# Qualitative Research in an Era of AI: A Pragmatic Approach to Data Analysis, Workflow, and Computation[1] [2] [3]

This is a pre-print of an article scheduled for publication in the *Annual Review of Sociology*, Volume 52 (2026). This version has been submitted for review but has not yet undergone final assignment or copyediting. The final published version may differ from this manuscript. Please cite the final published version with DOI when it becomes available. Feedback on this draft, or related resources to include in associated materials are welcome and can be sent to corey.abramson@rice.edu

Corey M. Abramson, Ph.D., *Associate Professor of Sociology, Co-Director of the Center for Computational Insights into Inequality and Society (CIISR), Rice University*

Zhuofan Li, Ph.D., *Assistant Professor of Sociology, Virginia Tech*

Tara Prendergast, *Ph.D. Candidate in Sociology, University of Arizona*

Daniel Dohan, Ph.D., *Professor of Health Policy, Surgery, and Humanities and Social Sciences at the Philip R. Lee Institute for Health Policy Studies, UC San Francisco*

**Abstract**

Rapid computational developments—particularly the proliferation of artificial intelligence (AI)—increasingly shape social scientific research while raising new questions about in-depth qualitative methods such as ethnography and interviewing. Building on classic debates about using computers to analyze qualitative data, we revisit longstanding concerns and assess possibilities and dangers in an era of automation, AI chatbots, and "big data." We first historicize developments by revisiting classical and emergent concerns about qualitative analysis with computers. We then introduce a typology of contemporary modes of engagement—*streamlining workflows, scaling up projects, hybrid analytical approaches, and the sociology of computation*—alongside rejection of computational analyses. We illustrate these approaches with detailed workflow examples from a large-scale ethnographic study and guidance for solo researchers. We argue for a pragmatic sociological approach that moves beyond dualisms of technological optimism versus rejection to show how computational tools—simultaneously dangerous *and* generative—can be adapted to support longstanding qualitative aims when used carefully in ways aligned with core methodological commitments.

**Keywords:** *qualitative research, qualitative data analysis, computational social science (CSS), computational ethnography, mixed methods, artificial intelligence (AI), large language models (LLMs), open science, computational grounded theory, natural language processing (NLP), machine learning (ML)*

---

[1] We would like to thank Martín Sánchez-Jankowski and Dan Dohan, who offered feedback on the initial proposal, alongside support to 'update' their classic 1998 ARS article on computing and ethnography. We'd like to extend gratitude to Claude Fischer and Ronald Breiger as well for their support and discussions on this work. We would like to acknowledge the generous feedback from Kieran Turner, Andy McCumber, Alissa Sideman, Kelsey Gonzalez, Alma Hernandez, Ignacia Arteaga, Melissa Ma, Brandi Ginn, Stephen Zaparista, Yuhan (Victoria) Nian, and Zain Khemani. The authors are grateful to David Grusky and the full American Voices Project (AVP) team for their support in working with that large-scale qualitative data, the Qualitative Data Repository at Syracuse University for supporting work to verify associated papers, and the Russell Sage Foundation for facilitating opportunities to interact and share work. ASA events on "Trends in Mixed -Methods Research" and a panel on "Computational and Mathematical Approaches to Qualitative and Quantitative Data" organized by Laura Nelson provided additional inspiration.

[2] A version of this title first appeared in a blog post authored by one of the current authors: "A Silicon Cage? Qualitative Research in the Era of AI," Culture of Medicine, December 2023. The present article substantially expands and formalizes some ideas introduced there, as well as in prior methodological works.

[3] This research was supported in part by National Institute on Aging of the National Institutes of Health (NIA/NIH) award DP1AG069809 (Dohan PI). Content and views are those of the authors not of NIH.

# Introduction

Rapid computational developments, including the proliferation of 'artificial intelligence' (AI), increasingly shape contemporary life and sociological research(Bail 2024; Burrell and Fourcade 2021; Edelmann et al. 2020; Pugh 2024). The shifting technological landscape raises questions for traditionally in-depth qualitative methods like participant observation and in-depth interviewing. Building on debates about the role of computers in qualitative data analysis following the 25th anniversary of early debates about this issue, we revisit classic concerns (Dohan and Sánchez-Jankowski 1998; Franzosi 1998; Mohr and Duquenne 1997) and discuss the possibilities and dangers of combining qualitative research with computation in an era of automation, AI chatbots, and 'big data.'

We do not attempt to predict whether computational developments will yield more good than harm—a task involving prophecy as much as review. Instead, we describe pragmatic uses of current tools, outlining decision points and tradeoffs, and situating them within a broader sensibility that extends to computational technologies more broadly, and how they might be engaged sociologically.

First, we review and historicize methodological developments, noting that social scientists have used (and debated) computation in qualitative research for decades. New technologies rekindle longstanding concerns, such as the imposition of quantitative logics (Biernacki 2014; Burawoy 1998), distancing from data (Paula 2020), and the loss of contextual granularity (Grigoropoulou and Small 2022), alongside emergent issues like AI hallucinations and algorithmic bias (Grossmann et al. 2023; O'Neil 2016).

Second, we turn to emergent scholarship at the intersection of qualitative research and computational social science (CSS). We introduce a growing body of work that argues that contemporary computation can enhance, facilitate or expand qualitative research and analysis when used thoughtfully and with defined purpose. We examine four broad approaches to engaging with computational analysis and workflow in historically qualitative research (e.g. ethnography, in-depth interviewing, historical analyses)—*streamlining/augmenting workflow*, *scaling up* (e.g. Chandrasekar et al. 2024; Edin et al. 2024), *hybrid analytical approaches* involving mixed-methods (e.g. Abramson et al. 2018; Nelson 2020; Pardo-Guerra and Pahwa 2022), and the *sociology of computation* (e.g. Christin 2020). These approaches, alongside the rejection of technologies, differ in their aims and means of technology use, shaping strategies of engagement.

Third, we provide both general and specific examples of when and how various computational tools are being integrated into qualitative studies today. We describe how technologies can be used to process and analyze information like ethnographic fieldnotes, historical documents and interview transcripts. We also discuss how they shape project design even when not intended—e.g. how algorithms mediate our literature reviews—and identify further possibilities for analysis, publication, and data sharing. Detailing our work on a large-scale ethnographic study that combines human interpretation with computational procedures, we explain the workflow and how key aspects can be adapted for different aims, as well as single-researcher studies (Arteaga et al. forthcoming 2025; Li and Abramson 2025).

We conclude by noting that debates over the deployment and repurposing of analytical technologies are a perennial topic in the history of social science. Sociology has a deep tradition of pragmatic and sometimes subversive adaptations of available tools. This legacy runs from W.E.B. Du Bois's (1900) use of data visualization to challenge racial inequality to C. Wright Mills's (1959) transformation of the filing cabinet into a systematic engine for imaginative inquiry. Our contribution is to articulate a sociological approach that moves beyond dualism of technological optimism or rejection to show how computation can be adapted to support longstanding qualitative aims. In so doing, we model navigating computational tools that can be both dangerous *and* generative, constraining *and* useful in ways that remain aligned with core methodological commitments.

## Qualitative Analysis and Computing: Debates, Longstanding and New

Qualitative research encompasses methods that are not easily reducible to numerical data, such as ethnography and in-depth interviews (Small and Calarco 2022). These methods are essential for capturing meanings and occurrences that might otherwise be missed (Glaser and Strauss 1967; Lamont and Swidler 2014), illuminating inequalities (Burawoy 1998; Collins 2000), documenting sequences of behavior in situ (Jerolmack and Khan 2014; Sánchez-Jankowskii 2002), and preserving ecological validity (Cicourel 1982; Gans 1999). Qualitative research draws on diverse traditions – from phenomenology to realism, intersectional feminism to grounded theory – that have distinct philosophical priors (Gong and Abramson 2020; Steinmetz 2005)

This diversity creates confusion around terms. The idea of 'qualitative coding'—tagging text data with a theme, concept, category, or variable – provides a useful example. While some critique coding as reductive categorization (Biernacki 2014) and others see it as central to theorizing (Strauss and Corbin 1990),we frame it as indexing: a systematic way to tag and retrieve text without displacing or supplanting deeper interpretation (Li and Abramson 2025). This aligns with a pragmatic, iterative approach common in many strains of contemporary sociology that integrate prior theory with emergent findings rather than adhering to a strict inductive or deductive logic (Abramson and Sánchez-Jankowski 2020; Deterding and Waters 2021; Lichtenstein and Rucks-Ahidiana 2023). Our approach uses indexing to identify patterns and navigate complexity without reducing text to codes. But the term 'coding' itself can attract Rorschach-like critiques for varied positions, even though they vary in research paradigms.

Despite the variation in approach, some challenges and imperatives are shared. In addition to the challenge of navigating definitional issues, qualitative scholars concerned with systematic inquiry face the same practical task of qualitative data analysis: how to examine and share information that cannot be reduced to numbers alone without compromising the purpose of collecting in-depth data in the first place (Abramson and Dohan 2015).Computational tools can and have been used in a variety of ways to accomplish this.

*Qualitative Data Analysis Software*

The use of computational methods for managing and analyzing qualitative data has a history that predates generative AI (i.e. computer systems that generate new content by learning patterns from existing data). For decades, researchers have used commercial software, home-made retrieval systems in markdown, notebooks and word processors, spreadsheets, and occasionally (computational) code to analyze and write about people. Early forms of Computer-Assisted Qualitative Data Analysis Software (CAQDAS) were designed to streamline labor-intensive tasks involved in managing complex text data--like ethnographic field notes, interview transcripts, and historical documents (Silver 2018). These tools supported core processes such as aggregating, structuring, tagging/coding, memoing, and retrieving text from fieldnotes, transcripts, and documents (Dohan and Sánchez-Jankowski 1998)—functions previously carried out by scholars like C. Wright Mills and W.E.B. Du Bois with bureaucratic technologies such as filing cabinets, carbon-paper and folders (Abramson 2024). While sociological applications have sometimes integrated text with computational modeling (Mohr and Duquenne 1997; Roberts, et al. 2022), most qualitative uses of computers have emphasized reducing time burdens and reliance on memory, preserving materials for future analysis, and making established practices more efficient rather than fundamentally rethinking the possibilities of analysis.

**Definitions and Acronyms: Artificial Intelligence and Its Associates**

Artificial intelligence (AI*)* refers to technologies designed to perform tasks that historically relied on human abilities—this can include recognizing patterns, extracting text from archival images, summarizing interviews, or generating synthetic content and simulations. Some subfields commonly used in qualitative research workflows include machine learning (ML) for text classification, natural language processing (NLP**)** for parsing language data, and computer vision for analyzing visual materials. Large language models (LLMs) are a subset of AI: deep learning systems trained on mass-scale text data to predict and/or generate language (GPT is a commercial example). While sometimes integrated into qualitative analysis, LLMs pose various methodological and ethical challenges.

Over time tools expanded to include possibilities for quantification, visualization, and automation as well as emergent forms of new scholarship (Bjerre-Nielsen and Glavind 2022; Li and Abramson 2025; Peponakis et al. 2023). The current moment abounds with technologies that alter, extend or reconfigure capabilities for qualitative data analysis. Recent work demonstrates how purposeful machine learning can scale researcher-created codes (concepts, themes, categories) using inductive, deductive, and iterative approaches (Li, Dohan, and Abramson 2021; Nelson et al. 2021; Pardo-Guerra and Pahwa 2022). AI tools are being used to aid indexing of satellite imagery and historical texts (Law and Roberto 2025), transcribe and summarize audio, visualize and triangulate patterns in field-research (Abramson et al. 2018; Hanson and Theis 2024), identify cultural concepts embedded in large text datasets (Boutyline and Arseniev-

Koehler 2025; Kozlowski and Evans 2025), support counterfactual checks and retrieval (Davidson 2024; Li et al. 2021; Rodriguez and Spirling 2022; Stuhler, Ton, and Ollion 2025), and simulate humans beings (Kozlowski and Evans 2025).

Technological developments also enhance longstanding concerns and expose new dangers. In particular, the degree of automation and growth of complex systems under the umbrella term of AI, which includes applications from narrow machine learning to commercial chat bots and robotics systems, exacerbates some concerns about the use of computers in qualitative research.

### *Concerns and Critiques*

Even when CAQDAS supports conventional aims such as aggregation, organization, coding, and retrieval, it has faced persistent critiques, including:

- **The imposition of a singular logic.** Critics argue that computational tools impose a standardized logic on researchers that results in quantifying and structuring fluid, context-dependent data in ways that prioritize measurability rather than deep reading and interpretation (Biernacki 2014; Coffey, Beverley, and Paul 1996).This concern persists even as software now allows full reading and linking of complete documents—the core point being that tools (or their absence) shape our engagement with data even if they are flexible enough to allow both grounded theory and positivist inquiry. Biernacki (2014) argued that coding itself risks reducing text to variables, though others note contemporary ‘codes’ can function more like interactive hashtags that index text without replacing it (Li and Abramson 2025; Deterding and Waters 2021).
- **Distancing from data.** Qualitative work depends on close, immersive engagement with source materials, often drawn from in-depth direct human interactions (Lareau and Rao 2016; Fine and Abramson 2020). Interfaces that decontextualize quotes or excerpts have been criticized for creating analytic and emotional distance. Even when human review remains central, some worry that insights from slow, manual iteration may be lost leaning toward what they see as positivist error (Paulus and Marone 2024).
- **Decontextualization.** Computational methods risk stripping qualitative data of its meaning when context is reduced or ignored. An utterance or fieldnote gains its significance from a network of social, linguistic and historical context (Cicourel 1982; Geertz 2000). Context can vanish when data is quantified and separated from the textual representation, risking superficial or erroneous interpretations akin to mining for “statistical rather than sociological significance” (Bourdieu 1984; Lichtenstein and Rucks-Ahidiana 2023).

