Machine Learning and the Analysis of Culture

Sophie Mützel, Étienne Ollion

Machine Learning and the Analysis of Culture

Sophie Mützel and Étienne Ollion

February 26, 2024

Submitted version

This is a draft of a chapter to appear in the forthcoming book *The Oxford Handbook of the Sociology of Machine Learning,* edited by Christian Borch and Juan Pablo Pardo-Guerra.

Abstract

The focus of this chapter is on how machine learning (ML) impacts the analysis of culture in sociology. It shows how ML has greatly advanced the analysis of culture, with new tools enabling a massive and fine-grained extraction of information from textual and audiovisual troves as well as data analysis, operationalizing long-standing cultural sociology concepts. It also indicates that this renewed interest is building on already fertile ground, as sociologists of culture have long used and reflected on formal models when analyzing culture. The chapter suggests that as the toolbox of ML approaches expands, so will the need for methodological reflection on the datasets and algorithms used, analyzed, and interpreted. The chapter also suggests that ML techniques can serve as catalysts to generate new theoretical insights. The chapter's conclusion discusses the potential of ML research to generate new theoretical insights abductively and advocates for methodological reflexivity.

Keywords: culture, machine learning, topic modeling, word embeddings, large language models (LLMs), unsupervised and supervised models, frames, schema, text, sound, images, theory

Authors:

Sophie Mützel is Professor of Sociology at the Department of Sociology, University of Lucerne, Switzerland.

Étienne Ollion is Professor of Sociology at l'Ecole Polytechnique in Paris, France.

1. Introduction

Sociologists have a rich history of exploring culture, understood both as a process of meaning-making and as a set of social practices. This chapter presents an overview of how current applications of machine learning (ML) are impacting both the analysis of culture and the field of cultural sociology. It shows how new tools offer new ways to operationalize existing sociological concepts in the study of culture. Despite all of the innovations, the chapter also indicates that these new tools fit into a long tradition of measuring culture in the human and social sciences. From content analysis to various forms of relational approaches to studying culture, research has used methods of measurement and models of formalization before turning to interpretation.

The chapter suggests that as the toolbox of ML approaches expands, so will the need for methodological reflection on the datasets and algorithms used, analyzed, and interpreted. The chapter also suggests that ML techniques can serve as catalysts to generate new theoretical insights. Applications of ML in the study of culture have shown that deductive approaches of hypothesis testing with a limited set of variables do not hold. Instead, research has pointed to unsupervised ML as an inductive data mining strategy to "'discover' unnoticed, surprising regularities" (Evans & Aceves, 2016), which then need interpretation using qualitative insights and deep knowledge of the empirical case to turn them into theoretical sociological insights.

Current research indicates a tacking back-and-forth between computational and qualitative analysis to develop interpretation and new theoretical insights. Studies have moved between computational analyses and hermeneutically grounded "thick reading" as "computational hermeneutics" (Mohr, Wagner-Pacifici, & Breiger, 2015), between "atheoretical induction and theory-led deduction" as "forensic social sciences" (Goldberg, 2015; McFarland, Lewis, & Goldberg, 2016, p. 21), or have combined "inductive grounded theory with deductive quantitative tests" to detect, refine, and confirm patterns as "computational grounded theory" (Nelson, 2020). Current research in cultural sociology using ML thus follow abductive logics, when using nuances of large datasets, the tools of machine learning, and tacking back-and-forth between computational analysis and qualitative interpretation "to locate surprising empirical findings" and to discover "innovative and creative theoretical insights" (Brandt & Timmermans, 2021, p. 192).

The chapter begins with a brief history of the measurement of culture in sociology through the lens of relational approaches, which have been central to this endeavor. It then investigates the culture of data, looking both at how culture is embedded in data and how researchers need to build a culture of data. The following section explains how ML has greatly advanced the analysis of culture in corpora, with new tools enabling a massive and fine-grained extraction of information from textual and audiovisual troves. ML algorithms are useful not only for data extraction but also for data analysis; the final section investigates how recent and older tools are currently being used to operationalize long-standing cultural sociology concepts. The conclusion discusses the potential of ML research to generate new theoretical insights abductively and advocates for methodological reflexivity.

