pp. 95–119.

Puddephatt, A.J. (2005) ‘Mead has never been modern: Using Meadian theory to extend the constructionist study of technology’, *Social Epistemology*, 19(4), pp. 357–380.

Radder, H. (1992) ‘Normative Reflexions on Constructivist Approaches to Science and Technology‘, *Social Studies of Science*, *22*(1), 141-173.

Raji, I.D., Bender, E.M., Paullada, A., Denton, E. and Hanna, A. (2021) ‘AI and the Everything in the Whole Wide World Benchmark’, *arXiv preprint*, arXiv:2111.15366. doi:10.48550/arXiv.2111.15366.

Ribes, D. (2019) ‘How I learned what a domain was’, *Proceedings of the ACM on Human–Computer Interaction*, 3(CSCW), pp. 1–12.

Ribes, D. (2019) ‘STS, meet data science, once again’, *Science, Technology & Human Values*, 44(3), pp. 514–539.

Ribes, D., Hoffman, A.L., Slota, S.C. and Bowker, G.C. (2019) ‘The logic of domains’, *Social Studies of Science*, 49(3), pp. 281–309.

Ribes, D. and Monteiro, E. (in press) ‘More instrument than data’, *Science, Technology, & Human Values.*

Richta, R. (ed.) (1969) *Civilization at the Crossroads: Social and Human Implications of the Scientific and Technological Revolution*. White Plains, NY: International Arts and Sciences Press.

Roberge, J. and Castelle, M. (2021) *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Palgrave Macmillan. doi:10.1007/978-3-030-56286-1.

Roberts, L.G. (1963) *Machine Perception of Three-Dimensional Solids*. PhD Thesis, Massachusetts Institute of Technology.

Rodu, J. and Baiocchi, M. (2023) ‘When black box algorithms are (not) appropriate’, *Observational Studies*, 9(2), pp. 79–101.

Rosenblatt, F. (1957) *The Perceptron: A Perceiving and Recognizing Automaton*. Ithaca, NY: Cornell Aeronautical Laboratory.

Sambasivan, N., et al. (2021) ‘Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI’, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.

Samuel, A.L. (1959) ‘Some studies in machine learning using the game of checkers’, *IBM Journal of Research and Development*, 3(3), pp. 210–229.

Scheuerman, M.K., Denton, E. and Hanna, A. (2021) 'Do datasets have politics? Disciplinary values in computer vision dataset development', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), pp. 1–37.

Scheuerman, M.K., Wade, K., Lustig, C., et al. (2020) 'How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis', *Proceedings of the ACM on Human-Computer Interaction*, 4(1), pp. 1–35.

Schiller, D. (1999) *Digital Capitalism: Networking the Global Market System*. Cambridge, MA: MIT Press.

Schuilenburg, M. and Peeters, R. (2020) *The Algorithmic Society: Technology, Power, and Knowledge*. London: Routledge.

Seaver, N. (2022) *Computing Taste: Algorithms and the Makers of Music Recommendation*. Chicago, IL: University of Chicago Press.

Seaver, N. (2021) 'What is patchwork ethnography?', *Qualitative Communication Research*, 10(1), pp. 1–21.

Seaver, N. (2017) 'Algorithms as culture: Some tactics for the ethnography of algorithmic systems', *Big Data & Society*, 4(2).

Sevilla, J., Heim, L., Ho, A., et al. (2022) 'Compute trends across three eras of machine learning', *arXiv preprint,* arXiv:2202.05924. doi:10.48550/arXiv.2202.05924.

Shankar, S., Halpern, Y., Breck, E., et al. (2017) 'No classification without representation: Assessing geodiversity issues in open data sets for the developing world', in *Proceedings of the 2017 NIPS Workshop on Machine Learning for the Developing World*.

Shapin, S. and Schaffer, S. (1985) *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.

Shibutani, T. (1955) 'Reference groups as perspectives', *American Journal of Sociology*, 60(6), pp. 562–569.

Silverman, D. (2013) *Doing Qualitative Research: A Practical Handbook*. 4th ed. London: SAGE Publications.

Simon, H.A. (1996) *The Sciences of the Artificial*. 3rd ed. Cambridge, MA: MIT Press.

Sismondo, S. (2010) *An Introduction to Science and Technology Studies*. 2nd ed. Chichester: Wiley-Blackwell.

Slezak, P. (1989) 'Scientific discovery by computer as empirical refutation of the Strong Programme', *Social Studies of Science*, 19(4), pp. 563–600.

Sloane, M., Danks, D. and Moss, E. (2024) 'Tackling AI hyping', *AI and Ethics*, 4, pp. 669–677. doi:10.1007/s43681-024-00481-y

Smits, M. (2001) Langdon Winner: Technology as a shadow constitution. In H. Achterhuis (Ed.), *American Philosophy of Technology*. Bloomington: Indiana University Press.

Sohn-Rethel, A. (1978) *Intellectual and Manual Labour: A Critique of Epistemology*. London: Macmillan.

Srnicek, N. (2017) *Platform Capitalism*. Cambridge: Polity Press.

Star, S.L. (2010) 'This is not a boundary object: Reflections on the origin of a concept', *Science, Technology & Human Values*, 35(5), pp. 601–617.

Star, S.L. (1999) 'The ethnography of infrastructure', *American Behavioral Scientist*, 43(3), pp. 377–391.

Star, S.L. (1995) 'The politics of formal representations: Wizards, gurus, and organizational complexity', in Star, S.L. (ed.) *Ecologies of Knowledge*. Albany, NY: SUNY Press, pp. 88–118.

Star, S.L. (1991) 'Power, technology and the phenomenology of conventions: On being allergic to onions', in Law, J. (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp. 26–56.

Star, S.L. (1989) 'The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving', in Gasser, L. and Huhns, M.N. (eds.) *Distributed Artificial Intelligence, Vol. 2*. London: Morgan Kaufmann, pp. 37–54.

Star, S.L. (1985) 'Scientific work and uncertainty', *Social Studies of Science*, 15(2), pp. 391–427.

Star, S.L. and Gerson, E.M. (1987) 'The management and dynamics of anomalies in scientific work', *The Sociological Quarterly*, 28(2), pp. 147–169.

Star, S.L. and Griesemer, J.R. (1989) 'Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39', *Social Studies of Science*, 19(3), pp. 387–420.

Star, S.L. and Strauss, A. (1999) 'Layers of silence, arenas of voice: The ecology of visible and invisible work', *Computer Supported Cooperative Work (CSCW)*, 8(1–2), pp. 9–30.

Stehr, N. (2022) *Knowledge Capitalism*. New York: Routledge.

Strauss, A. (1993) *Continual Permutations of Action*. New York: Aldine de Gruyter.

Strauss, A. (1985) 'Work and the division of labour', *The Sociological Quarterly*, 26(1), pp. 1–19.

Strauss, A. (1978) 'A social world perspective', *Studies in Symbolic Interaction*, 1, pp. 119–128.

Strauss, A. and Corbin, J. (1994) ‘Grounded theory methodology’, in Denzin, N.K. and Lincoln, Y.S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, pp. 273–285.

Strauss, A. and Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.

Striphas, T. (2015) ‘Algorithmic culture’, *European Journal of Cultural Studies*, 18(4–5), pp. 395–412.

Strübing, J. (2017) ‘Anselm L. Strauss – Action/Work as Process and Perspective‘. in Jacobsen, M. H. (ed.) *The interactionist imagination*. Basingstoke: Palgrave Macmillan, pp. 263–289.

Strübing, J. (1998) ‘Bridging the Gap: On the Collaboration between Symbolic Interactionism and Distributed Artificial Intelligence in the Field of Multi-Agent Systems Research’. *Symbolic Interaction*, *21*(4), 441–463.