Many new concerns with the use of computing in qualitative analysis focus on the rise of large language models (LLMS), which have been used alongside or as a replacement for traditional techniques:

- **Technical unreliability and bias.** LLMs often produce inconsistent results, shifting with small prompt changes or model updates, typically without transparency (Bisbee et al. 2024; Chen, Zaharia, and Zou 2024) They are prone to "hallucinations"—plausible but incorrect outputs—which may be underestimated given intuitive interfaces (Farquhar et al. 2024). Training corpora (text data used to construct the models) introduce algorithmic bias (Ashwin, Chhabra, and Rao 2025), while failures of algorithmic fidelity obscure data nuance(Amirova et al. 2024). Results are difficult to replicate absent validation standards (Abdurahman et al. 2025). LLMs struggle with irony, sarcasm, and cultural nuance without specialized tuning (Weber and Prietl 2021).

- **Social harms and flawed interpretation.** Without curation and alignment models can reproduce stereotypes and biases (Omiye et al. 2023; Tan and Celis 2019), misrepresent groups (Santurkar et al. 2023), and amplify particular viewpoints as 'common sense' (Bender et al. 2021). In health care and organizational psychology, algorithmic sorting by race worsens disparities (Hurd, Payton, and Hood 2024; O'Neil 2016). LLMs rely on exploitative labor for data training and curation, reinforce racial hierarchies (Bender and Hanna 2025; Benjamin 2019), and erode human roles (Pugh 2024; Zuboff 2019).When used as proxies for human interpretation, they risk reproducing normative narratives, fabricating data, and extending classification systems rooted in historical, racialized disparities (Crawford 2021; Farber, Abramson, and Reich 2025; Hammer and Park 2021).
- **Security risks and corporate dependency.** Using third-party commercial LLMs to analyze sensitive data raises security and confidentiality risks when cloud chatbots handle qualitative data (Bail 2024; Chopra and Haaland 2023; Farber et al. 2025). The concentration of AI development in a few corporations fosters dependency and global inequality, while opaque decisions in training and precarious labor arrangements sustain the systems ( Burrell and Fourcade 2021; Spirling 2023). Deploying LLMs as "synthetic participants" raises dilemmas around informed consent about data uses from human subjects and risks substituting simulated for human and environmental costs of large-scale computation remain substantial e(Bender et al. 2021; Kozlowski and Evans 2025; Strubell, Ganesh, and McCallum 2020).

*Moral Panic or Legitimate Concerns?*

Some concerns with traditional CAQDAS never manifested. Scholars used CAQDAS for varied paradigms (ranging from conventional science to post-modern critique), and while early programs often decontextualized text, most contemporary software allows reading full excerpts in context. Automation has been used to assist with initial indexing, but in sociological analysis this does not replace deeper interpretive readings or the application of more complex concepts, making reduction a choice rather than a built-in feature (Deterding and Waters 2021; Li et al. 2021). Yet in the current moment, both project size and use of AI technologies are expanding.

Advocates for "scaling up" qualitative data to allow larger projects argue this can facilitate comparative or representative analyses alongside traditional projects and emphasize the importance of maintaining connections to the underlying text in context—a concern echoed in "big data" arenas, where scholars argue for scaling down to situate meanings alongside numbers (Abramson et al. 2024; Breiger 2015; Edin et al. 2024; Nelson 2020). New technologies may sharpen this tension with smaller scale work rather than remove it. Even in a project involving a single researcher, machine learning can scale human categories en masse or expand qualitative iteration (Li et al. 2021). Like other technologies (word dictionaries, statistics), tools can be applied to varied ends.

Some describe responses to AI as a general moral panic, arguing overstated fears spill into legitimate inquiries in ways that can thwart progress by association (Simon, Altay, and Mercier 2023; Spirling 2023; Yadlin and Marciano 2024). Yet many scholars who use advanced computing technologies underscore substantive risks, particularly with large-scale commercial LLMs. A point of disagreement is whether the risks are perceived to be manageable, with proponents pointing to precedent in scientific machine learning, inductive statistics, and other technologies adaptable to varied uses (Bail 2024; Davidson 2024).

Emerging uses illustrate pragmatic possibilities for tools that are simultaneously problematic and useful. These include secure local deployments of fine-tuned, context-dependent systems with human review (Abramson 2024; Li et al. 2021) and subversive applications that critically engage technology (Benjamin 2019). The expansion of AI into varied arenas of life—essays generated for college classes, blurbs on websites, emails and text messages—adds weight and emotional valence.

The technologies may be modern, but the debate about technological engagement is longstanding. Proponents and critics alike have long noted that science can serve harmful ends—not only physics and biology but also ethnography, statistics, and modern computing (Foucault 1977; Jasanoff 2004; Noble 2018). Opting out may be impossible or itself impose costs (Weber 1978). Computers now permeate the research process—literature reviews, writing, dissemination, data management—making avoidance nearly impossible. Many concerns about technology apply to standard practices, and more broadly to life under global capitalism. As Weber observed in his account of the "iron cage," once institutionalized, technological systems become difficult, even impossible, to avoid without incurring new constraints, and are deeply intertwined with social life.

Our position is that—like qualitative research itself—specific uses and issues vary in meaningful ways. Employing local language models (e.g., run on a local computer, without a commercial service, possibly fine-tuned to be receptive to data) to check patterns in human-coded data differs markedly from pasting text wholesale into ChatGPT. Yet both could (generously) be described as qualitative data analysis, and both (correctly) as using AI. Wholesale opting out may be mythologizing our ability to conduct scholarship outside real world contexts, while equating all uses risks erasing human agency. Ethnography has been used for both colonial expansion and documenting inequality; physics to build weapons and to generate renewable energy. Computing may extend surveillance and inequality (likely), but it might also support evidence to mitigate those very processes (Bail 2024).

To ground this discussion, we now turn to the range of ways scholars are engaging with emerging technologies to support, extend, or transform qualitative research.

## A Typology of Qualitative–Computational Engagement

Combinations of qualitative methods and computation are tremendously varied—reflecting the diversity of qualitative traditions and changing computational tools themselves. Following Merton’s (1938) discussion of social structure and anomie—which theorizes how people understand goals (valued ends) and means (institutionalized possibilities) and adjust through strategies of conformity, innovation, ritualism, etc.—we organize qualitative-computational approaches into genres of engagement. These ideal-typical styles that vary in whether they extend widely valued qualitative goals (understanding and interpreting human action) or expand toward new ones (predicting outcomes). They also vary in whether they rely on conventional means (software for extending codes) or require expanded computational techniques (using AI to create interactive simulations). This typology helps situate contemporary practices and clarify how scholars engage computation.

## Genres of Engagement

**Streamlining (extending goals, extending means).** Many approaches use computational tools to extend existing qualitative practices without radically altering them. The goal is continuity with valued practices, with technology serving to improve efficiency or consistency in tasks such as coding, retrieval, and comparison. Sociological examples include flexible and context coding that involve basic automations in QDA software for medium to large data sets (Deterding and Waters 2021; Lichtenstein and Rucks-Ahidiana 2023), and human-hybrid machine-learning approaches that improve label consistency while maintaining human judgment in text classification (Li et al. 2021). These methods extend established qualitative workflows to facilitate or enhance interpretive analysis rather than replace it with numerical information. Recent analyses suggest that many works use new LLM technologies with conventional approaches—such as inductive coding and thematic analysis—though sometimes automating in ways that extend beyond streamlining (Chandrasekar et al. 2024)

**Scaling Up (extending goals, expanding means).** Other projects extend the goals of qualitative inquiry while expanding how they are achieved, by focusing on how computational infrastructure can scale up the size, reach, and accessibility of qualitative data. These approaches align with open science initiatives that encourage transparency, sharing, and replication through expanded tools for de-identification, sharing, validation and reuse (Abramson et al. 2018; Freese, Rauf, and Voelkel 2022; Kirilova and Karcher 2017).

Examples include the American Voices Project (2021), which collected over 1500 interviews across the United States in its initial wave using advanced survey sampling techniques for representativeness (Edin et al. 2024), comparative ethnographic projects that use shared protocols alongside new technologies to produce policy data (Bernstein and Dohan 2020) ,and approaches that rethink aggregation and sampling for both heterogeneity and generalizability (DeLuca, Clampet-lundquist, and Edin 2016; Sykes, Verma, and Hancock 2018). Scaling does not a priori require altering research goals—such as the interpretive analysis of lived experience or charting behaviors in situ—but it does rely on computational expansion and conventions to manage, share, and analyze the resulting volume of "big" qualitative data in order to move beyond prior pre-computational attempts at scaling like Mass observation and the Human Relation Area Files (Abramson and Dohan 2015; Karcher et al. 2021; Murphy, Jerolmack, and Smith 2021).

**Hybrid approaches (expanded goals, expanded means).** A growing body of work combines qualitative interpretation with computational tools to expand qualitative research in more fundamental ways—aiming to reconfigure what is possible, or the types of questions that can be asked. This includes computational grounded theory for identifying themes (Nelson 2020), computational ethnography for charting patterns of speech and behavior within and across cases (Abramson et al. 2018), and the extended computational case method which examines data construction itself (Pardo-Guerra and Pahwa 2022). Each uses multiple methods to connect macro patterns and micro data. Such approaches, developed by scholars studying culture and inequality (Abramson et al. 2018; Nelson 2020), knowledge construction (Pardo-Guerra and Pahwa 2022), and health (Bernstein and Dohan 2020), allow researchers to examine larger data sets and connect multilevel patterns. They aim to bridge dualities between history and biography, objectivity and subjectivity, individual and group by showing both patterns and specific manifestations in micro-data (Breiger 1974; Du Bois 1996; Mills 1959; Mohr and Duquenne 1997).

Hybrid approaches may be combined with attempts at scaling using extant or prospective data (Abramson et al. 2024) or used on data of traditional scope to analyze data from multiple vantages in exploratory, confirmatory or explanatory ways (Garrett et al. 2019; Mische and Mart 2025). Hybrid approaches—rooted in sociology's pragmatic tradition and history of multi-method triangulation (Du Bois 1996; Small 2011)—reflect an expansion of goals to align with multi-method inquiry.

**Studying computation itself.** Qualitative methods' uneasy relationship with computing is compounded by work that makes the sociology of computing and digital life objects of inquiry in their own right (Boellstorff 2008; Burrell and Fourcade 2021; Christin 2020). While social science has long studied sociotechnical systems, large-scale computing introduces new methodological challenges. Algorithms are often opaque—hidden by organizational secrecy, user illiteracy, and, in AI, a lack of interpretability even for experts (Burrell 2016)

Studying them forces ethnographers to shift across levels of analysis, from individuals engaging code to platform economies, and to reassemble diffuse networks of human and non-human actors (Rosa 2022; Seaver 2017). Even when not the focus, algorithms' ubiquity demands engagement: researchers may need to "enroll" them in producing qualitative data (Christin 2020). For example, recommendation systems so shape social media that understanding how they translate data into user experience is now prerequisite for studying creators and influencers (Mears 2023). Researchers must likewise account for the extent and automation of content moderation (Roberts 2019). In the platform economy, algorithmic control makes architecture itself central to many service jobs, online and offline (Griesbach et al. 2019).

Algorithms also mediate the visibility and circulation of scholarship—through platforms like Google Scholar and citation metrics—shaping what becomes legible in the academic field and how scholars respond beyond methodological choices (Fourcade and Healy 2024; Pardo-Guerra 2022). These dynamics have led digital sociologists to integrate computational and mixed methods into domains traditionally reliant on qualitative data, in order to theorize human experience in an algorithmic society and to study phenomena like digitally enabled social change (Earl and Kimport 2013) .The approach may expand tools or scaling, but its typological position is defined by making tools an empirical focus.

**Rejection (rejected goals, rejected means).** Finally, not all qualitative scholars embrace computational methods, even at the level of software for organization. Some reject computation outright, either on epistemological grounds or in defense of a humanist orientation. These perspectives argue that attempts to formalize, automate, or quantify qualitative work distort its interpretive aims and risk reinforcing technocratic or ‘positivist’ logics (Biernacki 2014; Burawoy 1998). This stance often represents a rejection of the "social science turn" in favor of traditions that prioritize meaning, resistance, and humanism (Clifford and Marcus 1986).Rejection thus resists aligning qualitative research with computational or data-intensive science, which is perceived as the encroachment of a singular, technocratic logic or even the dangers of empiricism.

The typology, excluding rejection, is represented in table 1 below.

**Table 1: Genres of Qualitative – Computational Engagement in Research***

| | **Research Goals** | |
|---|---|---|
| **Technical Means** | *Extension* | *Expansion* |
| *Extension* | **Streamlining**<br>Using software more efficiently, without radically altering qualitative practices | **Sociology of Computation**<br>Defined by topical area<br><br>May involve new tools |
| *Expansion* | **Scaling Up**<br><br>Large-scale fieldwork<br>Team based interviews<br>Shared repositories | **Hybrid Approaches**<br><br>Mixed methods<br>In-depth text analysis and computational analysis together |

**Not shown: Rejection, of social scientific goals and computational means.*

Taken together, this typology highlights both continuity and divergence in how sociologists and other researchers approach computation in qualitative research and analysis. Some seek efficiency, others seek scale, others pursue hybrid innovation, and still others refuse or critique the turn to computation. Each response reflects a position within broader debates over the possibilities and limits of qualitative research in a computational era. In the next section, we turn to practical workflows and case studies that illustrate applications of technology to qualitative research and data analysis.