2. A relational approach to measuring culture

Sociologists have long been interested in using texts to study social phenomena. Content analysis has been used to systematically find and measure specific constructs of interest, based on dictionaries,

indices, and coding schemes (Krippendorff, 2012). Similarly important has been the early use of lexicometric analyses, which measure the frequency of words in a given corpus, to describe and quantify "the manifest content of communication" (Berelson, 1952, p. 18). Meanwhile, cultural sociologists have long been interested in identifying *latent* patterns in how individuals, groups, or organizations make meaning. Their research has focused on cultural practices and dynamics using a variety of data types, including surveys, newspaper reports, directories, observation, and interviews, as well as musical scores, flags, and recipes.

One strand of such a cultural sociology has sought to measure culture (e.g., Jepperson & Swidler, 1994; Mohr, 1998; Mohr et al., 2020; Rawlings & Childress, 2021). For measuring culture, relational theories (e.g., White, 1992) paved the way because cultural elements do not exist in isolation. Rather, cultural elements and their manifestations in practices and discourses exist in specific contexts and in relation to each other.

Given this theoretical focus, methods to study relations have shaped the quest to measure culture. Several overviews highlight the fundamental role of network analysis for the analysis of culture (e.g., DiMaggio, 2011; Fuhse & Mische, 2024; Pachucki & Breiger, 2010). One way to measure culture is to use formal network techniques to analyze relations within and between cultural artifacts, e.g., texts or other kinds of cultural forms. The central idea here is that cultural forms are relationally composed and that these relations can be systematically mapped and measured in ways that reveal both the organization of those cultural elements and their embedding. Research typically breaks down cultural forms in texts into distinct observable components, such as concepts, categories, practices, narratives, events, and genres, and then employs formal techniques to map their relationships to one another or to other types of entities (Mohr et al., 2020).[1]

Characteristic for the empirical analysis of culture is the search for pattern, part of a larger sociological embrace of descriptive modes (Savage, 2009). Network analysis and other relational approaches to formally measure meaning have been foundational in this search for patterns. Recently, other types of algorithms and methods have contributed to the formal analysis of culture.

3. The Culture of Data

Yet, before turning to how ML algorithms are used to analyze culture, it is necessary to reflect on where the data to be analyzed come from, how they are produced, and how datasets are curated and constructed. This section suggests that cultural sociology working with ML tools requires a reflection on the culture embedded in the data (culture in data). At the same time, it also needs to develop a culture of data, to anticipate where biases and errors may exist in data and algorithms.

[1] Examples include the analysis of identities in narrative networks and semantic triplets (e.g., Bearman & Stovel, 2000; Franzosi, 1997), mental models in concept networks (e.g., Carley, 1994), and discourse and practices in Galois lattices (e.g., Mische & Pattison, 2000; Mohr & Duquenne, 1997). Similarly, in his classic work, Bourdieu uses taste as the object of formal analysis, surveys, and ethnographies as data sources, and applies the relational method of multiple correspondence analysis to show cultural distinctions in French society as objective relations (1984). Other studies have measured discursive practices in spoken, written, or digital texts that signal and performatively constitute relations with *other* actors. Here, communicative interactions serve as relational units that can be observed, aggregated, compared, and analyzed using network measures (e.g., Mische, 2008).

Most of the contemporary ML algorithms are trained and tested on a limited number of benchmark datasets (Koch, Denton, Hanna, & Foster, 2021). For the development of the field of ML, the use of these datasets was instrumental since they allowed researchers to immediately evaluate and compare the performance of each algorithm. However, these benchmark datasets also come with problems: they are appropriated for different tasks than originally intended, and, because of their context of origin, they contain certain biases. A now-classic paper (Buolamwini & Gebru, 2018) demonstrates that two commonly used facial detection algorithms are built on "pale male" datasets, i.e., they disproportionally contain photographs of light-skinned men. In turn, these algorithms perform best on white men and systematically misrecognize women and men of color. Similar problems due to lack of diversity, reappropriation, and existing historical cultural and social biases in standard benchmark datasets also apply to textual data and, accordingly, affect results of language models by reproducing and potentially even amplifying biases and social prejudices in their results.

Indeed, language models are biased (e.g., Caliskan, Bryson, & Narayanan, 2017). This holds for word embedding models pertaining to gender and ethnic biases (e.g., Basta, Costa-Jussà, & Casas, 2019) but also to biases regarding socioeconomic status, age, physical appearance, sexual orientation, religious sentiment, and political leanings (Rozado, 2020). Also, large language models (LLMs) exhibit similar patterns of encoded bias since stereotypes (Nangia, Vania, Bhalerao, & Bowman, 2020), discriminatory language, and derogatory associations along gender, race, ethnicity, and disability status are encoded in training datasets (Bender, Gebru, McMillan-Major, & Mitchell, 2021).