Suchman, L. (2023) ‘The uncontroversial ‘thingness’ of AI‘. *Big Data & Society*, 10(2). doi:10.1177/20539517231206794

Suchman, L. (2022) ‘Imaginaries of omniscience: Automating intelligence in the US Department of Defense‘. *Social Studies of Science*, *53*(5), 761-786.

Suchman, L. and Trigg, R.H. (1993) ‘Artificial intelligence as craftwork’, in Chaiklin, S. and Lave, J. (eds.) *Understanding Practice: Perspectives on Activity and Context*. Cambridge: Cambridge University Press, pp. 144–178.

Suchman, L.A. (1987) *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Sutton, R. (2019) ‘The bitter lesson’, *Blog post*, 13 March. Available at: https://www.incompleteideas.net/IncIdeas/BitterLesson.html

Sveningsson Elm, M. (2009) ‘How do various notions of privacy influence decisions in qualitative research?’, *Research Ethics*, 5(3), pp. 94–103.

Szeliski, R. (2010) *Computer Vision: Algorithms and Applications*. London: Springer.

Tavory, I. (2020) ‘Interviews and inference: Making sense of interview data in qualitative research’, *Qualitative Sociology*, 43(4), pp. 449–465. doi:10.1007/s11133-020-09464-6.

Tavory, I. and Timmermans, S. (2014) *Abductive Analysis: Theorizing Qualitative Research*. Chicago, IL: University of Chicago Press.

Thakkar, D., Ismail, A., Kumar, P., et al. (2022) ‘When is machine-learning data good?: Valuing in public-health datafication’, in *CHI Conference on Human Factors in Computing Systems*. New Orleans, LA, pp. 1–16.

The Nobel Prize. (2024) 'The Nobel Prize in Physics 2024 – Press Release', *NobelPrize.org*. (Accessed: 2 October 2025)

Thomas, W.I. and Thomas, D.S. (1928) The child in America: Behavior problems and programs. New York: Knopf, pp. 571–572.

Thrift, N. (2005) *Knowing Capitalism*. London: Sage.

Thylstrup, N.B. (2022) 'The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains', *Media, Culture & Society*, 44(4), pp. 655–671.

Timmermans, S. and Epstein, S. (2010) 'A world of standards but not a standard world: Toward a sociology of standards and standardization', *Annual Review of Sociology*, 36, pp. 69–89.

Timmermans, S. and Tavory, I. (2022) *Data Analysis in Qualitative Research: Theorizing with Abductive Analysis*. Chicago, IL: University of Chicago Press.

Timmermans, S. and Tavory, I. (2012) 'Theory construction in qualitative research: From grounded theory to abductive analysis', *Sociological Theory*, 30(3), pp. 167–186.

Torralba, A., Isola, P. and Freeman, W.T. (2024) *Foundations of Computer Vision*. Cambridge, MA: MIT Press.

Traweek, S. (1988) *Beamtimes and Lifetimes: The World of High-Energy Physicists*. Cambridge, MA: Harvard University Press.

Turkle, S. (2009) *Simulation and Its Discontents*. Cambridge, MA: MIT Press.

Turner, A. (2019) 'Using data from Git and GitHub in ethnographies of software development', in Öhman, C. and Watson, D. (eds.) *The 2018 Yearbook of the Digital Ethics Lab*. Cham: Springer International Publishing, pp. 35–49.

Turner, S.P. (1994) *The Social Theory of Practices: Tradition, Tacit Knowledge, and Presuppositions*. Chicago, IL: University of Chicago Press.

Unruh, D.R. (1980) 'The nature of social worlds', *Pacific Sociological Review*, 23(3), pp. 271–296.

van den Broek, E., Sergeeva, A. and Huysman, M. (2021) 'On the sociotechnical nature of machine learning: An illustration in the context of job-matching', *Big Data & Society*, 8(2).

van den Scott, L.-J. K., Sanders, C. B., and Puddephatt, A. J. (2017) 'Reconceptualizing users through enriching ethnography'. In R. Fouché, U. Felt, C. A. Miller, and L. Smith-Doerr (Eds.), *The Handbook of Science and Technology Studies* (4th ed., pp. 501–528). MIT Press.

van der Vlist, F., Helmond, A., and Ferrari, F. (2024) 'Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence'. *Big Data & Society*, *11*(1).

van Dijk, J. (1991) *De Netwerkmaatschappij*. Houten: Bohn Stafleu van Loghum.

Varshney, L.R., Keskar, N.S. and Socher, R. (2019) 'Pretrained AI models: Performativity, mobility, and change', *arXiv preprint,* arXiv:1909.03290.

Varoquaux, G., Luccioni, S. and Whittaker, M. (2025) 'Hype, sustainability, and the price of the bigger-is-better paradigm in AI', *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '25), 23–26 June, Athens, Greece. New York, NY: ACM, pp. 1–15. doi:10.1145/3715275.3732006

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) 'Attention is all you need', *arXiv preprint,* arXiv:1706.03762. doi:10.48550/arXiv.1706.03762.

Vercellone, C. (2007) 'From formal subsumption to general intellect: Elements for a Marxist reading of the thesis of cognitive capitalism', *Historical Materialism*, 15(1), pp. 13–36.

Vertesi, J. (2016) 'From the 'new' sociology of technology to the new sociology of technology?', *Blog post*, 21 April.

Vertesi, J. (2015) *Seeing Like a Rover: How Robots, Teams, and Images Craft Knowledge on Mars*. Chicago, IL: University of Chicago Press.

Vertesi, J. (2014b) 'Seamful spaces: Heterogeneous infrastructures in interaction', *Science, Technology & Human Values*, 39(2), pp. 264–284.

Vertesi, J., Ribes, D., DiSalvo, C., et al. (2019) *digitalSTS: A Field Guide for Science & Technology Studies*. New Haven: Princeton University Press.

Vincenti, W.G. (1990) *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*. Baltimore, MD: Johns Hopkins University Press.

Vipra, J., and West, S. M. (2023). 'Computational Power and AI'. AI Now Institute.

Wajcman, J. (2002) 'Addressing Technological Change: The Challenge to Social Theory'. *Current Sociology*, *50*(3), 347-363.

Webster, F. (2002) *Theories of the Information Society*. 2nd ed. London: Routledge.

Whalley, P. (1986) *The Social Production of Technical Work: The Case of British Engineers*. Albany, NY: SUNY Press.

Winner, L. (1986) *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.

Winner, L. (1980) 'Do artifacts have politics?', *Daedalus*, 109(1), pp. 121–136.

Winner, L. (1977) *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA: MIT Press.

Wirth, N. (1976) *Algorithms + Data Structures = Programs*. Englewood Cliffs, NJ: Prentice Hall.

Woolgar, S. (1991) 'Configuring the user: The case of usability trials', in Law, J. (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp. 57–99.

Woolgar, S. (1991) 'The turn to technology in social studies of science'. *Science, Technology & Human Values*, 16(1), 20-50.

Woolgar, S. (1989) 'A Coffeehouse Conversation on the Possibility of Mechanizing Discovery and Its Sociological Analysis'. *Social Studies of Science*, *19*(4), 658–668.

Woolgar, S. (1988) *Science: The Very Idea*. London: Tavistock.

Woolgar, S. (1985) 'Why not a sociology of machines? The case of sociology and artificial intelligence', *Sociology*, 19(4), pp. 557–572.

Woolgar, S., & Lezaun, J. (2013) 'The wrong bin bag: A turn to ontology in science and technology studies?'. *Social Studies of Science*, 43(3), 321–340.

Woolgar, S. and Neyland, D. (2013) *Mundane Governance: Ontology and Accountability*. Oxford: Oxford University Press.

Yang, F., Zamzmi, G., Angara, S., et al. (2023) 'Assessing inter-annotator agreement for medical image segmentation', *IEEE Access*, 11, pp. 21300–21312.