## Workflow and Uses of Computation in Qualitative Research (2025 edition, with general principles)

In the current moment of proliferating AI, the use of computation in qualitative research spans the entire project lifecycle, from study design to dissemination, even as skepticism and uncertainty about possible uses abound. Much like the earlier adoption of CAQDAS (Coffey et al. 1996; Dohan and Sánchez-Jankowski 1998), current tools—from shared data sets like the American Voices project, open-source visualization resources, and commercial LLMs—can reshape practices, addressing or introducing challenges of rigor, interpretation, and transparency over the course of a project.

We first provide an overview of workflow possibilities, illustrated with examples of current uses. Table 2 presents these applications in roughly chronological order (from project start to finish), distinguishing whether computational tools are broadly assistive (basic technologies predating contemporary applications, but involving a computer), automated (narrowly defined tasks, often integrating aspects of machine learning), or agentic (broader execution by computing systems that heavily involve LLMs). Rows can be thought of as phased with decision points in research, and columns how computers may be integrated.

**Table 2: Workflow– Stages and Computational Tools**

| | Assistive | Automated | Agentic |
|---|---|---|---|
| **Research Design** | Citation management<br>Project records<br>Project management<br>Version control | Readability checks<br>Data-assisted sampling<br>Simulating sample-size | Literature review |
| **Data Collection** | Participant/site tracking<br>Hyperlink field artifacts<br>E-consent capture<br>Digital diary or EMA<br>Survey apps<br>Cloud back-up/sync | Multi-media aggregation<br>Sensor or geospatial logging<br>Timestamping/geolocation<br>Live transcription | Adaptive or event based SMS prompts |
| **Data Processing** | Interview transcription<br>Transcript editing<br>File-format normalization<br>Data versioning<br>Data tracking | Scanned docs/images to text<br>A/V speech-to-text pipelines<br>Context/metadata auto-capture<br>Entity/event annotation layers<br>De-identification workflows<br>Batch uploads via APIs/CLI<br>Text into data frames/databases<br>Metadata tagging<br>Quality checks<br>Validation rules | Adaptive or event based reminders |
| **Data Analysis** | Human coding<br>Quote retrieval<br>Memo writing | List/regex scripts coding<br>Inter-coder reliability tests<br>Pattern examination/analysis<br>Visualizing patterns<br>Counterfactual checks<br>Data linking<br>Network overlays<br>Network simulations | LLM assisted coding<br>LLM assisted memos<br>ML classifiers<br>ML embeddings<br>Augmented Retrieval<br>Semantic Q&A<br>Typology validation |

| | | | |
|---|---|---|---|
| **Writing & Presentation** | Triangulation<br>Consistency checks<br>Real-time writing collab | Retrieval of analytic products<br>Generating visuals<br>Citation formatting/insertion<br>Plain-language summaries<br>Accessibility audits<br>Embedded runnable notebooks | Assisted writing |
| **Sharing & Preservation** | Replication code<br>Notebooks<br>Codebooks<br>DOI archiving<br>Long-term preservation | Containerized analytic spaces<br>Interactive data portals/APIs<br>Tiered access controls<br>Encryption for sensitive data | Simulated participants |

Consider, some present examples:

**Research Design.** At the outset of projects, computational tools are increasingly used in designing studies, reviewing literature, editing grants and IRB applications, and testing instruments like interview guides (Farber et al. 2025; Parker, Richard, and Becker 2023).

**Data Collection.** New applications extend computation into the field itself, with automated collection of interviews by chatbots, platforms for synchronizing fieldnotes, and reminders for structured data like time diaries (Astrupgaard et al. 2024; Chopra and Haaland 2023).

**Data Processing.** Once collected, qualitative data increasingly undergo systematic preprocessing with computational assistance, to deidentify and index information, particularly for projects conducted at scale (Li and Abramson 2025).

**Data Analysis.** The greatest proliferation of uses lies in analysis, which includes both traditional approaches that are streamlined and multi-method approaches combining computational modeling alongside in-depth interpretation of large or small scale qualitative data (Abramson et al. 2018; Nelson et al. 2020).

**Writing and Presentation.** AI and automation support editing and even text writing or generating text as 'synthetic' participants (Kozlowski and Evans 2025) as well as allowing pattern visualization (Abramson et al. 2024; Hanson and Theis 2024).

**Sharing and Preservation.** The expansion of infrastructure enables the technical possibility for making qualitative findings more shareable (Abramson and Dohan 2015; Edin et al. 2025; Murphy et al. 2021; Freese et al. 2022).

The online appendix contains a more detailed discussion with added references and debates. Our aim here is not to judge which applications are useful or correct, but to note how such tools have crept into our repertoire with purpose, as well as drift. We now turn to an example of our own purposive application to workflow.

## Workflow Example: Team Based Ethnography, Supported by Computational Social Science (CSS)

This section describes our use of technical tools in the DISCERN project—a five-year NIH-funded study (PI: Dohan) of how older adults living with dementia and family care partners navigate experiential, cultural, institutional, and structural circumstances across three U.S. states. Data include over 300 day-long site visits, 119 in-depth interviews, and longitudinal immersion by multiple fieldworkers in communities and institutions, building on team-based protocols described in prior works (Abramson and Dohan 2015).

Our approach integrates computational tools with traditional qualitative methods in ways that are iterative, grounded in human insights, and transparent while meeting human subjects and institutional requirements—using computers to facilitate data aggregation, indexing, analysis and verification—i.e., computational ethnography (Abramson et al. 2018). Though DISCERN is the focal case, aspects of this workflow are adoptable for individual and team-based studies (Li and Abramson 2025). We present work chronologically, noting where software enters, whether AI is used, how human review is maintained, and why specific steps matter for validity and interpretive depth.

**Research Design:** The study was designed to answer substantive questions about how people living with dementia (PLWD) and care partners make sense of and navigate cognitive decline, with attention to similarities and differences by background and context among the >20 million U.S. adults and care partners directly affected by Alzheimer's disease and related dementias (ADRD).

Site selection drew on pilot observations, aggregate demographic data and existing findings to maximize expected heterogeneity across regions and institutions, including disconfirming as well as confirmatory cases (Small and Calarco 2022; Bernstein and Dohan 2020). Project management used Slack (team chat, coordination), REDCap (secure regulatory database, linking names and pseudonyms), Git (code version control), and Box (papers, logistics, instruments), as well as weekly meetings and quarterly day-long retreats. A team wiki in Box provided core documentation summarizing protocols and linked to files like interview guides.

LLMs were not used here. For literature work we relied on Google Scholar, libraries, and bibliographic software rather than AI summarization, though API-based tools are more common today. All interviewers were fluent for languages in which they conducted interviews, and had training on field observation as well as qualitative interviewing. Interviewer training simulations used standardized patients (medical actors), not LLMs, to practice complex interactions with PLWD and care partners while respecting participants' time.

**Data Collection:** Interviews were conducted by trained fieldworkers familiar with sites and with the challenges of interviewing older adults, including those with mild cognitive impairment (with added consent and medical ethicist oversight). Transcription was performed by cleared professionals rather than automated services—more expensive, but precise and secure (and harder to find at present). Fieldnotes were typed by fieldworkers and uploaded to a Box system for archiving.

To enable cross-system integration, a shared file-naming convention linked site, date, data type, participant ID, and researcher initials. Data type markers in files (e.g., fn for fieldnote; iv1/iv2/iv3 for interview waves) carried into QDA software and Python/R data frames. Unique IDs linked data to REDCap. Substantive identifiers (e.g., OA017 for older adult #17, MP002 for medical provider 2) enabled linking across interviews, notes, and survey metadata, with pseudonyms substituted before publication.

Dates took the form YYYYMMDD so when sorted they could be presented from newest to oldest. For example, CAsite1_20220908_iv1_OA013_XX indicates an interview with a subject from CAsite1 (a senior living facility in California), the date, that this is a baseline interview (iv1), with subject OA013 (older adult), conducted by interviewer XX. This means alphabetical sorting, even without software, groups first by site, then date, etc. (the order of variable content sorted by "_" can change depending on team preference). This allows auditing data, and easy initial groupings by factors like year or demographics in QDA software or extraction of values in Python, as well as pairing with metadata and subsequent sorting and subsetting (Li and Abramson 2025).

Adjacent digital instruments (e.g., diaries, geolocation) would be possible but were not deployed in this study. Consent records, site logs, and tracking sheets were stored in secure systems separate from qualitative data, but downloadable as metadata (REDCap). No data were uploaded to commercial AI systems.

**Data Processing:** Formatting and cleaning preserved text integrity and interoperability in ATLAS.ti, Python, and archival storage. Transcripts and notes were inspected in Sublime Text to remove stray characters, enforce UTF-8 encoding, and standardize breaks and whitespace. ATLAS.ti projects and mirrored folders supported movement between interactive QDA and machine-readable exports without altering raw text (Li and Abramson 2025).

A QDA dictionary indexed topics relevant to dementia (terminology, symptoms, care contexts, social ties) by listing keywords for automated coding using grep/regex (always marked as auto_ to reflect the potential for false positives). Dictionaries are transparent, versioned, and served as scaffolds for coding and modeling—not replacements for reading or in-depth indexing. Python scripts validated file encoding, subject IDs for tracking across data types, and de-identification, with human verification always central.

Standardization (lowercasing, punctuation normalization, controlled stop-wording) supported computational procedures while preserving verbatim text for human reading. Embeddings (BERT, RoBERTa; more recently Gemma2-2B) identified synonyms for dictionary expansion. Fieldworkers contributed to keyword lists, and iterative dictionary growth preserved links back to original paragraphs for later reconstruction in QDA. This is only one strategy, but important for studies that produce more data than can be reasonably hand coded (Abramson and Dohan 2015; Li et al. 2021).

**Data Analysis:** Our analysis proceeded within a broadly realist paradigm, using computational tools to extend an iterative qualitative logic informed by insights from time with people, listening to them, talking with them, and observing them (Abramson and Sánchez-Jankowskii 2020). We treat our coding protocol as a multi-layered approach to indexing—a process for systematically tagging data for retrieval and analysis without replacing interpretive reading. To address classic concerns about context-stripping and to capture overlapping concepts, all indexing occurs at the paragraph level. This preserves the immediate context often lost in sentence-level analysis, a crucial feature for interview and ethnographic data. This helps provide a useful balance between precision and recall, false positives and false negatives, given the scale and content of our data (Li et al. 2021). All indexed data remain linked to the full source document via positional information in both our QDA software and our Python data schema (see Appendix).

Our workflow combines three genres of indexing—all of which involve writing analytical memos on patterns and specifics—often used within the same project depending on papers and team goals:

1. **List/Dictionary Indexing:** For straightforward concepts, we use dictionary-based tools like regex searchers for high-accuracy automation (e.g., tagging mentions of "clinical trials").
2. **Full Human Coding:** For complex, emergent, or highly interpretive concepts (e.g., identity), we rely on traditional human coding.
3. **Hybrid ML Classification:** To scale researcher-defined concepts (e.g., discussions of social networks), we train a local language model on human-coded examples, followed by rigorous human review (Li et al. 2021).

This layered approach supported several analytical moves. In ATLAS.ti, we used flexible coding for interpretive retrieval, memoing on topics like migration or healthcare challenges to support close reading and comparative case construction (Deterding and Waters 2021; Dohan and Sánchez-Jankowski 1998), often visualized with heatmaps, networks, and other visuals of cases or language (Abramson and Dohan 2015).

For some projects, we then used supervised scaling of human codes, where human-coded examples trained locally fine-tuned classifiers (e.g., RoBERTa; Gemma2-2B) to extend labels across the corpus, with all machine classifications undergoing human review before use and *always* paired with in depth reading (Li et al. 2021). Finally, we used supervised and unsupervised tools for pattern discovery and representation, building semantic networks to visualize concept co-occurrence (Abramson et al. 2024; Boutyline and Soter 2021) and using embeddings for similarity searches and charting meaning (Evans and Kozlowski 2025), with iterative human review to correct automation errors (now sometimes called "human in the loop") and ensure findings were grounded in the source text (Li and Abramson 2025).

Weekly meetings connected analytical work with fieldworker findings. Field teams shared observations and reports; analysts presented outputs (heatmaps, networks, site profiles); memos recorded provisional explanations and disconfirming evidence. We discussed reflexivity and positionality, discussed the comments of communities after sharing findings, unpacking challenges and insights, alongside similarities and differences in what we found. Identifiable outputs were restricted internally, and public presentations were de-identified. All computational analyses ran on a secure local Linux server; no commercial LLM APIs were used even though visualization, machine learning and classification could all reasonably be referred to as AI (Abramson 2024).

**Writing and Dissemination:** Text analysis and ethnographic interpretation were mutually informing, a key component of computational ethnography. Heatmaps summarized case-by-code patterns in analyses of words and cases (Abramson and Dohan 2015; Arteaga et al.); semantic networks provided meso-level maps linked to in-depth narratives; targeted retrieval grounded claims in verbatim passages and linked back to the original. Findings were reported as patterns across interviews and observations but we also focused on exemplary and disconfirming cases when relevant to research questions. Pseudonyms and de-identification were used to protect participants and meet IRB requirements. AI was not used in writing, though local models checked for typos on deidentified transcripts (sometimes, successfully). Community partners engaged with visuals and findings as in earlier projects.