The presence of bias in these models has drastic consequences when deployed in real-world analyses and applications. Facial recognition systems, trained on "pale male" datasets, have been shown to contribute to the propagation of errors, prejudices, discrimination, and exclusion, e.g., when used in systems of crime prediction and policing (Noble, 2018) or in hiring and job evaluation (O'Neil, 2016). In order to increase transparency and accountability and to mitigate unwanted yet encoded societal biases, stereotypical associations, and negative sentiment towards specific groups, researchers in academia and industry have emphasized the need to document the characteristics of datasets used in training and analyses (e.g., Mitchell et al., 2019; Whittaker et al., 2018). They suggest accompanying each dataset with information on its motivation, composition, collection process, and recommended uses (e.g., Bender & Friedman, 2018; Gebru et al., 2021).[2]

While a consequential problem in commercial applications, the existence of bias in data and models can turn out to be an empirical boon for social scientists. The presence of bias in these models has turned them into a tool for large-scale analysis of social biases, stereotypes, and attitudes by investigating cultural understandings embedded in the data. Studies using historical data show how gender stereotypes and attitudes toward ethnic minorities have evolved (Garg, Schiebinger, Jurafsky, & Zou, 2018) and how some stereotypes about social class have remained stable (Kozlowski, Taddy, & Evans, 2019). Nelson (2021) shows how bias can be used analytically in a study of first-person narratives from the US South before 1920. Combining inductive and ML approaches in one research design, Nelson uses biased cultural understandings of the dataset to highlight how different biases intersect in the narratives. Luo, Gligorić, & Jurafsky (2023) use supervised ML methods to assess stereotypes associated with each cuisine in the United States based on restaurant-rating website

[2] Studies also point out that an analysis of biases needs to extend beyond datasets to include all participants in their production and the choices that contribute to the development of algorithmic procedures (e.g., Jaton, 2017 for a case study).

reviews. They show that positive reviews, as well as terms such as authenticity and cleanliness, are disproportionately attributed to restaurants that serve food from old immigration countries rather than those that serve food from more recently arrived migrant groups.

Developing a Culture of Data

In addition to understanding the potential biases encoded in datasets used for training and further analyses, sociologists also need to develop a culture *of* data. This entails considering bias and errors at all stages of data production and curation, thus assessing all potential sources of distortion between observation and analysis. Sociological data literacy in ML based research includes becoming familiar with the benefits and drawbacks, with the production and curation of each dataset, as well as with the possible re-use of data.

Becoming familiar includes the development of standards and guidelines, similar to those established protocols used in other methods, i.e., surveys, ethnography, or interviews. The field has begun to systematically develop such guidelines and standards for ML research methods in general (Kapoor et al., 2023) and, separately, for data quality. Hurtado Bodell, Magnusson, & Mützel (2022) propose a framework for assessing total corpus quality that identifies all stages—study design, data collection, processing, and, finally, analysis—at which potential errors in working with a digitized dataset can occur. They use digitized Swedish newspapers to demonstrate the framework yet underscore its application to other projects that use sound, images, or digital data. Others similarly develop a framework for digital trace data (Sen et al., 2021).

When using existing datasets or pre-trained algorithms, data sources and their curation are not always transparent. A growing body of research demonstrates that people who clean and test the data used to train widely used ML applications work under stressful conditions and for low pay. Like platform moderators (Gillespie, 2018, 2020; Roberts, 2019), ML annotators are typically outsourced, poorly paid gig workers (Tubaro, Casilli, & Coville, 2020), often living in the Global South. Following instructions that describe in graphic detail what constitutes the most vile or toxic content of the internet, their job is to refine training data, models, and their outputs at high speeds (GPAI, 2023; Perrigo, 2023). Moreover, inquiries into encoded cultural and language-based biases of trained LLMs used in publicly used generative AI models find systematic cultural disparities in outputs and censored answers (e.g., Ghosh & Caliskan, 2023; Urman & Makhortykh, 2023). These practices indicate the solidification of a culture of global inequality as part of a larger culture of data.