Ziewitz, M. (2016) 'Governing algorithms: Myth, mess, and methods', *Science, Technology, & Human Values*, 41(1), pp. 3–16.

Ziewitz, M. and Lynch, M. (2018) 'It's important to go to the laboratory: Malte Ziewitz talks with Michael Lynch', *Engaging Science, Technology, and Society*, 4, pp. 343–365. doi:10.17351/ests2018.220.

Zwart, S. (2022) 'Engineering Epistemology: Between Theory and Practice'. *Engineering Studies*, *14*(2), 79–86.

Zylinska, J. (2023) *The Perception Machine: Our Photographic Future between the Eye and AI*. Cambridge, MA: MIT Press.

Appendix

Table 2. Interviews

#	Name	Date & Format	Affiliation	Themes
1	Lisa	2022-01-10, Online (43 m)	National Research Center	Organization, dataset documentation
2	Ben	2022-02-15, Online (1 h 3 m)	National Research Center; Academic Lab	Data annotation, computer vision
3	Ben	2022-05-06, Online (1 h 13 m)	National Research Center; Academic Lab	Data annotation, computer vision
4	Yevgeny	2022-05-31, Online (41 m)	Tech Company	Datasets, machine learning, backpropagation
5	Cassandra	2022-06-15, On-site (1 h 30 m)	National Research Institute	AI in healthcare, data privacy, developer ecosystem and tooling
6	Patrick	2022-08-29, On-site (58 m)	National Research Institute	Computer vision, drones, reinforcement learning
7	Ken	2022-09-14, Online (25 m)	Big Tech	Dataset, benchmarks
8	Jane	2022-09-30, Online (25 m)	Academic Lab	Dataset, annotation, legal, benchmarks
9	Patrick	2022-10-03, On-site (4 h)	National Research Institute	Computer vision, programming, object detection
10	Sara	2022-10-04, On-site (52 m)	Tech Company	Computer vision, model training, developer ecosystem and tooling, Saas
11	Tom	2022-12-01, Online (1 h 7m)	Academic Dataset Team	Computer vision, benchmark, data annotation
12	Patrick	2023-06-28, Online + screen go-along (2 h)	National Research Institute	Machine learning, model evaluation
13	Stewart	2023-10-26, Online (90 m)	Developer Community	Three-dimensional reconstruction, neural radiance fields
14	Simon	2023-10-26, Online (1 h)	Startup Company	Neural radiance fields, data, startup culture
15	Jack	2023-11-01, On-site (30 m)	Visiting Professor	Multi-agents in virtual reality, large language models
16	Carl	2023-11-02, On-site (30 m)	MedTech Founder	Computer vision, technical dependencies, startup
17	Karl	2023-12-04, Online (55 m)	PhD, Tech Company	Computer vision, dataset, models

#	Name	Date & Format	Affiliation	Themes
18	Maria	2023-12-06, Online (40 m)	PhD Student	Foundation models, lab organization
19	Stefan	2024-10-03, Online (1 h)	PhD Student	Computer vision, lidar, diffusion models
20	Hyeonju	2024-10-04, On-site, 2×30 min	Professor	Generative AI, computer vision, scaling, superintelligence
21	Mira	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
22	Jianyu	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
23	Liang	2024-10-04—2024-10-31, On-site (30)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
24	Amahle	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
25	Sean	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
26	Sean	2024-10-04—2024-10-31, On-site (1 h)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
27	Sean	2024-10-04—2024-10-31, On-site (1 h)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
28	Logan	2024-10-04—2024-10-31, On-site (1 h)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
29	Seokjin	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
30	Jihwan	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
31	Zhenghao	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
32	Dylan	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
33	Kyle	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)

#	Name	Date & Format	Affiliation	Themes
34	Mason	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
35	Charlotte	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)
36	Isaac	2024-10-04—2024-10-31, On-site (30 min)	PhD Student	Various computer vision techniques (NeRFs, diffusion models, self-supervised learning, reinforcement learning, scaling)

Table 3. Events

#	Event	Date & Format	Affiliation	Themes
1	Academic Seminar	2022-10-13, On-site (53 m)	Academic Lab	Machine learning, object detection
2	Public Event	2022-11-09, On-site (2 h)	Academic Panel	AI in healthcare
3	Academic Seminar	2022-09-22, Online (53 m)	Academic Lab	Computer vision, machine learning, ecology
4	Academic Lab	2022-10-06, On-site (1 day)	Academic Lab	Computer vision
5	Academic Workshop	2022-11-21, On-site (1 week)	Academic Lab	Computer vision, machine learning, Python
6	Digital Fieldwork	2021–2022, Online (recurring)	Consortium	Dataset materials incl. annotation guidelines
7	Academic Conference	2023-03-13–15, On-site	Academic and Industry	Deep learning, computer vision, image processing, sensors
8	Academic Seminar	2023-05-11, Online (1 h)	Academic Lab	Computer vision, neural radiance fields, remote sensing, environment
9	Academic Seminar	2023-05-25, Online, Seminar (1 h)	Academic Lab	Computer vision, satellite data, wildfire, object detection
10	AI & Industry Panel	2023-08-31, On-site (2 h)	Academic and Industry	Artificial intelligence, computer vision, business
11	Academic Seminar	2023-09-14, Online (1 h)	Research Scientist	Federated learning, large-scale machine learning
12	Real-World AI Panel	2023-09-14, On-site (2 h)	Academic and Industry	Artificial intelligence, computer vision, medtech
13	Academic Seminar	2023-09-14, Online (1 h)	Research Scientist	Federated learning, large-scale machine learning
14	AI Office Hours	2023-10-23, On-site (3 h)	Academic and Industry	Networking, fundraising, startups
15	Public Conference	2023-10-24, On-site (1 day)	Academic and Industry	Foundation models, computer vision, three-dimensional reconstruction
16	Demo & Panel Day	2023-10-28, On-site (1 day)	Academic and Industry	Computer vision, three-dimensional reconstruction, Generative AI, startups
17	Keynote	2023-10-30, Online (2 h)	Big Tech	Tech Founder, Generative AI, multimodality
18	Panel	2023-10-31, On-site (2 h)	Academic, Public Administration, and Industry	National computing infrastructure, computer vision, politics
19	Computer History Museum	2023-11-01, Guided Tour (1 h)	Museum	Computers, hardware, history
20	Demo	2023-11-02, Online (1 h)	Big Tech	Object detection, model tutorial
21	Hackathon	2023-11-03–04, On-site (2 days)	Trade School, Incubator	Innovation, computer vision, startup
22	Academic Seminar	2023-11-30, Online (1 h)	Research Scientist	Computer vision, ecology, data monitoring
23	Academic Seminar	2024-01-18, On-site (2 h)	Professor and Director, Nordic AI Lab	Computer vision, image analysis

#	Event	Date & Format	Affiliation	Themes
24	AI Conference	2024-10-01–02, On-site (2 days)	Industry	Infrastructure, deep learning, large-scale computing
25	Academic Lab	2024-10-03–31 On-site (1 month)	Academic Lab	Computer vision, deep learning, three-dimensional reconstruction
26	Robotics Conference	2024-10-11, On-site (1 day)	Academic	Robotics, computer vision, three-dimensional reconstruction

Public Summary in Swedish

Artificiell intelligens beskrivs ofta som om den lär sig själv eller är autonom. Men det krävs ett omfattande och ofta osynligt arbete av människor. Under fyra år följde jag några fall av det här arbetet på nära håll. Målet var att se vad människor gör när de utvecklar, testar och inför system för datorseende.

Jag genomförde fältarbete genom att studera AI-forskare på plats i deras labb. Jag följde flera olika projekt: ett konsortium som byggde ett stort jämförelsedataset, och projekt som anpassade färdigtränade modeller för specifika ändamål som fågelövervakning, fordonsdetektion och stöd i strålbehandling. Totalt genomförde jag 36 intervjuer samt observerade arbetsmöten och digitala arbetsmiljöer.