**Box 1: A Pragmatic Workflow for Solo and Smaller-Scale Research**

*The core principles of our workflow can be used for solo researchers, and follow comparative ethnographic projects done with modest resources like student licenses to QDA software and text editors (c.f. Abramson 2015).*

- **Focus on Open-Source Tools:** If you have or can develop the skills, consider free tools in software like Python or R for text processing and analysis. Securely run open-source language models on a local machine to avoid the costs and privacy risks of commercial APIs.
- **Prioritize a Scalable Data Structure:** A consistent file-naming convention and a simple, interoperable data schema from the outset help keep track of all your data—with or without computational tools. Further, using simple file types and formatting ensures your data remains machine-readable as your project grows. A text editor and open source code can provide powerful analyses, at low or 0 cost.
- **Strategic Automation:** Don't try to automate everything (or, if you are in a pragmatic tradition, don't try to do everything with codes– use tags to index what is necessary without reducing explanation to your indexing system). Use dictionary-based indexing for simple, recurring concepts can be captured with high accuracy, and your underlying data remains the full text rather than the classifiers attached. Reserve your limited time for deep coding to complex, interpretive themes; writing memos and time in the field are key to insights, even though less technical.
- **Leverage QDA Software for Integration:** Consider software like ATLAS.ti or MAXQDA or interactive markdown editors (noteplan, obsidian), as a hub to link memos, raw text, and the outputs of your computational scripts, maintaining the connection between different layers of analysis. This parallels work done by sociologists, who sought to archive and curate their own experiences and data. Also, make sure you back up your data in a secure location.
- **Memos**: 'Coding' is important for indexing or theorizing in several traditions, but memos on your findings are the essential link between analysis and writing for many of us– including those who reject software or the code label. We use memos in not only our qualitative work, but programming, statistical analysis, and policy research—as a step between data and final writings.

## Discussion

Qualitative work and computation have long been intertwined, from early formal modeling and qualitative data analysis (Dohan and Sánchez-Jankowskii 1998; Mohr and Duquenne 1997) to narratives (Franzosi 1998), digital repositories (Murphy et al. 2021), and recent interpretive-computational projects (Abramson et al. 2024; Nelson 2020) as well as quantified text (Roberts et al 2022). We have discussed enduring concerns (whether software imposes a singular logic), technical possibilities (scaling and visualization), and pitfalls (environmental and scientific costs) in an era of AI.

Though predicting futures is fraught, one projection seems clear: whether computation is cast as a pinnacle of human achievement, a mechanized cage of algorithms, or a toolkit with potential to harm or help, it is unlikely to disappear so long as society continues in its current form. It saturates everyday life, the worlds we study, and our research, even if we might hope otherwise. How we respond is a matter of bounded agency. Responses will vary even within qualitative fields—a strength, perhaps (Abramson and Gong 2020).Our bias is toward pragmatic and purposeful engagement with tools, which, if used well, can yield accurate insights into health, inequality, culture, labor, and science, even as they carry potential for harm—much like research itself (Abramson 2024).

*Keeping up with the Algo-Joneses*

In discussing the challenges of technology, one theme that comes up but is rarely mentioned is the frenetic pace of development. The phrase *keeping up with the Joneses* refers to the pressure to match others, often in a competitive way. This remains a very real concern in an era of not just computation but quantified scholarship, constantly changing models, and the moral valence of either embracing or rejecting algorithmic tools. Hence, keeping up with the "algo-Joneses" (algo signifying algorithm). Classical sociologists noted that such dynamics can become endemic to modernity—producing systemic inequality and driving individuals to pursue unattainable goals while losing sight of broader social forces (Durkheim 1997; Weber 1904). Contemporary studies of digital society extend these observations, documenting how rankings, feeds, metrification, and digital traces affect life chances (Fourcade and Healy 2024; Mears 2023; Pardo-Guerra 2022).[4]

Keeping up with "the algo-Joneses" can feel futile and mandatory. Yet, as we have shown, possibilities for thoughtful engagement exist alongside challenges. The point is neither to weigh down scholars with despair nor to offer hollow optimism, but to recognize the coexistence of contradictory currents—itself a central feature of sociology. Technology has long been both a tool of creativity and a avenue of control. The question is how to respond when the algo-Joneses continue to bring home new tools, some useful, others alarming.

*Considerations from the Present and Future*

[4] For qualitative scholarship, the issue is close to home. The pressures to adopt new technologies can be understood through a scientific field shifting towards easily quantifiable outputs over long-term ethnographic work (Pardo-Guerra 2022). Google Scholar, for instance, systematically under-indexes book-to-book citations, devaluing the currency of ethnographic monographs while amplifying the metrics of shorter, digitized outputs. The effect, already in motion before the rise of generative AI, exacerbates pressure for everyone to produce more rather than better, and encourages conservative uses of technology for efficiency rather than improving inquiry. Those earlier in their careers, or in less secure environments, are in a more precarious position as the stakes involve the ability to continue to work in social science and earn a living with hard won skills (Fine and Abramson 2021).

We offer the following suggestions, based on reflections using, teaching, and studying technology as applied to qualitative information in fieldwork, interviews, documents, and other sources of qualitative information within and beyond the social sciences:

1. **Computational AND qualitative literacy is essential.** Just as quantitative and qualitative literacy are both necessary for a multi-method field like sociology, knowledge of computational methods is now important to engaging with the present and future of social inquiry. Algorithms, text analysis, simulations, and AI are ubiquitous. Even those who reject quantification should understand these systems to engage critically. This is not without difficulties—some technologies are opaque and the field is shifting—but this remains an important challenge. While working by hand may be necessary in our craft (with value in statistics as well as qualitative research), only knowing how to engage in such a way is unlikely to remain sufficient.
2. **Local systems matter, and help address downsides.** Running small language models on a desktop may lack the spectacle or interface of GPT outputs, but it enables practical uses: fine-tuning transformer models to check codes, visualizing patterns, and validating analysis securely without uploading sensitive data. While all tools are implicated in commodity chains, local use avoids the environmental and confidentiality risks of high-performance commercial systems that some assume are endemic to all computational work.
3. **Human judgment is central.** Automating key decisions risks undermining interpretation. At the same time, burdensome manual tasks can sap the capacity for higher-level reasoning. Where the boundary lies is contested, and perhaps context-dependent, but human oversight is indispensable. Sometimes called 'human in the loop' in the computational world, this principle has always been central to qualitative analysis from the earliest days of QDA, and is core to the qualitative aspects of data analysis.
4. **Disagreement is inevitable.** To dismiss computation entirely—or, conversely, to dismiss those who resist—risks closure rather than dialogue. Whether you use QDA or not, STATA/R or Python, seems an odd distinction when weighed against shared interests like understanding and addressing social inequality. One of the most powerful uses of fieldwork is perspective.
5. **Sociology has a history of novel and even subversive methods.** From Du Bois' data visualizations at the World's Fair to the repurposing of computer systems to chart inequality across America during COVID, tools developed for other ends have long been appropriated for sociological aims. This is perhaps possible for AI too, at least in its broadest senses.
6. **Shared formats.** A key barrier to replication, data sharing, and the transparent integration of computational tools is the lack of a simple, durable data format. Proprietary file types like .qdpx and corporate software are often too complex and bloated. The field would benefit from adopting a simple, common data_frame format—we provide an example of our schema in our appendix—that works across QDA software and programming languages, enhancing interoperability and decreasing the need to rely on expensive commercial software. Such a format doesn't have to do everything, just enough to allow communication and perhaps examine targeted verifications which have been occurring around large scale projects.
7. **Open, critical engagement.** Rather than racing to keep up, scholars might maintain openness to emergent technologies while critically assessing their limits and possibilities.

What is useful in one case may not generalize, but sharing tools and methods can strengthen collective practice—even if it is not called 'open science' or 'science.'

*Concluding thoughts*

The issues raised here—of technological change, institutional response, and methodological conservation—underscore the continued value of a sociological vantage. We write in a moment of rapid transformation and unsettled times, that call for consideration of both intended and potential uses of computers, influence and risk, danger and possibility. To adopt a pragmatic approach to qualitative data analysis, workflow, and computation is neither an embrace of technology as progress nor a rejection based on dystopian applications. Rather, it reflects the view that dualisms are a luxury, and that any approach to social life now entails a position on computation—whether explicit or implicit, deliberate or de facto—and that this is unlikely to change.

To paraphrase Weber: the future is uncertain, but it will likely include computers—and AI—in ways that both constrain and enable social life beyond the intentions of those who live today. Our aim has not been to pre-emptively adjudicate whether such developments advance a greater good—a question beyond this paper's scope—but to connect divergent positions to concrete methodological issues, ethical considerations, and practical strategies for qualitative and computational research.

Our hope is that by:

1. Clarifying how qualitative and computational research have and continue to intersect,
2. Offering a typology of modes of engagement in an era of automation and AI,
3. Walking through grounded, real-world workflows, and
4. Reflecting on the challenges and possibilities of pragmatic sociological practices,

—we have offered not just points to think about, but tools to work with, and perhaps support inquiry and human agency.

For implementation details, additional references, and open-source code, please see the appendix and linked repository.

## References


1. Abdurahman, Suhaib, Alireza Salkhordeh Ziabari, Alexander K. Moore, Daniel M. Bartels, and Morteza Dehghani. 2025. "A Primer for Evaluating Large Language Models in Social-Science Research." *Advances in Methods and Practices in Psychological Science* 8(2):25152459251325174. doi:10.1177/25152459251325174.

2. Abramson, C. M., and D. Dohan. 2015. "Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data." *Sociological Methodology* 272–319. doi:10.1177/0081175015578740.

3. Abramson, Corey M. 2015. *The End Game: How Inequality Shapes Our Final Years*. Cambridge, Massachusetts: Harvard University Press.

4. Abramson, Corey M. 2024. “From Carbon Paper to Code: Crafting Sociology in an Age of AI.” *Contexts: Guest Post*. (link).

5. Abramson, Corey M., and Neil Gong. 2020. “Introduction: The Promise, Pitfalls, and Practicalities of Comparative Ethnography.” in *Beyond the Case: The Logics and Practices of Comparative Ethnography*, edited by C. M. Abramson and N. Gong. New York: Oxford University Press.

6. Abramson, Corey M., Jacqueline Joslyn, Katharine A. Rendle, Sarah B. Garrett, and Daniel Dohan. 2018. “The Promises of Computational Ethnography: Improving Transparency, Replicability, and Validity for Realist Approaches to Ethnographic Analysis.” *Ethnography* 19(2):254–84. doi:10.1177/1466138117725340.

7. Abramson, Corey M., Zhuofan Li, Tara Prendergast, and Martín Sánchez-Jankowski. 2024. “Inequality in the Origins and Experiences of Pain: What ‘Big (Qualitative) Data’ Reveal About Social Suffering in the United States.” *RSF: The Russell Sage Foundation Journal of the Social Sciences* 10(5):34. doi:10.7758/RSF.2024.10.5.02.

8. Abramson, Corey M., and Martín Sánchez-Jankowski. 2020. “Conducting Comparative Participant Observation: Behavioralist Procedures and Techniques.” in *Beyond the Case: The Logics and Practices of Comparative Ethnography*, edited by C. M. Abramson and N. Gong. New York: Oxford University Press.

9. Amirova, Aliya, Theodora Fteropoulli, Nafiso Ahmed, Martin R. Cowie, and Joel Z. Leibo. 2024. “Framework-Based Qualitative Analysis of Free Responses of Large Language Models: Algorithmic Fidelity.” *PLOS ONE* 19(3):e0300024. doi:10.1371/journal.pone.0300024.

10. Annette Lareau and Aliya Hamid Rao. 2016. “It’s about the Depth of Your Data - Contexts.” *Contexts*. https://contexts.org/blog/its-about-the-depth-of-your-data/.

11. Arteaga, I., A. Hernández de Jesús, B. Ginn, Corey M. Abramson, and Daniel Dohan. “Understanding How Social Context Shapes Decisions to Seek Institutional Care: A Qualitative Study of Latinx Families Navigating Progressive Cognitive Decline.” *The Gerontologist*.

12. Ashwin, Julian, Aditya Chhabra, and Vijayendra Rao. 2025. “Using Large Language Models for Qualitative Analysis Can Introduce Serious Bias.” *Sociological Methods & Research* 00491241251338246. doi:10.1177/00491241251338246.

13. Astrupgaard, Sofie L., August Lohse, Emilie M. Gregersen, Jonathan H. Salka, Kristoffer Albris, and Morten A. Pedersen. 2024. “Fixing Fieldnotes: Developing and Testing a Digital Tool for the Collection, Processing, and Analysis of Ethnographic Data.” *Social Science Computer Review* 42(5):1223–43. doi:10.1177/08944393231220488.

14. Bail, Christopher A. 2024. “Can Generative AI Improve Social Science?” *Proceedings of the National Academy of Sciences* 121(21). doi:10.1073/pnas.2314021121.

15. Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜.” Pp. 610–23 in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM.

16. Bender, Emily M., and Alex Hanna. 2025. *The AI Con: How to Fight Big Tech’s Hype and Create the Future We Want*. New York: Harper.

17. Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK Medford, MA: Polity.

18. Bernstein, Alissa, and Daniel Dohan. 2020. “USING COMPUTATIONAL TOOLS TO ENHANCE COMPARATIVE ETHNOGRAPHY: LESSONS FROM SCALING ETHNOGRAPHY FOR BIOMEDICINE.” in *Beyond the Case: The Logics and Practices of Comparative Ethnography*, edited by C. M. Abramson and N. Gong. New York: Oxford University Press.

19. Biernacki, Richard. 2014. “Humanist Interpretation Versus Coding Text Samples.” *Qualitative Sociology* 37(2):173–88. doi:10.1007/s11133-014-9277-9.

20. Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson. 2024. “Synthetic Replacements for Human Survey Data? The Perils of Large Language Models.” *Political Analysis* 32(4):401–16. doi:10.1017/pan.2024.5.

21. Bjerre-Nielsen, Andreas, and Kristoffer Lind Glavind. 2022. “Ethnographic Data in the Age of Big Data: How to Compare and Combine.” *Big Data & Society* 9(1):20539517211069893. doi:10.1177/20539517211069893.

22. Boellstorff, Tom. 2008. *Coming of Age in Second Life : An Anthropologist Explores the Virtually Human*. Princeton, N.J: Princeton University Press.

23. Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

24. Boutyline, Andrei, and Alina Arseniev-Koehler. 2025. “Meaning in Hyperspace: Word Embeddings as Tools for Cultural Measurement.” doi:10.1146/annurev-soc-090324-024027.

25. Boutyline, Andrei, and Laura K. Soter. 2021. “Cultural Schemas: What They Are, How to Find Them, and What to Do Once You’ve Caught One.” *American Sociological Review* 86(4):728–58. doi:10.1177/00031224211024525.

26. Breiger, Ronald L. 1974. “The Duality of Persons and Groups.” *Social Forces* 53(2):181–90.

27. Breiger, Ronald L. 2015. “Scaling Down.” *Big Data & Society* 83(7):42–44. doi:10.1177/2053951715602497.

28. Burawoy, Michael. 1998. “The Extended Case Method.” *Sociological Theory* 16(1):4–33.

29. Burrell, Jenna. 2016. “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms.” *Big Data & Society* 3(1):2053951715622512. doi:10.1177/2053951715622512.

30. Burrell, Jenna, and Marion Fourcade. 2021. “The Society of Algorithms.” *Annual Review of Sociology* 47(1):213–37. doi:10.1146/annurev-soc-090820-020800.

31. Chandrasekar, Abinaya, Sigrún Eyrúnardóttir Clark, Sam Martin, Samantha Vanderslott, Elaine C. Flores, David Aceituno, Phoebe Barnett, Cecilia Vindrola-Padros, and Norha Vera San Juan. 2024. “Making the Most of Big Qualitative Datasets: A Living Systematic Review of Analysis Methods.” *Frontiers in Big Data* 7. doi:10.3389/fdata.2024.1455399.

32. Chen, Lingjiao, Matei Zaharia, and James Zou. 2024. “How Is ChatGPT’s Behavior Changing Over Time?” *Harvard Data Science Review* 6(2). doi:10.1162/99608f92.5317da47.

33. Chopra, Felix, and Ingar Haaland. 2023. “Conducting Qualitative Interviews with AI.”

34. Christin, Angèle. 2020. “The Ethnographer and the Algorithm: Beyond the Black Box.” *Theory and Society* 49(5–6):897–918. doi:10.1007/s11186-020-09411-3.

35. Cicourel, Aaron V. 1982. “Interviews, Surveys, and the Problem of Ecological Validity.” *The American Sociologist* 17(1):11–20.

36. Clifford, James, and George E. Marcus. 1986. *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley CA: University of California Press.

37. Coffey, Amanda, Holbrook Beverley, and Atkinson Paul. 1996. “Qualitative Data Analysis: Technologies and Representations.” *Sociological Research Online* 1(1):80–91. doi:10.5153/sro.1.

38. Collins, Hill Patricia. 2000. “Black Feminst Thought.” *Black Feminist Thought* 335.

39. Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

40. Davidson, Thomas. 2024. “Start Generating: Harnessing Generative Artificial Intelligence for Sociological Research.” *Socius* 10:23780231241259651. doi:10.1177/23780231241259651.

41. DeLuca, Stefanie, Susan Clampet-lundquist, and Kathryn Edin. 2016. “Want to Improve Your Qualitative Research? Try Using Representative Sampling and Working in Teams.” *Contexts (Online)*. https://contexts.org/blog/its-about-the-depth-of-your-data/.

42. Deterding, Nicole M., and Mary C. Waters. 2021. “Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach.” *Sociological Methods & Research* 50(2):708–39. doi:10.1177/0049124118799377.

43. Dohan, Daniel, and Martín Sánchez-Jankowski. 1998. “Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations.” *Annual Review of Sociology* 24(1):477–98. doi:10.1146/annurev.soc.24.1.477.

44. Dohan, Daniel, and Martin Sanchez-Jankowski. 1998. “Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations.” *Annual Review of Sociology* 24:477–98.

45. Du Bois, W. E. B. 1900. “The Exhibit of American Negroes.” *Exposition Universelle (Paris, 1900)*.

46. DuBOIS, W. E. B., and Isabel Eaton. 1996. *The Philadelphia Negro: A Social Study*. University of Pennsylvania Press.

47. Durkheim, Emile. 1997. *Suicide: A Study in Sociology*. New York: Free Press.

48. Earl, Jennifer, and Katrina Kimport. 2013. *Digitally Enabled Social Change: Activism in the Internet Age*. Acting with Technology. Cambridge, MA: The MIT Press.

49. Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. “Computational Social Science and Sociology.” *Annual Review of Sociology* 46(Volume 46, 2020):61–81. doi:10.1146/annurev-soc-121919-054621.

50. Edin, Kathryn J., Corey D. Fields, David B. Grusky, Jure Leskovec, Marybeth J. Mattingly, Kristen Olson, and Charles Varner. 2024. “Listening to the Voices of America.” *RSF: The Russell Sage Foundation Journal of the Social Sciences* 10(4):1–31. doi:10.7758/RSF.2024.10.4.01.

51. Farber, Orly, Corey Abramson, and Amanda Reich. 2025. “Utilizing Artificial Intelligence to Facilitate Qualitative Surgical Research.” *Annals of Surgery (Open)*.

52. Farquhar, Sebastian, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. 2024. “Detecting Hallucinations in Large Language Models Using Semantic Entropy.” *Nature* 630(8017):625–30. doi:10.1038/s41586-024-07421-0.

53. Fine, Gary Alan, and Corey M. Abramson. 2020. “Ethnography in the Time of Covid-19. Vectors and the Vulnerable.” *Etnografia e Ricerca Qualitativa* 13(2):165–165.

54. Foucault, Michele. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Random House.

55. Fourcade, Marion, and Kieran Healy. 2024. *The Ordinal Society*. Cambridge, MA: Harvard University Press.

56. Franzosi, Roberto. 1998. “Narrative Analysis—Or Why (and How) Sociologists Should Be Interested In Narrative.” *Annual Review of Sociology* 24(1):517–54. doi:10.1146/annurev.soc.24.1.517.

57. Freese, Jeremy, Tamkinat Rauf, and Jan Gerrit Voelkel. 2022. "Advances in Transparency and Reproducibility in the Social Sciences." *Social Science Research* 107:102770. doi:10.1016/j.ssresearch.2022.102770.

58. Gans, Herbert J. 1999. "Participant Observation in the Era of 'Ethnography.'" *Journal of Contemporary Ethnography* 28(5):540-48-540–48.

59. Garrett, Sarah B., Corey M. Abramson, Katharine A. Rendle, and Daniel Dohan. 2019. "Approaches to Decision-Making among Late-Stage Melanoma Patients: A Multifactorial Investigation." *Supportive Care in Cancer* 27(3):1059–70. doi:10.1007/s00520-018-4395-7.

60. Geertz, Clifford. 2000. "Thick Description: Towards an Interpretive Theory of Culture." Pp. 3-30-3–30 in. New York: Basic Books.

61. Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine Publishing Company.

62. Gong, Neil, and Corey M. Abramson. 2020. "Conclusion: A Comparative Analysis of Comparative Ethnographies." P. 0 in *Beyond the Case: The Logics and Practices of Comparative Ethnography*, edited by C. M. Abramson and N. Gong. Oxford University Press.

63. Grigoropoulou, Nikolitsa, and Mario L. Small. 2022. "The Data Revolution in Social Science Needs Qualitative Research." *Nature Human Behaviour* 6(7):904–6. doi:10.1038/s41562-022-01333-7.

64. Grossmann, Igor, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. 2023. "AI and the Transformation of Social Science Research." *Science* 380(6650):1108–9. doi:10.1126/science.adi1778.

65. Hanson, Kenneth R., and Nicholas Theis. 2024. "Networked Participants, Networked Meanings: Using Networks to Visualize Ethnographic Data." *Sociological Methodology* 54(1):142–56. doi:10.1177/00811750231195338.

66. Hurd, Thelma C., Fay Cobb Payton, and Darryl B. Hood. 2024. "Targeting Machine Learning and Artificial Intelligence Algorithms in Health Care to Reduce Bias and Improve Population Health." *The Milbank Quarterly* 102(3):577–604. doi:10.1111/1468-0009.12712.

67. Jasanoff, Sheila, ed. 2004. *States of Knowledge: The Co-Production of Science and the Social Order*. Routledge.

68. Jerolmack, C., and S. Khan. 2014. "Talk Is Cheap: Ethnography and the Attitudinal Fallacy." *Sociological Methods & Research* 43(2):178–209. doi:10.1177/0049124114523396.

69. Karcher, Sebastian, Dessislava Kirilova, Christiane Pagé, and Nic Weber. 2021. "How Data Curation Enables Epistemically Responsible Reuse of Qualitative Data." *The Qualitative Report* 26(6):1996–2010. doi:10.46743/2160-3715/2021.5012.

70. Kirilova, Dessi, and Sebastian Karcher. 2017. "Rethinking Data Sharing and Human Participant Protection in Social Science Research: Applications from the Qualitative Realm." *Data Science Journal* 16(0). doi:10.5334/dsj-2017-043.

71. Kozlowski, Austin C., and James Evans. 2025. "Simulating Subjects: The Promise and Peril of Artificial Intelligence Stand-Ins for Social Agents and Interactions." *Sociological Methods & Research* 54(3):1017–73. doi:10.1177/00491241251337316.

72. Lamont, Michèle, and Ann Swidler. 2014. "Methodological Pluralism and the Possibilities and Limits of Interviewing." *Qualitative Sociology* 37(2):153–71.

73. Law, Tina, and Elizabeth Roberto. 2025. “Generative Multimodal Models for Social Science: An Application with Satellite and Streetscape Imagery.” *Sociological Methods & Research* 54(3):889–932. doi:10.1177/00491241251339673.

74. Li, Zhuofan, and Corey M. Abramson. 2025. “Ethnography and Machine Learning: Synergies and New Directions.” in *The Oxford Handbook of the Sociology of Machine Learning*, edited by C. Borch and J. P. Pardo-Guerra. Oxford University Press.

75. Li, Zhuofan, Daniel Dohan, and Corey M. Abramson. 2021. “Qualitative Coding in the Computational Era: A Hybrid Approach to Improve Reliability and Reduce Effort for Coding Ethnographic Interviews.” *Socius* 7. doi:10.1177/23780231211062345.

76. Lichtenstein, Matty, and Zawadi Rucks-Ahidiana. 2023. “Contextual Text Coding: A Mixed-Methods Approach for Large-Scale Textual Data.” *Sociological Methods & Research* 52(2):606–41. doi:10.1177/0049124120986191.

77. Mears, Ashley. 2023. “Bringing Bourdieu to a Content Farm: Social Media Production Fields and the Cultural Economy of Attention.” *Social Media + Society* 9(3):20563051231193027. doi:10.1177/20563051231193027.

78. Merton, Robert K. 1938. “Social Structure and Anomie.” *American Sociological Review* 3(5):672–82.

79. Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.

80. Mische, Ann, and Şehrazat G. Mart. 2025. “Imagined Interventions: Critical-Ideational Strategies across Multiple Futures in Public Interest Scenarios.” *Time & Society* 34(1):77–105. doi:10.1177/0961463X241310564.

81. Mohr, John W., and Vincent Duquenne. 1997. “The Duality of Culture and Practice: Poverty Relief in New York City, 1888--1917.” *Theory and Society* 26(2–3):305–56. doi:10.1023/A:1006896022092.

82. Murphy, Alexandra K., Colin Jerolmack, and DeAnna Smith. 2021. “Ethnography, Data Transparency, and the Information Age.” *Annual Review of Sociology* 47(1):annurev-soc-090320-124805. doi:10.1146/annurev-soc-090320-124805.

83. Nelson, Laura K. 2020. “Computational Grounded Theory: A Methodological Framework.” *Sociological Methods & Research* 49(1):3–42. doi:10.1177/0049124117729703.

84. Nelson, Laura K., Derek Burk, Marcel Knudsen, and Leslie McCall. 2021. “The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods.” *Sociological Methods & Research* 50(1):202–37. doi:10.1177/0049124118769114.

85. Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.

86. Omiye, Jesutofunmi A., Jenna C. Lester, Simon Spichak, Veronica Rotemberg, and Roxana Daneshjou. 2023. “Large Language Models Propagate Race-Based Medicine.” *Npj Digital Medicine* 6(1):195. doi:10.1038/s41746-023-00939-z.

87. O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

88. Pardo-Guerra, Juan Pablo. 2022. *The Quantified Scholar: How Research Evaluations Transformed the British Social Sciences*. New York (N.Y.): Columbia University press.

89. Pardo-Guerra, Juan Pablo, and Prithviraj Pahwa. 2022. “The Extended Computational Case Method: A Framework for Research Design.” *Sociological Methods & Research* 51(4):1826–67. doi:10.1177/00491241221122616.

90. Parker, Jessica, Veronica Richard, and Kimberly Becker. 2023. "Guidelines for the Integration of Large Language Models in Developing and Refining Interview Protocols." *The Qualitative Report* 28(12):3460–74. doi:10.46743/2160-3715/2023.6801.