Other developments include gig workers' use of publicly available generative AI models instead of human decision-making only to train or test AI models themselves. While this may foreshadow further use of synthetic data in research (Bail, 2023), it also requires additional inquiries into the culture of data for datasets used in commercial applications and research.

4. Extracting Culture from Corpora

One way to use ML algorithms for sociological analysis is to extract information from data, structure the resulting data, and reduce their complexity. This strategy, also known as feature extraction, builds on earlier efforts to reduce dimensionality when dealing with textual data. However, unlike previous approaches—such as manually assigning codes as in traditional content analysis or using mathematical, relational models to find and extract patterns in small datasets—current large-scale corpora contain both nuanced and messy information.

Due to the digital transformation of daily life, people are leaving an increasing number of traces that can be used by researchers (e.g., Bail, 2014; Edelmann, Wolff, Montagne, & Bail, 2020; Salganik, 2017). Additionally, digitization initiatives in archives around the world produce digitized datasets of old data for further analysis (Bearman, 2015). Such large-scale digital and digitized text datasets, both new and old, have challenged the limitations of earlier research techniques, such as labor-intensive manual coding in content analysis or using dictionaries or close reading of words and numbers. Instead, we are witnessing the advent of the "golden age of textual analysis" (e.g., Grimmer, Roberts, & Stewart, 2022; Ignatow & Mihalcea, 2016; Mohr, 1998, p. 366).

Currently, ML algorithms assist in converting massive amounts of data into numerical features, which can then be processed and further analyzed. Such ML techniques promise researchers "to do more with fewer resources" (Nelson, Burk, Knudsen, & McCall, 2021, p. 203) and to extract even more information than previously available. This promise holds great appeal for cultural sociologists because it can help operationalize important sociological terms like "frames," "stories," "narratives," "concepts," and "schemas," which, in turn, can create new opportunities for theoretically driven and empirically grounded research (Bonikowski & Nelson, 2022).

Finding patterns in texts with topic models

One of the methods cultural sociology has often used in the past decade is the unsupervised ML approach of topic modeling (based on Latent Dirichlet Allocation, LDA; e.g., Blei, 2012; Blei & Lafferty, 2009; Blei, Ng, & Jordan, 2003). Topic modeling yields groups of words that frequently co-occur together and collectively represent themes or latent "topics." This is based on a statistical model of language, i.e., on a probabilistic model of how words are distributed in a corpus and does not require *a priori* dictionaries or interpretive guidelines. The amount of human involvement in topic modeling algorithms is minimal: researchers must tell the algorithm how many topics to find in the corpus; the algorithm then produces that number of topics, the words that make up each topic, and the distribution of those topics across the entire corpus.[3]

Highlighted in a special issue of *Poetics* (Mohr & Bogdanov, 2013), topic models have become a preferred and powerful tool for many cultural sociologists to analyze large unclassified textual data sources for explorative analyses. There is an elective affinity between topic models and sociology since extracted topics have an intuitive resemblance to what other sociological inquiries yield as themes: they are relevant and interpretable as units of meaning. Indeed, topic modeling has been likened to grounded theory approaches (Baumer et al., 2017), which identify themes based on close readings. DiMaggio et al.'s (2013) insightful discussion highlights three additional significant advantages of topic models for sociologists of culture (578-582): (1) Most topics obtained from topic model solutions are substantively interpretable and can be read as "frames" in the sense that topics are semantic contexts, capturing a relationality of meaning; (2) topic model solutions are able to capture the polysemy of terms, as in "not all banks sit in parks"; and (3) topic model solutions are able to capture heteroglossia in texts, i.e., the copresence of different voices, styles, and perspectives because each document consists of multiple topics.

Substantive applications of topic modeling in the study of culture are manifold. They include, amongst others, studies on funding in the arts (DiMaggio et al., 2013); valuation of restaurant

[3] To be sure, topic modeling is a statistically rooted clustering technique and, thus, only arguably, a machine learning method. We include it here because topic models have been instrumental in the transformation of cultural sociology to analyze large unclassified textual data before moving towards further methods rooted in ML models.

experiences (Mützel, 2015b) or music (Light & Odden, 2017); decision-making during financial crises (Fligstein, Brundage, & Schultz, 2017); economic thought (Erikson, 2021); socio-political conflicts (Karell & Freedman, 2019, 2020); sociological knowledge production (Heiberger, Munoz-Najar Galvez, & McFarland, 2021); scientific innovations (Mützel, 2022); attitudes in health care (Miner et al., 2023); climate talk negotiations (Gray & Cointet, 2023).