Centralt i analysen är begreppet kunskapsproduktion. Med det avser jag de praktiska sätt genom vilka teknisk kunskap frambringas, stabiliseras och görs användbar. Avhandlingen fokuserar på kunskapsproduktionen inom tre huvudområden: när modeller byggs i labbet, när dataset skapas som utvärderingsramar, och när färdiga modeller görs användbara i nya lokala sammanhang.

I labbet handlar arbetet ofta om att sätta ihop kedjor av tekniska steg. Forskarna väljer modellarkitektur, förbereder data, ställer in algoritmer och testar parametrar. Jag kallar denna typ av praktiskt arbete för pipeline-svetsning: att sammanfoga komponenter så att de fungerar tillsammans. När modellen lär sig uppstår ”vikter” vilket kan förstås som interna inställningar som kodar vad modellen lärt sig. Dessa vikter är svåra att tolka direkt. För att förstå och bedöma modellen utvecklar forskarna gemensamma rutiner. De följer mätkurvor, visualiserar mellanresultat och diskuterar avvikelser. På så vis skapas kunskap om modellens inre genom praktiska interaktioner. Denna kollektiva uppmärksamhet skapar medvetenhetskontexter: delade sätt att tolka modellens beteende och avgöra när och varför den fungerar som förväntat.

När ett benchmark-dataset byggs upp blir kunskapsproduktionen synlig på ett annat sätt. Att skapa ett dataset kräver många val: vad som räknas som relevant data, hur situationer ska avgränsas, hur materialet ska spelas in och märkas upp. Detta dataset krävde att dess skapare behövde enas om vad som räknas som vardagliga aktiviteter. Dessa val förhandlas fram i detaljerade protokoll och instruktioner. Jag beskriver detta som samordnings- och anpassningsarbete: att forma gemensamma standarder och procedurer så att data kan användas för jämförande utvärdering. Valen som görs under insamling och annotering avgör vilka slags situationer och moment modeller senare kan känna igen. Resultatet blir sällan en neutral spegling av verkligheten, istället är datasetet en förenklad och selekterad bild.

I arbetet med att föra in färdiga modeller i nya sammanhang visar kunskapsproduktionen sig som en lokaliseringsprocess. Stora modeller tränas ofta centralt. När de ska användas lokalt krävs finjustering med lokala data, omfattande dataarbete och samarbete med domänexperter. Jag kallar detta för implementeringsbanor: de praktiska processer som tar en modell från ett generellt till ett kontextspecifikt bruk. Under dessa banor spelar nätverk och trovärdighet en viktig roll. Forskarna använder sina kontakter för att få tillgång till data, förklara tekniken för slutanvändare och tolka regler. Jag beskriver denna dimension som en social licens: förmågan att samla stöd och resurser som behövs för att arbetet ska kunna fortsätta. Kunskapen om hur en modell ska anpassas är alltså både teknisk och social.

Mitt bidrag är att ge namn åt och synliggöra dessa processer. Den empiriska analysen leder till flera begrepp som fångar kunskapsproduktionens processer: pipeline-svetsning, delegerad konstruktion eller att skapa förutsättningar för maskininlärning, medvetenhetskontexter, samordnings- och anpassningsarbete, implementeringsbanor och social licens. Dessa begrepp hjälper till att göra synligt vad som annars riskerar att framstå som teknisk magi eller isolerad intelligens.

Publications at the Department of Sociology
Lund University

Ordering and current prices at
bokshop.lu.se
Books are delivered against invoice or card payment.

Lund Dissertations in Sociology (ISSN 1102–4712)

1 Chylicki, Pawel *Att upphöra med brott: vägar ut ur den kriminella karriären* 275 pages ISBN 9179661823 (1992)
2 Eðvarðsson, Ingi Rúnar *Printing in Action: General Printing in Iceland and Sweden* 231 pages ISBN 9179661858 (1992)
3 Sverrisson, Árni *Evolutionary Technical Change and Flexible Mechanization: Entrepreneurship and Industrialization in Kenya and Zimbabwe* 274 pages ISBN 9179662404, 0862383404 (1993)
4 Sederblad, Per *Arbetsorganisation och grupper: studier av svenska industriföretag* 285 pages ISBN 9179662501 (1993)
5 Tjörvason, Sævar *Demokratiskt deltagande och kognitiv socialisation* 196 pages ISBN 9179662471 (1993)
6 Lema, Antoine *Africa Divided: The Creation of "Ethnic Groups"* 200 pages ISBN 9179662552, 086238348X (1993)
7 Ericson, Birgitta & Johansson, Britt-Marie *Bostadsbyggandet i idé och praktik: om kunskaper och föreställningar inom byggsektorn* 387 pages ISBN 917966282X (1994)
8 Rafnsdóttir, Guðbjörg Linda *Kvinnofack eller integrering som strategi mot underordning: diskussion kring kvinnliga fackföreningar på Island* 224 pages ISBN 9179663125 (1995)
9 Knoll, Thomas *Not a Piece of Cake: Ambivalence about Female-Headed Families* 352 pages ISBN 9179663133, 0862383889 (1995)
10 Pekkola, Sari *Magical Flutes: Music Culture and Music Groups in a Changing Bolivia* 254 pages ISBN 9179663583, 0862384273 (1996)
11 Cvetkova, Cveta *Codes of power* 309 pages ISBN 9179663745, 0862384435 (1996)

12 Cuesta, Marta [photography: Alejandro Parellada & Marta Cuesta] *Flowers to the ocean: a study of everyday practices, rituals and identity in Candomblé* 222 pages ISBN 9179664067, 0862384753 (1997)

13 Neergaard, Anders *Grasping the Peripheral State: A Historical Sociology of Nicaraguan State Formation* 401 pages ISBN 91-89078-00-4 (1997)

14 Jannisa, Gudmund *The Crocodile's Tears: East Timor in the Making* 328 pages ISBN 91-89078-02-0 (1997)

15 Naranjo, Eduardo *Den auktoritära staten och ekonomisk utveckling i Chile: Jordbruket under militärregimen 1973–1981* 429 pages ISBN 91-89078-03-9 (1997)

16 Wangel, Arne *Safety Politics and Risk Perceptions in Malaysian Industry* 404 pages ISBN 91-89078-06-3 (1997)

17 Jönhill, Jan Inge *Samhället som system och dess ekologiska omvärld: En studie i Niklas Luhmanns sociologiska systemteori* 521 pages ISBN 91-89078-09-8 (1997)

18 Lindquist, Per *Det klyvbara ämnet: Diskursiva ordningar i svensk kärnkraftspolitik 1972–1980* 445 pages ISBN 91-89078-11-X (1997)

19 Richard, Elvi *I första linjen: Arbetsledares mellanställning, kluvenhet och handlingsstrategier i tre organisationer* 346 pages ISBN 91-89078-17-9 (1997)

20 Einarsdotter-Wahlgren, Mia *Jag är konstnär! En studie av erkännandeprocessen kring konstnärskapet i ett mindre samhälle* 410 pages ISBN 91-89078-20-9 (1997)

21 Nilsson-Lindström, Margareta *Tradition och överskridande: En studie av flickors perspektiv på utbildning* 165 pages ISBN 98-89078-27-6 (1998)

22 Popoola, Margareta *Det sociala spelet om Romano Platso* 294 pages ISBN 91-89078-33-0 (1998)

23 Eriksson, Annika *En gangster kunde kanske älska sin mor… Produktionen av moraliska klichéer i amerikanska polis- och deckarserier* 194 pages ISBN 91-89078-36-5 (1998)