91. Paula, O'Kane. 2020. "Demystifying CAQDAS: A Series of Dilemmas." Pp. 133–52 in *Advancing Methodological Thought and Practice*. Vol. 12, edited by T. R. Crook, J. Lê, and A. D. Smith. Emerald Publishing Limited.

92. Paulus, Trena M., and Vittorio Marone. 2024. "'In Minutes Instead of Weeks': Discursive Constructions of Generative AI and Qualitative Data Analysis." *Qualitative Inquiry* 31(5):395–402. doi:10.1177/10778004241250065.

93. Peponakis, Manolis, Sarantos Kapidakis, Martin Doerr, and Eirini Tountasaki. 2023. "From Calculations to Reasoning: History, Trends and the Potential of Computational Ethnography and Computational Social Anthropology." *Social Science Computer Review* 089443932311676. doi:10.1177/08944393231167692.

94. Pugh, Allison. 2024. "The Last Human Job | Princeton University Press." https://press.princeton.edu/books/hardcover/9780691240817/the-last-human-job.

95. Roberts, M. E. Grimmer, J. Stewart, B. M. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press.* Princeton University Press.

96. Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven (Conn.): Yale University press.

97. Rodriguez, Pedro L., and Arthur Spirling. 2022. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *The Journal of Politics* 84(1):101–15. doi:10.1086/715162.

98. Rosa, Fernanda R. 2022. "Code Ethnography and the Materiality of Power in Internet Governance." *Qualitative Sociology* 45(3):433–55. doi:10.1007/s11133-022-09517-3.

99. Sanchez-Jankowski, Martin. 2002. "Representation, Responsibility and Reliability in Participant Observation." Pp. 144–60 in *Qualitative Research in Action*, edited by T. May. London: Sage.

100. Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. "Whose Opinions Do Language Models Reflect?" Pp. 29971–4 in *Proceedings of the 40th International Conference on Machine Learning*. PMLR.

101. Seaver, Nick. 2017. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data & Society* 4(2):205395171773810. doi:10.1177/2053951717738104.

102. Silver, Christina. 2018. "CAQDAS at a Crossroads: Choices, Controversies and Challenges." Pp. 1–13 in *Computer Supported Qualitative Research*, edited by A. P. Costa, L. P. Reis, F. N. de Souza, and A. Moreira. Cham: Springer International Publishing.

103. Simon, Felix M., Sacha Altay, and Hugo Mercier. 2023. "Misinformation Reloaded? Fears about the Impact of Generative AI on Misinformation Are Overblown." *Harvard Kennedy School Misinformation Review*. doi:10.37016/mr-2020-127.

104. Small, Mario Luis. 2011. "How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature." *Annual Review of Sociology* 37(1):57-86-57–86. doi:10.1146/annurev.soc.012809.102657.

105. Small, Mario Luis, and Jessica McCrory Calarco. 2022. *Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research*. Oakland, California: University of California Press.

106. Spirling, Arthur. 2023. "Why Open-Source Generative AI Models Are an Ethical Way Forward for Science." *Nature* 616(7957):413–413.

107. Steinmetz, George. 2005. *The Politics of Method in the Human Sciences*. Duke University Press.

108. Strauss, Anselm L., and Juliet M. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications.

109. Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2020. "Energy and Policy Considerations for Modern Deep Learning Research." *Proceedings of the AAAI Conference on Artificial Intelligence* 34(09):13693–96. doi:10.1609/aaai.v34i09.7123.

110. Stuhler, Oscar, Cat Dang Ton, and Etienne Ollion. 2025. "From Codebooks to Promptbooks: Extracting Information from Text with Generative Large Language Models." *Sociological Methods & Research* 54(3):794–848. doi:10.1177/00491241251336794.

111. Sykes, Bryan L., Anjuli Verma, and Black Hawk Hancock. 2018. "Aligning Sampling and Case Selection in Quantitative-Qualitative Research Designs: Establishing Generalizability Limits in Mixed-Method Studies." *Ethnography* 19(2):227–53. doi:10.1177/1466138117725341.

112. Tan, Yi Chern, and L. Elisa Celis. 2019. "Assessing Social and Intersectional Biases in Contextualized Word Representations." in *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

113. The Ghost in the Algorithm: Racial Colonial Capitalism and the Digital Age. 2021a. Pp. 221–49 in *Political Power and Social Theory*. Emerald Publishing Limited.

114. Weber, Jutta, and Bianca Prietl. 2021. "AI in the Age of Technoscience: On the Rise of Data-Driven AI and Its Epistem-Ontological Foundations." in *The Routledge Social Science Handbook of AI*. Routledge.

115. Weber, Max. 1978. *Economy and Society*. Vol. I. Berkeley: University of California Press.

116. Weber, Max. n.d. *The Protestant Ethic and the Spirit of Capitalism*. New York: Routledge Classics.

117. Yadlin, Aya, and Avi Marciano. 2024. "Hallucinating a Political Future: Global Press Coverage of Human and Post-Human Abilities in ChatGPT Applications." *Media, Culture & Society* 46(8):1580–98. doi:10.1177/01634437241259892.

118. Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, NY: PublicAffairs.

# APPENDIX & SUPPLEMENTS

## I. WORKFLOW

### TABLE 2 EXEGISIS

*Research Design.* At the outset of projects, computational tools are increasingly used in designing studies and instruments. For instance, in some qualitative studies large language models (LLMs) are used to synthesize literature for systematic reviews through both custom API and commercial systems (Parker et al. 2023). LLMs are also used to test the readability of interview questions prior to piloting, to help with initial translations on recruitment fliers, and to check for typos. Many of these options existed in some form in word processing software, but the tools have expanded and barriers to entry are lower. Some simulate potential participants or scenarios to stress-test protocols (Kapania et al. 2024; Kuipers 1986) and use in field experiments is expanding. Project management and version control systems (e.g., OSF, GitHub) support collaborative design and documentation, sometimes including automations but also usable without AI (Farber et al. 2025).

*Data Collection.* New applications extend into the field itself. Automated transcription services (Zhou 2025) and digital consent platforms may aid in the capture of field materials or digital documents and, representing a bridge between surveys, time diaries, and qualitative accounts, researchers have employed mobile platforms to collect diaries and ecological momentary assessments, augmented by automated reminders or AI-generated prompts (Farber et al. 2025). In a controversial data collection application, AI agents have been deployed as interviewers (Chopra and Haaland 2023). The argument is that chatbots can conduct structured text collection, sometimes enabling participants to feel more comfortable sharing sensitive experiences than with a human (Chopra and Haaland 2023). At present, to our knowledge, this remains a rhetorical question, though increasing information is suggesting the conventional wisdom that in-person interviews produce more accurate data is not substantiated in tests and the technology is changing quickly (Johnson et al. 2019).

*Data Processing.* Once collected, qualitative data increasingly undergo systematic preprocessing with computational assistance, particularly for projects conducted at scale. Through a combination of named entity recognition (NER) and human review of the AI processing on a local computer, current technologies can facilitate de-identification of transcripts to protect confidentiality. Scripts written in Python or R can automate the conversion of text, audio, and images into standardized formats for computational analysis and easier human reading, while metadata tagging and automated quality checks can enable analysis by data type, demographic group, or topic in qualitative data analysis software (Li and Abramson 2025). These steps extend earlier practices of structure and organization (Dohan and Sánchez-Jankowski 1998) into more formalized, computationally reproducible pipelines that can be shared as reproducible code in accordance with the principles of open science (Freese and Peterson 2017).

*Data Analysis***.** The greatest proliferation of uses lies in analysis, at least by current estimates. For instance, hybrid approaches to coding integrate machine learning classifiers or LLM-assisted coding into qualitative research. Uses range from largely inductive (searching for emerging patterns) to deductive (applying existing categories), in ways that reflect various approaches to social science. Computational grounded theory (Nelson 2020; Alqazlan et al. 2025) exemplifies an approach in which algorithmic analyses help surface patterns inductively. Analysis is refined through human interpretive reading to parallel a focus on emergent themes. In another approach, expanded iterative qualitative coding, iterative coding (Deterding and Waters 2021) – already augmented with basic automation in flexible coding – has been expanded to include hybrid pipelines that allow language models to use human codings to learn how to classify text. Human researchers review the results and conduct higher-order interpretation, using both deductive concepts (e.g. study domains or hypotheses) and inductive findings (e.g. emergent patterns), as part of the indexing process (Li et al. 2021). This can be done with high consistency and accuracy using machine learning on secure local computers, sometimes identifying relevant data humans may have missed. For example, Yoon and McCumber (2024) used a combination of natural language processing methods to inductively construct a topic-by-geography typology of New York Times Travel articles, which then guides their qualitative analysis of the symbolic hierarchy in the cultural narratives about places. Finally, the deductive application of categories, once a hallmark of quantitative content analysis and some formal approaches can be applied at scale with good accuracy using a variety of models (Than et al. 2025).

*Data generation and simulation.* With the rise of large language models, some researchers have begun experimenting with synthetic data in the social sciences. A common design is to prompt models to simulate "empirically realistic, culturally situated human subjects" or "personas" by reproducing real survey responses from the General Social Survey or World Values Survey (Bisbee et al. 2024; Boelaert et al. 2025; Kozlowski and Evans 2025). [5]

*Writing and Presentation.* AI and automation support writing and dissemination, though both are points of contention. LLMs (both offline and commercial) can draft memos, generate figures, summarize and can create tables from coded or uncoded data (Farber et al. 2025). Collaborative platforms allow for live editing, and literature agents automate the ability to web scrape scholarly articles and link to bibliographic databases. Word processors have long included spell checks, but increasingly use advanced tools or apps (e.g. Grammarly) to edit, rewrite, and even write academic work, raising possibilities for clearer writing, as well as concerns about subtle adjustments to scholarly positions (and fabrication). In terms of presentation, computational tools allow for novel visualizations of qualitative data from ethnoarray heatmaps that show patterns of cases (Abramson and Dohan 2015) to embeddings and network analyses (Hanson and Theis 2024; Basov et al. 2021). A rich tradition in cultural sociology has been using network, algebraic and vector structures to visualize spaces of cultural meanings (Bearman and Stovel 2000; Breiger 2000; Carley 1994; Mohr 1994). Recently, using word embeddings (vector representations of language) to map culture in historical texts and subject speech as part of broader examinations of

[5] These outputs, however, tend to be more biased, less variable, temporally unstable, and more likely to flatten or essentialize identity groups than real human responses (Argyle et al. 2023; Bisbee et al. 2024; Boelaert et al. 2025; Kozlowski and Evans 2025; Wang et al. 2024; Zhang et al. 2025). Commercial models are also shaped by "model alignment"—human-driven, often undisclosed modifications intended to make outputs more consistent with provider and user expectations (Ji et al. 2023). While alignment may reduce harmful language and bias, it also makes models less useful when the analytic goal is to study prejudice, misinformation, or other social processes (Lyman et al. 2025). Even for qualitative researchers who do not adopt data generation as a research design, these issues matter for data analysis. Platform users increasingly employ LLMs to answer open-ended survey questions (Zhang et al. 2025), bots already account for an estimated 20% of social media interactions, and many are becoming LLM-powered—injecting synthetic text into qualitative research materials (Ng and Carley 2025; Radivojevic et al. 2024). While some organizations resist LLMs for data safety reasons (Serbu 2024), others—including universities—are embracing them (Singer 2025), accelerating the entry of synthetic text into administrative records, archival sources, and the broader qualitative corpus. As synthetic content becomes more pervasive, researchers may need to account for it not only in study design but in every stage of qualitative data analysis.

patterns of meaning, is gaining traction in a variety of fields (Boutyline and Arseniev-Koehler 2025; Koslowski et al. 2019).

*Sharing and Preservation:* The expansion of infrastructure enables the technical possibility for making qualitative findings more shareable, with code and analytic notebooks enabling replication—or websites linking text segments for interaction– though questions of understanding, representation and interpretation remain central to sociological approaches (Abramson and Dohan 2015; Murphy et al 2021). Researchers are also using AI to document their work and workflow in new ways. These include analytic processes like "promptbooks" (Stuhler et al. 2025) and extending the idea of a living codebook that captures indexing procedures in the era of AI through charting the evolution of computer instructions to index themes and patterns (Reyes et al. 2024).

Taken together, the applications above highlight computation-enabled workflow changes and possibilities, ranging from incremental to potentially transformative. Some are common (e.g., spell checking, using Google Scholar for literature), others specialized or emergent. Many practices using computation remain recognizably qualitative: iterative coding, interpretive reading, theorizing from cases, combining induction and deduction to speak to explanations, visualizing data, and triangulating analyses have deep roots in sociology. Yet each stage of the research process is increasingly mediated by computational tools that promise efficiency, scalability, and transparency—while raising enduring concerns about validity, bias, and epistemological fit (Ashwin, Chhabra, and Rao 2023; Amirova et al. 2024)—whether their use is purposeful or simply determined by the technology itself. For example, even basic web use, of systems like google search, are mediated by algorithms in ways beyond user control (Nobel 2018; Pasquale 2016).

## II. GENERALIZED WORKFLOW & TOOLS

[see also Abramson and Dohan 2015; Li and Abramson 2025]

<u>Research Design</u>

- Problem-focused: social and scientific issues → study aims → site assessment/selection → rapport with key members → interviews/observations (network referrals when useful) → add sites/actors to capture diverse and disconfirming cases.
- Logic: iterative, not strictly inductive or deductive.
- Tools: planning (Slack, Box); data tracking (REDCap); analysis (R/Python for quant; ATLAS.ti for qual); archiving (Box).