24 Abebe Kebede, Teketel *'Tenants of the State': The Limitations of Revolutionary Agrarian Transformation in Ethiopia, 1974-1991* 364 pages ISBN 91-89078-38-1 (1998)

25 Leppänen, Vesa *Structures of District Nurse – Patient Interaction* 256 pages ISBN 91-89078-44-6 (1998)

26 Idof Ståhl, Zeth *Den goda viljans paradoxer: Reformers teori och praktik speglade i lärares erfarenheter av möten i skolan* 259 pages ISBN 91-89078-45-4 (1998)

27 Gustafsson, Bengt-Åke *Symbolisk organisering: En studie av organisatorisk förändring och meningsproduktion i fyra industriföretag* 343 pages ISBN 91-89078-48-9 (1998)

28 Munk, Martin *Livsbaner gennem et felt: En analyse af eliteidrætsudøveres sociale mobilitet og rekonversioner of kapital i det sociale* rum 412 pages ISBN 91-89078-72-1 (1999)

29 Wahlin, Lottie *Den rationella inbrottstjuven? En studie om rationalitet och rationellt handlande i brott* 172 pages ISBN 91-89078-85-3 (1999)

30 Mathieu, Chris *The Moral Life of the Party: Moral Argumentation and the Creation of meaning in the Europe Policy Debates of the Christian and Left-Socialist Parties in Denmark and Sweden 1960–1996* 404 pages ISBN 91-89078-96-9 (1999)

31 Ahlstrand, Roland *Förändring av deltagandet i produktionen: Exempel från slutmonteringsfabriker i Volvo* 165 pages ISBN 91-7267-008-8 (2000)

32 Klintman, Mikael *Nature and the Social Sciences: Examples from the Electricity and Waste Sectors* 209 pages ISBN 91-7267-009-6 (2000)

33 Hultén, Kerstin *Datorn på köksbordet: En studie av kvinnor som distansarbetar i hemmet* 181 pages ISBN 91-89078-77-2 (2000)

34 Nilsén, Åke *"En empirisk vetenskap om duet": Om Alfred Schutz bidrag till sociologin* 164 pages ISBN 91-7267-020-7 (2000)

35 Karlsson, Magnus *Från Jernverk till Hjärnverk: Ungdomstidens omvandling i Ronneby under tre generationer* 233 pages ISBN 91-7267-022-3 (2000)

36 Stojanovic, Verica *Unga arbetslösas ansikten: Identitet och subjektivitet i det svenska och danska samhället* 237 pages ISBN 91-7267-042-8 (2001)

37 Knopff, Bradley D. *Reservation Preservation: Powwow Dance, Radio, and the Inherent Dilemma of the Preservation Process* 218 pages ISBN 91-7267-065-7 (2001)

38 Cuadra, Sergio *Mapuchefolket – i gränsernas land: En studie av autonomi, identitet, etniska gränser och social mobilisering* 247 pages ISBN 91-7267-096-7 (2001)

39 Ljungberg, Charlotta *Bra mat och dåliga vanor: Om förtroendefulla relationer och oroliga reaktioner på livsmedelsmarknaden* 177 pages ISBN 91-7267-097-5 (2001)

40 Spännar, Christina *Med främmande bagage: Tankar och erfarenheter hos unga människor med ursprung i annan kultur, eller Det postmoderna främlingskapet* 232 pages ISBN 91-7267-100-9 (2001)

41 Larsson, Rolf *Between Crisis and Opportunity: Livelihoods, diversification, and inequality among the Meru of Tanzania* 519 pages Ill. ISBN 91-7267-101-7 (2001)

42 Kamara, Fouday *Economic and Social Crises in Sierra Leone: The Role of Small-scale Entrepreneurs in Petty Trading as a Strategy for Survival 1960-1996* 239 pages ISBN 91-7267-102-5 (2001)

43 Höglund, Birgitta *Ute & Inne: Kritisk dialog mellan personalkollektiv inom psykiatrin* 206 pages ISBN 91-7267-103-3 (2001)

44 Kindblad, Christopher *Gift and Exchange in the Reciprocal Regime of the Miskito on the Atlantic Coast of Nicaragua, 20th Century* 279 pages ISBN 91-7267-113-0 (2001)

45 Wesser, Erik *"Har du varit ute och shoppat, Jacob?" En studie av Finansinspektionens utredning av insiderbrott under 1990-talet* 217 pages ISBN 91-7267-114-9 (2001)

46 Stenberg, Henrik *Att bli konstnär: Om identitet, subjektivitet och konstnärskap i det senmoderna samhället* 219 pages ISBN 91-7267-121-1 (2002)

47 Copes, Adriana *Entering Modernity: The Marginalisation of the Poor in the Developing Countries. An Account of Theoretical Perspectives from the 1940's to the 1980's* 184 pages ISBN 91-7267-124-6 (2002)

48 Cassegård, Carl *Shock and Naturalization: An inquiry into the perception of modernity* 249 pages ISBN 91-7267-126-2 (2002)

49 Waldo, Åsa *Staden och resandet: Mötet mellan planering och vardagsliv* 235 pages ISBN 91-7267-123-8 (2002)

50 Stierna, Johan *Lokal översättning av svenskhet och symboliskt kapital: Det svenska rummet i Madrid 1915–1998.* 300 pages ISBN 91-7267-136-X (2003)

51 Arvidson, Malin *Demanding Values: Participation, empowerment and NGOs in Bangladesh* 214 pages ISBN 91-7267-138-6 (2003)

52 Zetino Duartes, Mario *Vi kanske kommer igen, om det låser sig: Kvinnors och mäns möte med familjerådgivning* 246 pages ISBN 91-7267-141-6 (2003)

53 Lindell, Lisbeth *Mellan frisk och sjuk: En studie av psykiatrisk öppenvård* 310 pages ISBN 91-7267-143-2 (2003)

54 Gregersen, Peter *Making the Most of It? Understanding the social and productive dynamics of small farmers in semi-arid Iringa,* Tanzania 263 pages ISBN 91-7267-147-5 (2003)

55 Oddner, Frans *Kafékultur, kommunikation och gränser* 296 pages ISBN 91-7267-157-2 (2003)

56 Elsrud, Torun *Taking Time and Making Journeys: Narratives on Self and the Other among Backpackers* 225 pages ISBN 91-7267-164-5 (2004)

57 Jörgensen, Erika *Hållbar utveckling, samhällsstruktur och kommunal identitet: En jämförelse mellan Västervik och Varberg* 242 pages ISBN 91-7267-163-3 (2004)

58 Hedlund, Marianne *Shaping Justice: Defining the disability benefit category in Swedish social policy* 223 pages ISBN 91-7267-167-X (2004)

59 Hägerström, Jeanette *Vi och dom och alla dom andra andra på Komvux: Etnicitet, genus och klass i samspel* 234 pages ISBN 91-7267-169-6 (2004)

60 Säwe, Filippa A*tt tala med, mot och förbi varandra: Samtal mellan föräldrar och skolledning på en dövskola* 215 pages ISBN 91-7267-173-4 (2004)
61 Alkvist, Lars-Erik *Max Weber och kroppens sociologi* 271 pages ISBN 91-7267-178-5 (2004)
62 Winsvold, Aina *Når arbeidende barn mobiliserer seg: En studie av tre unioner i Karnataka, India* 300 pages ISBN 91-7267-183-1 (2004)
63 Thorsted, Stine *IT-retorik og hverdagsliv: Et studie af fødevarehandel over Internet* 219 pages ISBN 91-7267-186-6 (2005)
64 Svensson, Ove *Ungdomars spel om pengar: Spelmarknaden, situationen och karriären* 308 pages ISBN 91-7267-192-0 (2005)
65 Lundberg, Anders P. *Om Gemenskap: En sociologisk betraktelse* 248 pages ISBN 91-7267-193-9 (2005)
66 Mallén, Agneta *Trygghet i skärgårdsmiljö: En studie om rädsla för brott i Åland* 218 pages ISBN 91-7267-1955 (2005)
67 Ryding, Anna *Välviljans variationer: Moraliska gränsdragningar inom brottsofferjourer* 222 pages ISBN 91-7267-188-2 (2005)
68 Burcar, Veronika *Gestaltningar av offererfarenheter: Samtal med unga män som utsatts för brott* 206 pages ISBN 91-7267-207-2 (2005)
69 Ramsay, Anders *Upplysningens självreflexion: Aspekter av Theodor W. Adornos kritiska teori* 146 pages ISBN 91-7267-208-0 (2005)
70 Thelander, Joakim *Mutor i det godas tjänst: Biståndsarbetare i samtal om vardaglig korruption* 194 pages ISBN 91-7267-211-0 (2005)
71 Henecke, Birgitta *Plan & Protest: En sociologisk studie av kontroverser, demokrati och makt i den fysiska planeringen* 272 pages ISBN 91-7267-213-7 (2006)
72 Ingestad, Gunilla *Dokumenterat utanförskap: Om skolbarn som inte når målen* 180 pages ISBN 91-7267-219-6 (2006)
73 Andreasson, Jesper *Idrottens kön: Genus, kropp och sexualitet i lagidrottens vardag* 267 pages ISBN 91-628-7009-2 (2007)
74 Holmström, Ola *Skolpolitik, skolutvecklingsarena och sociala processer: Studie av en gymnasieskola i kris* 249 pages ISBN 91-7267-229-3 (2007)
75 Ring, Magnus *Social Rörelse: Begreppsbildning av ett mångtydigt fenomen* 200 pages ISBN 91-7267-231-5 (2007)
76 Persson, Marcus *Mellan människor och ting. En interaktionistisk analys av samlandet* 241 pages ISBN 91-7267-238-2 (2007)
77 Schmitz, Eva *Systerskap som politisk handling: Kvinnors organisering i Sverige 1968–1982* 362 pages ISBN 91-7267-244-7 (2007)
78 Lundberg, Henrik *Filosofisociologi: Ett sociologiskt perspektiv på filosofiskt tänkande* 225 pages ISBN 91-7267-245-5 (2007)

79 Melén, Daniel *Sjukskrivningssystemet: Sjuka som blir arbetslösa och arbetslösa som blir sjukskrivna* 276 pages ISBN 91-7267-254-4 (2007)

80 Kondrup Jakobsen, Klaus *The Logic of the Exception: A Sociological Investigation into Theological Foundation of Political with specific regard to Kierkegaardian on Carl Schmitt* 465 pages ISBN 91-7267-265-X (2008)

81 Berg, Martin *Självets garderobiär: Självreflexiva genuslekar och queer socialpsykologi* 230 pages ISBN 91-7267-257-9 (2008)

82 Fredholm, Axel *Beyond the Catchwords: Adjustment and Community Response in Participatory Development in Post-Suharto Indonesia* 180 pages ISBN 91-7267-269-2 (2008)

83 Linné, Tobias *Digitala pengar: Nya villkor i det sociala livet* 229 pages ISBN 91-7267-282-X (2008)

84 Nyberg, Maria *Mycket mat, lite måltider: En studie av arbetsplatsen som måltidsarena* 300 pages ISBN 91-7267-285-4 (2009)

85 Eldén, Sara *Konsten att lyckas som par: Populärterapeutiska berättelser, individualisering och kön* 245 pages ISBN 91-7267-286-2 (2009)

86 Bjerstedt, Daniel *Tryggheten inför rätta: Om rätten till förtidspension enligt förvaltningsdomstolarna under tre decennier* 240 pages ISBN 91-7267-287-0 (2009)

87 Kåhre, Peter *På AI-teknikens axlar: Om kunskapssociologin och stark artificiell intelligens* 200 pages ISBN 91-7267-289-7 (2009)

88 Loodin, Henrik *Biografier från gränslandet: En sociologisk studie om psykiatrins förändrade kontrollmekanismer* 118 pages ISBN 91-7267-303-6 (2009)

89 Eriksson, Helena *Befolkning, samhälle och förändring: Dynamik i Halmstad under fyra decennier* 212 pages ISBN 91-7267-313-3 (2010)

90 Espersson, Malin *Mer eller mindre byråkratisk: en studie av organisationsförändringar inom Kronofogdemyndigheten* 182 pages ISBN 91-7267-315-X (2010)

91 Yang, Chia-Ling *Othering Processes in Feminist Teaching: A case study of an adult educational institution* 184 pages ISBN 91-7267-318-4 (2010)

92 Isaksson, Anna *Att utmana förändringens gränser: En studie om förändringsarbete, partnerskap och kön med Equal-programmet som exempel* 206 pages ISBN 91-7267-321-4 (2010)

93 Hilding, Lars-Olof *"Är det så här vi är": Om utbildning som normalitet och om produktionen av studenter* 200 pages ISBN 91-7267-326-5 (2011)

94 Berg, Pernille *The Reluctant Change Agent: Change, Chance and Choice among Teachers Educational Change in The City* 196 pages ISBN 91-7267-331-1 (2011)

95 Rahman, Mashiur *Struggling Against Exclusion: Adibasi in Chittagong Hill Tracts, Bangladesh* 202 pages ISBN 91-7267-334-6 (2011)

96 Eklund, Lisa *Rethinking Son Preference: Gender, Population Dynamics and Social Change in the People's Republic of China* 218 pages ISBN 978-91-7473-108-8 (2011)

97 Gustavsson, Klas *Det vardagliga och det vetenskapliga: Om sociologins begrepp* 260 pages ISBN 91-7267-336-2 (2011)

98 Sjödin, Daniel *Tryggare kan ingen vara: Migration, religion och integration i en segregerad omgivning* 268 pages ISBN 91-7267-337-0 (2011)

99 Ringström, Jonas *Mellan sanning och konsekvens: En studie av den tredje generationens kognitiva beteendeterapier* 268 pages ISBN 921-7267-338-9 (2011)

100 Norstedt, Maria *Berättelser om stroke och arbetsliv: Att upptäcka styranderelationer* 204 pages ISBN 978-91-7473-182-8 (2011)

101 Anving, Terese *Måltidens paradoxer: Om klass och kön i vardagens familjepraktiker* 228 pages ISBN 91-7267-339-2 (2012)

102 Basic, Goran *Samverkan blir kamp: En sociologisk analys av ett projekt i ungdomsvården* 287 pages ISBN 91-7267-346-X (2012)

103 Zettervall, Charlotta *Reluctant Victims into Challengers: Narratives of a Kurdish Political Generation in Diaspora in Sweden* 315 pages ISBN 978-91-7473-412-6 (2013)

104 Apelmo, Elisabet *Som vem som helst: Kön, funktionalitet och idrottande kroppar* 272 pages ISBN 978-91-7473-408-9 (2012) Publisher: Bokförlaget Daidalos, Bergsjödalen 54b, 415 68, Göteborg, www.daidalos.se

105 Sandgren, Mikael *Europa som nation: En ny stil i nationalismens genre* 242 pages ISBN 91-7267-356-7 (2013)

106 Sandberg, Johan *Social Policy of Our Time? An Inquiry into Evidence, Assumptions, and Diffusion of Conditional Cash Transfers in Latin America* 182 pages ISBN 91-7267- 365-6 (2014) Cannot be ordered by Media-Tryck, only by the author.

107 Frees Esholdt, Henriette *Når humor, leg og lyst er på spil: Social interaktion på en multietnisk arbejdsplads* 260 pages ISBN 978-91-7623-215-6 (2015).