<u>Data Collection</u>

- Interviews and observations proceed as rapport develops; sampling adapts to emerging questions and disconfirming cases.
- Tools: field logs and files (Box), coordination (Slack), consent/participant tracking (REDCap).

Data Processing

- Transcription
- Deidentification
- Paragraph-level indexing, exportable in multiple formats.
- Storage pipeline links raw transcripts, ATLAS.ti projects, and Python/R notebooks; identifiers stay walled off from analytic text; de-identification checks run in this pipeline.

Data Analysis

- Collection and indexing co-evolve; each cycle informs the next round of fieldwork.
- Analysts move from full corpus → single paragraph without reformatting in ATLAS.ti or Python/R.
- New codes add cleanly with backward compatibility; subsets for analysis use etadata, and range from a few interviews to the full dataset.
- Computational tools visualize themes, surface latent patterns, validate typologies, and flag edge cases; qualitative reading and memoing remain central.

Writing & Presentation

- Computational outputs support memos, figures, and narrative, with links to source text.

Sharing & Preservation

- Files live in encrypted, regularly backed-up environments; personal identifiers are separate from analytic text.

## III. SCHEMA:

Our computational data_frame uses a unified schema, to make text machine readable, post-cleaning, and connected to metadata, often using tabulated QDA export. See Li, Dohan and Abramson 2021 and example of subsetting here.

```
# Schema with Python typing
schema = {
  "project": str, # List project, e.g. DISCERN
  "number": str, # Position information, e.g. chronological ID
  "reference": int, # position information, e.g. added information
  "text": str,  # Content, critical field: must not be empty, e.g. actual text
  "document": str,  # Data source, Critical field; e.g. OA010_iv1_20250411.txt
  "old_codes": list[str],  # Optional: codings, must be a list of strings, for codes not active
  "start_position": int, # Position information; character and QDA info
  "end_position": int, # Position information; character and QDA info
  "data_group": list[str],  # Optional, to differentiate document sets: Must be a list of strings
  "text_length": int, # Optional: NLP info
  "word_count": int, # Optional: NLP info
  "doc_id": str, # Optional: NLP info, unique paragrah level identifier
```

    "codes": list[str]  # critical for analyses with codes, Must be a list of strings; active filter codes
}

D21116 fx CG020So he grew up-- actually, they said he was born here, believe it or not, but at two years old he went back to China. I dont know where, was Jiangshan or somewhere, or maybe Quanzhou ph?, part-- or somewhere in China, laughs know where. So he-- and then he c

Find in Document: So he grew up--

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | project | number | reference | text | document | old_codes | start_positior | end_position | data_group | text_length | word_count | doc_id | codes |
| 21116 | discern | 264:13:00 | 43 | CG020So he | sfggp_20230 | ['_cg020', '_z | 43 | 43 | ['caregiver', 'c | 1422 | 274 | discern_243( | ['caregiver', 'migration'] |
| 21117 | discern | 264:14:00 | 197 | [PERSON] IN | sfggp_20230 | ['_intv_mm', ' | 197 | 197 | ['caregiver', 'c | 64 | 10 | discern_243( | ['caregiver', 'interviewer'] |
| 21118 | discern | 264:15:00 | 443 | [PERSON] IN | sfggp_20230 | ['_intv_mm', ' | 443 | 443 | ['caregiver', 'c | 25 | 4 | discern_243( | ['caregiver', 'interviewer'] |
| 21119 | discern | 264:16:00 | 173 | CG020 INTV_ | sfggp_20230 | ['_intv_mm', ' | 173 | 173 | ['caregiver', 'c | 195 | 35 | discern_2431 | ['caregiver', 'interviewer'] |

Here is a simple example of some rows without text, opened in excel from a .csv file. Text is post-deidentification.

## IV. EXAMPLES

### TEXT FORMATTING

Output deidentified, and checked for proper encoding, formatting for machine readability and interoperability between a data frame and QDA software.

```
✅ NIH/HIPAA Safe Harbor Standard MET

=== Final Assessment ===
✅ Deidentification process successful and meets NIH standards

Processed 78 files with 179 total replacements

Log file created: deidentification_20250513_162601.log
```

### *Interviews*

Two hard returns between interviewer and interviewee. One hard return between paragraphs on multi-paragraph answers. Always using ID: to indicate speaker, and allowing subtracting interviewer speech from modeling with natural language processing (Li and Abramson 2025).

> INTV_BG: Lovely. Do you know if she's in the memory care area, or if she is assisted living, separate from?
>
> CG021: She is technically in memory care, but she is possibly the person that needs the least care, from my observation, on her floor. Her floor is where other people come to have a meal, or the lounge, where they do the games, and the harp. I mean, it's really nice. There's a harpist that comes. It's really fancy <laughs>. Although, though, it is called memory care, yeah, because she has to have some help with, certainly, taking her medicines, and just nudging her to having good, healthy behavior. But I think they help her with laundry. There are some things that she was still doing at home. Like she could make her own tea and coffee, and heat up food. But she wouldn't really eat very well. Even with really healthy options, she would still want to just eat ice cream, root beer floats <laughs>.

### *Fieldnotes*

Use paragraphs for interactional sequences. Use “” for verbatim quotes. Indicate your field position. Use [] to indicate personal experiences or commentary. Use IDs for subjects, like

OA003. Python script adds Z_FN: and or IDS of subjects for data linking in QDA software and outside.

> Z_FN: I decided to approach the ladies at like 12:30pm to ask if they knew if the center was opened or when it would open. I approached them and in Spanish asked, "Disculpen, Saben si el centro esta abierto" – "Excuse me, Do you know if the center is open" they then continued to respond in Spanish. They said that it was open but that the staff was in

***DATA AGGREGATED IN ATLAS.ti,***

Screenshot of data indexed using dictionary, ml, and in-depth readings. From Abramson and Dohan 2015, data drawn from 12,000 pages of hand coded field-notes and interview transcripts collected as part of the Patient Deliberation Study. We often use QDA software to streamline interpretive analyses, curate data for machine learning, and help structure output into a machine readable format for computational analysis. (some initial tutorials [here](), currently updating public pages).

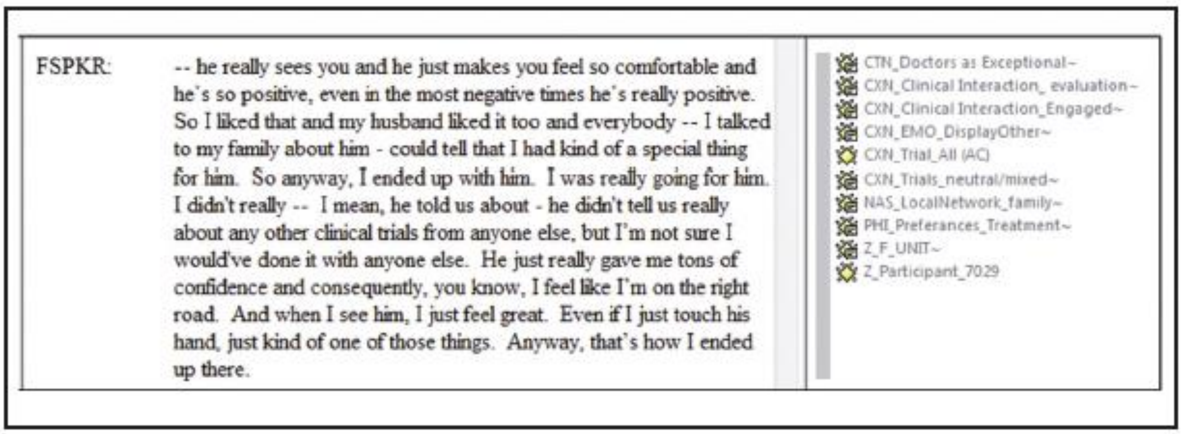


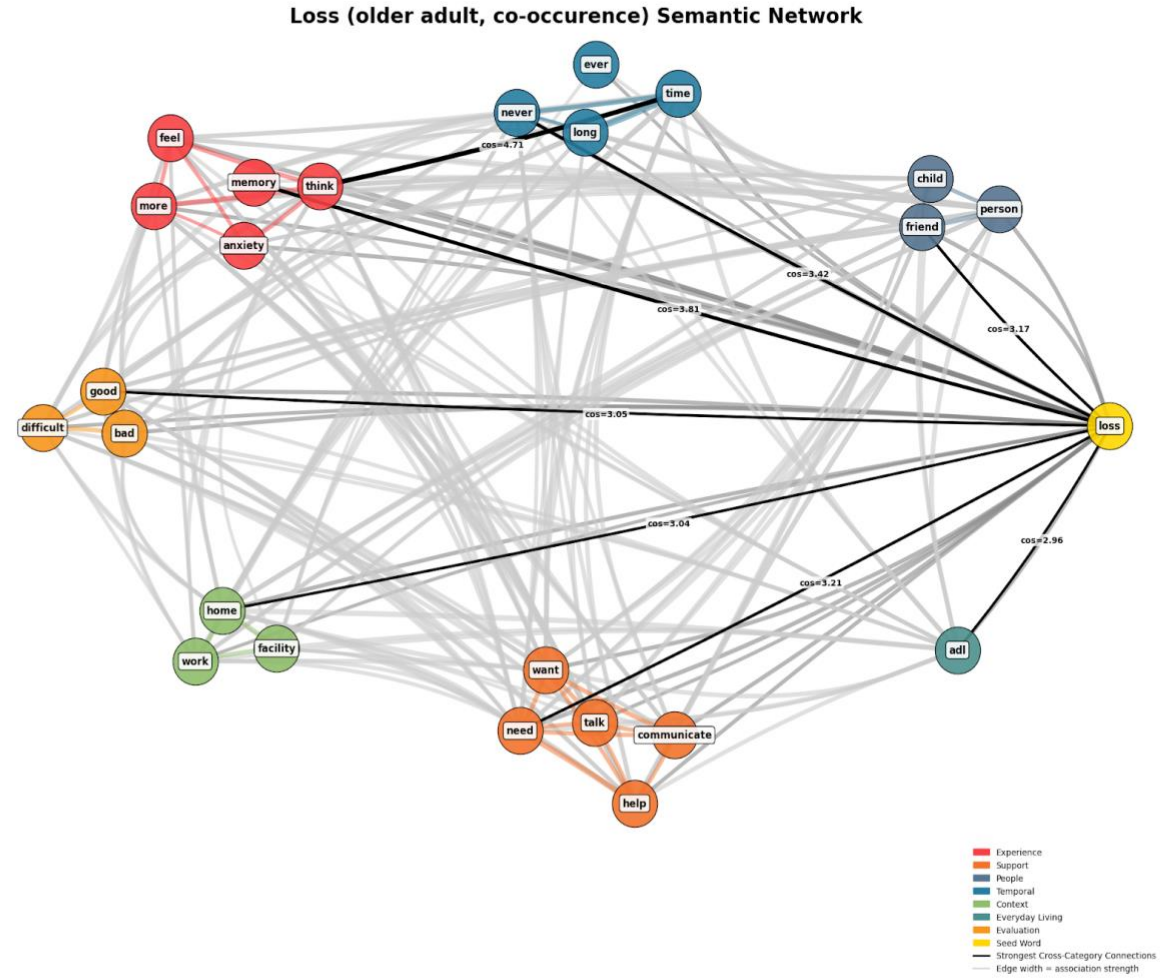

**Figure 2.** A single paragraph from the transcript of a Cancer Patient Deliberation Study patient interview, coded as an ATLAS.ti data set.