108 Vaide, Johan, *Contact Space: Shanghai. The Chinese Dream and the Production of a New Society* 215 pages ISBN 978-91-7623-231-6 (2015)

109 Kolankiewicz, Marta *Anti-Muslim Violence and the Possibility of Justice* 226 pages ISBN 978-91-7623-257-6 (2015)

110 Boethius, Susanne *Män, våld och moralarbete: Rapporter från män som sökt behandling för våld i nära relationer* 262 pages ISBN 978-91-7623-450-1 (2015)

111 Görtz, Daniel *Etnifierade polispraktiker: Hur etnicitet görs i polisens vardag* 337 pages ISBN 978-91-7267-380-9 (2015)

112 Stjärnhagen, Ola *Ekonomisk tillväxt i välfärdskapitalismen: En jämförande studie av BNP per capita-tillväxten i rika OECD-länder 1970–2000.* 168 pages ISBN 978-91-7623-490-7 (2015)

113 Hedman, Karl *Managing Medical Emergency Calls* 281 pages ISBN 978-91-7623- 690-1 (2016)

114 Solano, Priscilla *Assisting in the Shadows: Humanitarianism, Shelters and Transit Migration Politics* 250 pages ISBN 978-91-7753-102-9 (2017)

115 Ennerberg, Elin *Destination employment? Contradictions and ambiguities in Swedish labour market policy for newly arrived migrants* 232 pages ISBN 978-91-7753-204-0 (2017)

116 Sunnercrantz, Liv *Hegemony and the Intellectual Function Medialised Public Discourse on Privatisation in Sweden 1988–1993.* 338 pages ISBN 978-91-7753-471-6 (2017)

117 Kaya, Gökhan *Aspirations, Capital and Identity: Four studies on the determinants of life chances for young Swedes with an immigrant background* 163 pages ISBN 978-91-7753-771-7 (2018)

118 Hylmö, Anders *Disciplined reasoning: Styles of reasoning and the mainstream-heterodoxy divide in Swedish economics* 360 pages ISBN 978-91-7753-788-5 (2018)

119 Yakhlef, Sophia *United Agents: Community of Practice within Border Policing in the Baltic Sea Area* 220 pages ISBN: 978-91-7753-806-6 (2018)

120 Flower, Lisa *Loyalty Work: Emotional interactions of defence lawyers in Swedish courtrooms* 266 pages ISBN 978-91-7753-820-2 (2018)

121 Jack, Tullia *Negotiating Conventions: cleanliness, sustainability and everyday life* 166 pages ISBN 978-91-7267-400-4 (2018)

122 Hemmaty, Mona *On the Margins: Migrants, Status Mobility and Recent Turns in Swedish Migration Politics* 203 pages ISBN 978-91-7895-036-2 (2019)

123 Kloppenborg Møller, Henrik *Spectral Jade. Materiality, Conceptualisation, and Value in the Myanmar-China Borderland Jadeite Trade* 377 pages ISBN 978-91-7895-346-2 (2019)

124 Franzén, Alexandra *Brottslingar av en mycket speciell sort. Spionskandalen som en maktkamp mellan visselblåsare, grävande journalister och underrättelsetjänster* 237 pages ISBN 978-91-7895-943-3 (2021)

125 Sahlin Lilja, Hanna *The Emergence, Establishment and Expansion of Fear of Crime Research in Sweden* 300 pages ISBN 978-91-8039-025-5 (2021)

126 Klitgaard, Vibeke *Social (u)orden i distriktspsykiatrien - En systemteoretisk analyse af psykiatriske patienters kommunikation og adfærd samt stigmatisering fra omverdenen* 445 pages, ISBN 978-91-8039-063-7 (2021)

127 Ahlstrand, Rasmus *Structures of subcontracting: Work organisation, control, and labour in the Swedish building sector* 272 pages, ISBN 978-91-8039-108-5 (2022)

128 Flaherty, Colm *A politics of community: Identity, Stigma, and Meaning in the Extra-Parliamentary Left* 315 pages, ISBN 978-91-8039-246-4 (2022)

129 Quaglietta Bernal, Oriana *In Her Words: Women's Accounts of Managing Drug-related risk, Pleasure, and Stigma* 338 pages, ISBN 978-91-8039-391-1 (2022)

130 Rasan, Imad *Women in the public sphere in Egypt: 2011–2014* 331 pages, ISBN 978-91-8039-487-1 (2023)

131 Chan, Elton *The Last Urban Frontier: Commodification of Public Space and the Right to the City in Insurgent Hong Kong* 208 pages, ISBN 978-91-8039-592-2 (2023)

132 Crusefalk, Lars *Financial journeys: Reasoning about debt and money among young adults in Sweden* 314 pages, ISBN 978-91-8039-648-6 (2023)

133 Larsson, Marie *The work of contracepting: young people's experiences and practices with contraceptives in Sweden* 341 pages, ISBN 978-91-8039-854-1 (2023)

134 Alenius Wallin, Linn *Bonusbarnbarn och bonusförföräldrar: Om omsorgsrelationer mellan generationer i ombildade och valda familjer* 328 pages, ISBN 978-91-8104-129-3 (2024)

135 Karlsson, Daniel *At least I have this freedom: Subjectivity and self-precarization among digital freelancers in the Swedish cultural industries* 300 pages, ISBN 978-91-8104-173-6 (2024)

136 Kazi, Uzma *Swedish-(Par)Desi: Belonging Narratives of Second-Generation Pakistani Muslims in Sweden* 404 pages, ISBN 978-91-8104-231-3 (2024)

137 Edling, Staffan *Political Epistemology: Knowledge and Advocacy in Trade Union Research* 205 pages, ISBN 978-91-8104-249-8 (2024)

138 Morris, Freja *Tenacious Documents. Exploring the Allure of Documentation in Swedish Compulsory Schools* 232 pages, ISBN 978-91-8104-339-6 (2025)

139 Tollu, Aube *Queer intimacies: Rethinking Epistemology of Conducting Research with Jihadist Group Members* 210 pages, ISBN 978-91-8104-337-2 (2025)

140 Saaranen, Alexander *Trust in Time: How the Impressionable Years, Detrimental Events and Collective Experiences Shape Social Trust* 188 pages, ISBN 978-91-8104-461-4 (2025)

141 Kallos, Anna *Not a Real Job? Teenage Labor and Low-Wage Service Work in Sweden* 101 pages, ISBN 978-91-8104-598-7 (2025)

142 Bazan Royuela, David *Algorithmic interaction ritual chains on TikTok: scrolling through feedback loops* 302 pages, ISBN 978-91-8104-628-1 (2025)

143 Engdahl, Isak *Pixels and Weights: The Situated Work of Teaching Computers to See* 176 pages, ISBN 978-91-8104-691-5 (2025)