**DATA VISUALIZATION**

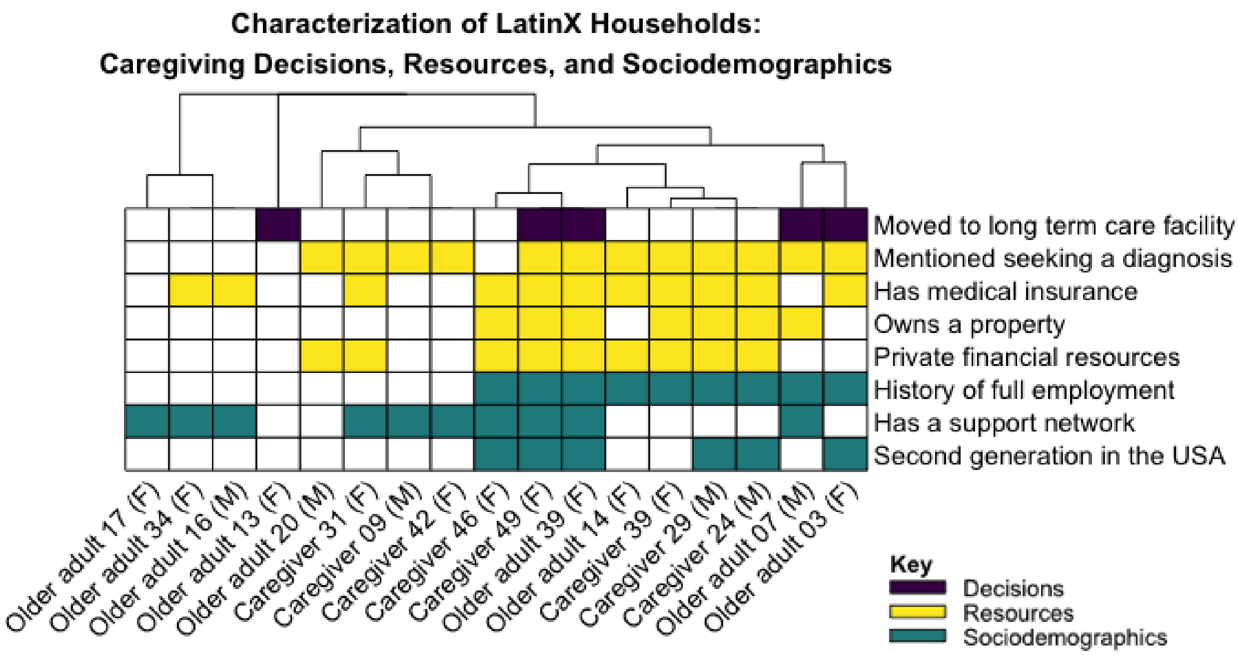
Below, execution block and output from an integrated development environment with Python; code version to be shared with article. These visuals use paragraph, n-word-window text segments, paragraphs, full documents, on whole cases (all data for an individual) depending on the analysis. Visualizations are always used alongside in-depth quotations, to show various levels of analysis, validate typologies, support claims, or provide evidence that expected outcomes did or did not manifest (Abramson et al. 2018; Abramson et al. 2024). The tools can be used for other approaches, but our use is iterative rather than purely inductive or deductive.

```python
if __name__ == "__main__":
    from pydantic import ValidationError
    import numpy as np
    import random

    # ======================== CONFIG ========================
    # ---- paths & stop-list ---------------------------------
    csv_path              = "data/1_cleaned_deidentified.csv"
    stop_list_path        = "input/stopwords.txt"
    use_custom_stoplist   = True

    # ----- core analysis ------------------------------------
    clustering_method     = 2           # 1=RoBERTa, 2=Jaccard, 3=PMI, 4=TF-IDF
    distance_metric       = "cosine"    # "default" | "cosine"
    window_size           = 15          # for co-occurence window size is the number of words before and after the target word that are included in the analysis
    num_words             = 25          # number of words to include in the analysis
    min_word_frequency    = 2           # does not show words with less than 2 occurences in the corpus
    reuse_clusterings     = False       # <-- extra attr

    # ----- preprocessing filters ----------------------------
    cross_pos_normalize   = True
    projects              = ["discern"]        # None for all
    data_groups           = ["interview"]      # None for all
    codes                 = ["older_adult"]      # None for all

    # ----- visualisation ------------------------------------
    layout                = "kamada_kawai"    # spring | kamada_kawai | circular | random | shell  <-- extra attr
    title                 = "Loss (older adult, co-occurence)"
    link_threshold        = 0.40        # <-- extra attr
    link_color_threshold  = 0.60        # <-- extra attr
    custom_colors         = True        # <-- extra attr

    # ----- seeds & colours ----------------------------------
    seed_words = {
        "loss": ["absence", "absent", "broken", "break", "breaks", "breaking",
```

[STANDARD OUTPUTS: TAG AND WORD BASED HEATMAPS, SEMANTIC NETWORKS (embedding or co-occurrence), DIMENSIONAL REDUCTION, SIMPLE VALIDATIONS AND LINGUISTIC STATISTICS; Integrating meta-data as filters]




From: Abramson, Corey M., Turner, K., Arteaga, I., Hernández de Jesús, A., Ginn, B., Nian, Y., Dohan, D. "Pragmatic Sensemaking: What Life With Dementia Reveals About Culture and Action." American Sociological Association Annual Meeting — Computational Approaches to Culture and Cognition (sponsored by the Section on Social Psychology). Tenth International Workshop on Social Network Analysis (Networks in culture, culture in networks). Naples, Italy. October 30-31. // Shows concept usage in discussions of loss, among older adults living with dementia. The larger paper examines practical challenges, the cultural schema people use to make sense of them, and their trajectories over time using longitudinal field methods and in-depth interviews. All networks, are paired with quotations and events to link general patterns and specific manifestations in the data. The code for the networks (runnable as co-occurrence, semantic similarity (with word embeddings), and PMI, and will be available in the code toolkit linked below.




Heatmap Case Visualization of Ethnographic Respondents (From Arteaga et al. 2025, in R, pre-print version) [code_link]

*OTHER RESOURCES*

**Open Source Tools: // in progress, expected by October with both python environment and collab versions [visualization toolkit]**

**Resource Page: // in progress, expected by October [github repo]**

**Tool list: [github repo] // in progress, expected by October**

*REFERENCES for APPENDIX*


Abramson, Corey, and Zhuofan Li. 2024. “Ethnography and Machine Learning: Synergies and New Directions.”

Abramson, Corey M., and Daniel Dohan. 2015. “Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data.” *Sociological Methodology* 45(1):272–319. doi:10.1177/0081175015578740.

Abramson, Corey M., Jacqueline Joslyn, Katharine A. Rendle, Sarah B. Garrett, and Daniel Dohan. 2018. “The Promises of Computational Ethnography: Improving Transparency, Replicability, and Validity for Realist Approaches to Ethnographic Analysis." *Ethnography* 19(2):254–84. doi:10.1177/1466138117725340.

Alqazlan, Lama, Zheng Fang, Michael Castelle, and Rob Procter. 2025. "A Novel, Human-in-the-Loop Computational Grounded Theory Framework for Big Social Data." *Big Data & Society* 8(2). doi:10.1177/20539517211039486.

Amirova, A., T. Fteropoulli, N. Ahmed, M. R. Cowie, and J. Z. Leibo. 2024. "Framework-Based Qualitative Analysis of Free Responses of Large Language Models: Algorithmic Fidelity." *PLOS ONE* 19(2). doi:10.1371/journal.pone.0300024.

Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. "Out of One, Many: Using Language Models to Simulate Human Samples." *Political Analysis* 31(3):337–51. doi:10.1017/pan.2023.2.

Arteaga, I., A. Hernández de Jesús, B. Ginn, Corey M. Abramson, and Daniel Dohan. "Understanding How Social Context Shapes Decisions to Seek Institutional Care: A Qualitative Study of Latinx Families Navigating Progressive Cognitive Decline." *The Gerontologist*.

Ashwin, Julian, Aditya Chhabra, and Vijayendra Rao. 2023. "Using Large Language Models for Qualitative Analysis Can Introduce Serious Bias." *arXiv*. doi:10.48550/arXiv.2309.17147.

Basov, Nikita, Wouter De Nooy, and Aleksandra Nenko. 2021. "Local Meaning Structures: Mixed-Method Sociosemantic Network Analysis." *American Journal of Cultural Sociology* 9(3):376–417. doi:10.1057/s41290-019-00084-9.

Bearman, Peter S., and Katherine Stovel. 2000. "Becoming a Nazi: A Model for Narrative Networks." *Poetics* 27(2–3):69–90. doi:10.1016/S0304-422X(99)00022-4.

Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson. 2024. "Synthetic Replacements for Human Survey Data? The Perils of Large Language Models." *Political Analysis* 32(4):401–16. doi:10.1017/pan.2024.5.

Boelaert, Julien, Samuel Coavoux, Étienne Ollion, Ivaylo Petev, and Patrick Präg. 2025. "Machine Bias. How Do Generative Language Models Answer Opinion Polls? <sup/>." *Sociological Methods & Research* 54(3):1156–96. doi:10.1177/00491241251330582.

Boutyline, Andrei, and Alina Arseniev-Koehler. 2025. "Meaning in Hyperspace: Word Embeddings as Tools for Cultural Measurement."

Breiger, Ronald L. 2000. "A Tool Kit for Practice Theory." *Poetics* 27(2):91–115. doi:10.1016/S0304-422X(99)00026-1.

Carley, Kathleen. 1994. "Extracting Culture through Textual Analysis." *Poetics* 22(4):291–312. doi:10.1016/0304-422X(94)90011-6.

Chopra, Felix, and Ingar Haaland. 2023. "Conducting Qualitative Interviews with AI." *SSRN*. doi:10.2139/ssrn.4646016.

Deterding, Nicole M., and Mary C. Waters. 2021. "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach." *Sociological Methods & Research* 50(2):708–39. doi:10.1177/0049124118799377.

Dohan, Daniel, and Martín Sánchez-Jankowski. 1998. "Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations." *Annual Review of Sociology* 24(1):477–98. doi:10.1146/annurev.soc.24.1.477.

Farber, Orly N., Corey M. Abramson, and Amanda J. Reich. 2025. "Utilizing Artificial Intelligence to Facilitate Qualitative Surgical Research." *Annals of Surgery Open* 6(2):e577. doi:10.1097/AS9.0000000000000577.

Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology* 43(1):147–65. doi:10.1146/annurev-soc-060116-053450.

Hanson, Kenneth R., and Nicholas Theis. 2024. "Networked Participants, Networked Meanings: Using Networks to Visualize Ethnographic Data." *Sociological Methodology* 54(1):142–56. doi:10.1177/00811750231195338.

Ji, Jiaming, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Lukas Vierling, Donghai Hong, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Juntao Dai, Xuehai Pan, Kwan Yee Ng, Aidan O'Gara, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. 2023. "AI Alignment: A Comprehensive Survey."

Johnson, David R., Christopher P. Scheitle, and Elaine Howard Ecklund. 2019. "Beyond the In-Person Interview? How Interview Quality Varies Across In-Person, Telephone, and Skype Interviews." *Social Science Computer Review* 39(6):1142–58. doi:10.1177/0894439319893612.

Kapania, Shivani, William Agnew, Motahhare Eslami, Hoda Heidari, and Sarah E. Fox. 2024. "Simulacrum of Stories: Examining Large Language Models as Qualitative Research Participants."

Kozlowski, Austin C., and James Evans. 2025. "Simulating Subjects: The Promise and Peril of Artificial Intelligence Stand-Ins for Social Agents and Interactions." *Sociological Methods & Research* 54(3):1017–73. doi:10.1177/00491241251337316.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. doi:10.1177/0003122419877135.

Kuipers, Benjamin. 1986. "Qualitative Simulation." *Artificial Intelligence* 29(3):289–338. doi:10.1016/0004-3702(86)90073-1.

Li, Zhuofan, and Corey M. Abramson. 2023. "Ethnography and Machine Learning: Synergies and New Directions." Pp. 245–72 in.

Li, Zhuofan, Daniel Dohan, and Corey M. Abramson. 2021. "Qualitative Coding in the Computational Era: A Hybrid Approach to Improve Reliability and Reduce Effort for Coding Ethnographic Interviews." *Socius: Sociological Research for a Dynamic World* 7. doi:10.1177/23780231211062345.

Lyman, Alex, Bryce Hepner, Lisa P. Argyle, Ethan C. Busby, Joshua R. Gubler, and David Wingate. 2025. "Balancing Large Language Model Alignment and Algorithmic Fidelity in Social Science Research." *Sociological Methods & Research* 54(3):1110–55. doi:10.1177/00491241251342008.

Mohr, John W. 1994. "Soldiers, Mothers, Tramps and Others: Discourse Roles in the 1907 New York City Charity Directory." *Poetics* 22(4):327–57. doi:10.1016/0304-422X(94)90013-2.

Murphy, Fiona, Michael Bar-Sinai, and Maryann E. Martone. 2021. "A Tool for Assessing Alignment of Biomedical Data Repositories with Open, FAIR, Citation and Trustworthy Principles." *PLOS ONE* 16(7):e0253538. doi:10.1371/journal.pone.0253538.

Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3–42. doi:10.1177/0049124117729703.
Ng, Lynnette Hui Xian, and Kathleen M. Carley. 2025. "A Global Comparison of Social Media Bot and Human Characteristics." *Scientific Reports* 15(1):10973. doi:10.1038/s41598-025-96372-1.
Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
Parker, Jessica, Veronica Richard, and Kimberly Becker. 2023. "Guidelines for the Integration of Large Language Models in Developing and Refining Interview Protocols." *The Qualitative Report* 28(9):2643–63. doi:10.46743/2160-3715/2023.6393.
Pasquale, Frank. 2016. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
Radivojevic, Kristina, Nicholas Clark, and Paul Brenner. 2024. "LLMs Among Us: Generative AI Participating in Digital Discourse."
Reyes, Victoria, Elizabeth Bogumil, and Levin Elias Welch. 2024. "The Living Codebook: Documenting the Process of Qualitative Data Analysis." *Sociological Methods & Research* 53(1):89–120. doi:10.1177/0049124120986185.
Serbu, Jared. 2024. "NARA Sees 'Unacceptable Risks' from Using ChatGPT for Agency Business." *Federal News Network*.
Singer, Natasha. 2025. "Welcome to Campus. Here's Your ChatGPT." *The New York Times*, June 7.
Stuhler, Oscar, Cat Dang Ton, and Etienne Ollion. 2025. "From Codebooks to Promptbooks: Extracting Information from Text with Generative Large Language Models." *Sociological Methods & Research* 54(3):794–848. doi:10.1177/00491241251336794.
Than, Nga, Leanne Fan, Tina Law, Laura K. Nelson, and Leslie McCall. 2025. "Updating 'The Future of Coding': Qualitative Coding with Generative Large Language Models." *Sociological Methods & Research* 54(3):849–88. doi:10.1177/00491241251339188.
Wang, Angelina, Jamie Morgenstern, and John P. Dickerson. 2024. "Large Language Models That Replace Human Participants Can Harmfully Misportray and Flatten Identity Groups."
Yoon, Hesu, and Andrew McCumber. 2024. "A Symbolic Hierarchy of Places: Global Inequalities in Tourism Narratives of the New York Times Travel Section." *Poetics* 102:101848. doi:10.1016/j.poetic.2023.101848.
Zhang, Simone, Janet Xu, and Aj Alvero. 2025. "Generative AI Meets Open-Ended Survey Responses: Research Participant Use of AI and Homogenization." *Sociological Methods & Research* 54(3):1197–1242. doi:10.1177/00491241251327130.
Zhou, Zhi-Hua. 2025. *Ensemble Methods: Foundations and Algorithms*. CRC Press.