1996:1 Forsberg, Pia *Välfärd, arbetsmarknad och korporativa institutioner: En studie av Trygghetsrådet SAF/PTK* 147 pages ISBN 91-89078-07-1
1996:2 Klintman, Mikael *Från "trivialt" till globalt: Att härleda miljöpåverkan från motiv och handlingar i urbana sfärer* 171 pages ISBN 91-89078-46-2
1996:3 Höglund, Birgitta *Att vårda och vakta: Retorik och praktik i en rättspsykiatrisk vårdkontext* 215 pages ISBN 91-89078-68-3
1997:1 Jacobsson, Katarina *Social kontroll i dövvärlden* 148 pages ISBN 91-89078-18-7
1997:2 Arvidsson, Adam *Den sociala konstruktionen av "en vanlig Människa": Tre betraktelser kring reklam och offentlighet* 122 pages ISBN 91-89078-26-8
1998:1 Lundberg, Magnus *Kvinnomisshandel som polisärende: Att definiera och utdefiniera* 136 pages ISBN 91-89078-40-3
1998:2 Stojanovic, Verica *Att leva sitt liv som arbetslös... Svenska och danska ungdomars relationer, ekonomi, bostadssituation och värdesättning av arbete* 48 pages ISBN 91-89078-54-3
1998:3 Wesser, Erik *Arbetsmarknad och socialförsäkring i förändring: En studie av långtidssjukskrivning och förtidspensionering på 90-talet* 150 pages 91-89078-57-8
1999:1 Radmann, Aage *Fotbollslandskapet: Fotboll som socialt fenomen* 167 pages ISBN 91-89078-81-0
1999:2 Waldo, Åsa *Vardagslivets resor i den stora staden* 288 pages ISBN 91-89078-88-8
1999:3 Säwe, Filippa *Om samförstånd och konflikt: Samtal mellan föräldrar och skolledning på en specialskola* 159 pages ISBN 91-89078-93-4
1999:4 Schmitz, Eva *Arbetarkvinnors mobiliseringar i arbetarrörelsens barndom: En studie av arbetarkvinnors strejkaktiviteter och dess inflytande på den svenska arbetarrörelsen* 138 pages ISBN 91-89078-99-3
2000:1 Copes, Adriana *Time and Space: An Attempt to Transform Relegated Aspects in Central Issues of the Sociological Inquiry* 177 pages ISBN 91-7267-003-7
2000:2 Gottskalksdottir, Bergthora *Arbetet som en port till samhället: Invandrarakademikers integration och identitet* 89 pages ISBN 91-7267-012-6
2000:3 Alkvist, Lars-Erik *Max Weber och rationalitetsformerna* 176 pages ISBN 91-7267-019-3
2001:1 Bergholtz, Zinnia *Att arbeta förebyggande: Tankar kring ett hälsoprojekt* 50 pages ISBN 91-7267-043-6
2005:1 Bing Jackson, Hannah *Det fragmenterede fællesskab: Opfattelser af sociale fællesskabers funktion og deres udvikling i det senmoderne samfund* 162 pages ISBN 91-7267-190-4
2005:2 Lundberg, Henrik *Durkheim och Mannheim som filosofisociologer* 88 pages ISBN 91-7267-200-5

1 Ekströmer, Melcher *Buskar och fåglar: en socialantropologisk studie av górale, bergsfolket i södra Polen* 228 pages ISBN 91-79661-47-5 (acad. thesis 1991)
2 Bergendorff, Sten *Faingu City: a modern Mekeo clan in Papua New Guinea* 279 pages ISBN 91-79663-60-5 (acad. thesis 1996)
3 Pérez-Arias, Enrique *Mellan det förflutna och framtiden: Den sandinistiska revolutionen i Nicaragua* 322 pages ISBN 91-89078-01-2 (acad. thesis 1997)
4 Karlsson, B. G. *Contested Belonging: An Indigenous People's Struggle for Forest and Identity in Sub-Himalayan Bengal* 318 pages ISBN 91-89078-04-7 (acad. thesis 1997)
5 Lindberg, Christer (red) *Antropologiska porträtt 2.* 342 pages ISBN 91-89078-05-5 (1997)
6 Gooch, Pernille *At the Tail of the Buffalo: Van Gujjar pastoralists between the forest and the world arena* 391 pages ISBN 91-89078-53-5 (acad. thesis 1998)
7 Persson, Johnny *Sagali and the Kula: A regional systems analysis of the Massim* 245 pages ISBN 91-89078-87-X (acad. thesis 1999)
8 Malm, Thomas *Shell Age Economics: Marine Gathering in the Kingdom of Tonga, Polynesia* 430 pages ISBN 91-89078-97-7 (acad. thesis 1999)
9 Johansson Dahre, Ulf *Det förgångna är framtiden: Ursprungsfolk och politiskt självbestämmande i Hawai'i* 228 pages Ill. ISBN 91-7267-107-6 (acad. thesis 2001)
10 Johnsdotter, Sara *Created by God: How Somalis in Swedish Exile Reassess the Practice of Female Circumcision* 301 pages ISBN 91-7267-127-0 (acad. thesis 2002)
11 Andersson, Oscar *Chicagoskolan: Institutionaliseringen, idétraditionen & vetenskapen* 336 pages ISBN 91-7267-153-X (acad. thesis 2003)
12 Carlbom, Aje *The Imagined versus the Real Other: Multiculturalism and the Representation of Muslims in Sweden* 234 pages ISBN 91-7267-154-8 (acad. thesis 2003)
13 Antoniusson, Eva-Malin *Överdosens antropologi: En kontextuell studie* 232 pages ISBN 91-7267-161-0 (acad. thesis 2003)
14 Parker, Peter *How Personal Networks Shape Business: An Anthropological Study of Social Embeddedness, Knowledge Development and Growth of Firms* 156 pages ISBN 91-7267-182-3 (acad. thesis 2004)
15 Lindberg, Christer (red) *Nya antropologiska porträtt* 355 pages ISBN 91-7267-182-3 (2005)
16 Sliavaite, Kristina *From Pioneers to Target Group: Social change, ethnicity and memory in a Lithuanian power plant community* 206 pages ISBN 91-7267-202-1 (acad. thesis 2005)

17 Göransson, Kristina *Conflicts and Contracts – Chinese Intergenerational Relations in Modern Singapore* 187 pages ISBN 91-7167-202-1 (acad. thesis 2006)
18 Bourgouin, France *The Young, the Wealthy, and the Restless: Trans-national Capitalist Elite Formation in Post-Apartheid Johannesburg* 342 pages ISBN 91-7267-249-8 (acad. thesis 2007)
19 Matsson, Anna *The Power to do Good: Post-Revolution, NGO Society, and the Emergence of NGO-Elites in Contemporary Nicaragua* 208 pages ISBN 91-7267-251-X (acad. thesis 2007)
20 Holm, Hilma *Knowledge as Action – An Anthropological Study of Attac Sweden* 144 pages ISBN 91-7267-317-6 (acad. thesis 2010)
21 Wittrock, Hanna *Säg inte mötesplats! – Teater och integration i ord och handling* 268 pages ISBN 91-7267-332-X (acad. thesis 2011)
22 Hedlund, Anna *Exile Warriors: Violence and Community among Hutu Rebels in the Eastern Congo* 244 pages ISBN 978-91-7473-983-1 (acad. thesis 2014)
23 Capelán Köhler, Annika *Fibre Formations: Wool as an anthropological site* 260 pages ISBN 978-91-7753-202-6 (acad. thesis 2017)
24 Granbom, Lotta *The Second Wave: The Urak Lawoi After the Tsunami in Thailand.* 356 pages ISBN 978-91-7753-397-9 (acad. thesis 2017)
25 Krantz, Anna Sarri *Tredje generationens överlevande – En socialantropologisk studie om minne, antisemitism och identitet i spåret av Förintelsen* 194 pages ISBN 978-91-7753-852-3 (acad. thesis 2018)
26 Abelin, Matthias *A Swedish Dilemma – A Culture and Rule of Law in Swedish Sickness Insurance* 256 pages ISBN 978-91-7895-018-8 (acad. thesis 2019)
27 Berglund, Anna *Ambiguous hopes: an ethnographic study of agricultural modernisation in a Rwandan village* 223 pages ISBN 978-91-7895-269-4 (acad. thesis 2019)
28 Östergren, Petra *Sweden's Ban on Sex-Purchase: Morality politics and the governance of prostitution* 312 pages ISBN 978-91-8104-159-0 (acad. thesis 2024)

Licentiate's Dissertations in Social Anthropology (ISSN 1404-7683)

1999:1 Parker, Peter *Cognition and Social Organisation: A Framework* 125 pages ISBN 91-89078-76-4
1999:2 Johansson Dahre, Ulf *Politik med andra medel: En antropologisk betraktelse av rättens politiska och ideologiska förhållanden* 137 pages ISBN 91-7267-006-1