The unsupervised method of topic modeling follows an inductive logic; it detects patterns in texts for explorative analyses. It does not include a measurement of frames based on theoretical expectations (Nelson et al., 2021). Rather, identified topics present "the lens through which one can see the data more clearly" (DiMaggio et al. 2013, p. 582), while contextual knowledge is needed to interpret the patterns. This inductive discovery of themes, however, can also be used to generate theory based on "anomalous and surprising empirical findings against a background of multiple existing sociological theories" (Timmermans & Tavory, 2012, p. 169). Rather than purely inductively, topic modeling can be used for "computational abductive analysis" (Karell & Freedman, 2019). In the study of rhetorics of radicalism, Karell and Freedman's topic model analysis yields results that fit existing assumptions, yet results also suggest new, contrasting concepts, thus generating new theoretical insights.[4]

After initial excitement about topic models to get information, find patterns, and figure out what those patterns mean, sociologists have identified several problems with such unsupervised approaches over the last decade. In general, unsupervised learning tools are wholly data-driven. This is beneficial for exploring a corpus, yet these tools cannot identify a priori specified theoretical concepts. Another limitation of topic modeling is that it does not grant researchers good control over what to extract; it remains a "black box" with occasionally cryptic output (Lee & Martin, 2015). Since researchers have little control over the modeling itself statistical validity is limited (Grimmer & Stewart, 2013).[5] Moreover, certain preprocessing and parametric modeling choices can impact topic modeling results (e.g., Denny & Spirling, 2018). Comparing hand-coded, dictionary, off-the-shelf supervised ML, and unsupervised topic modeling tools to identify the concept of inequality in a corpus of articles, Nelson et al. (2021) find that topic models can complement traditional approaches of coding complex and multifaceted sociological concepts. However, they cannot fully replace traditional, labor-intensive approaches.[6] And, last but not least, topic models rely on an unrealistic assumption about language since they treat each document as a "bag of words," thereby disregarding syntax, grammar, or order of words within the document (e.g., Shadrova, 2021). The combination of widespread interest in topic models and their limitations paved the way for the adoption of other ML approaches in the analysis of culture.

Renewals of text classification with LLMs

A prominent approach to extracting culture from texts is to use Large Language Models (LLMs), a supervised ML approach that first appeared in the late 2010s and that rejuvenated work on text classification. LLMs are models that have been trained on a large number of texts.[7] Through this training process, the models learn a representation of language. Initial language models (Brown et al., 2020; Devlin, Chang, Lee, & Toutanova, 2019; Liu et al., 2019) are built from *Transformers*

[4] Interestingly, models based on Bayesian inference, which builds on empirical updating, can be regarded as fundamentally similar in their logic to the logic of abductive inference (Ignatow, 2020).
[5] Research has argued to focus instead on semantic/internal and predictive/external validity (DiMaggio et al., 2013).
[6] However, more recent unsupervised and supervised models might replace human annotation altogether.
[7] Training data are based on all publicly available information online, mostly from common crawl.

(Vaswani et al., 2017), a neural network architecture capable of producing a better representation of language than previous models, and consequentially improving the accuracy of text annotation.

Examples of early supervised LLMs include BERT models (Bidirectional Encoder Representations from Transformers) in all their variations and languages. LLMs are publicly available and come pre-trained for a variety of tasks, including text classification and sequence labeling.[8] Fine-tuned with a subset of manually annotated text data, i.e., by providing a variety of examples of the pattern to be recognized or by giving instructions on how to carry out a task, such LLMs can classify texts all the way below the sentence level. Researchers can thus train and "fine-tune" LLMs for their research interests. For text classification tasks, LLMs have significantly outperformed all previous records on traditional benchmarks.

Used in a supervised way to classify texts, the promise of these LLMs is threefold: (1) theoretically, they can capture almost any pattern in a dataset; (2) they can do so with a high degree of precision; and (3) they can do that with minimal human training (Peters et al., 2018). Experimental evidence shows that LLMs can keep these promises, thus serving as valuable annotation tools. Do, Ollion, & Shen (2022) trained a series of classifiers, each attempting to recognize a specific text pattern in a news media corpus. One of the classifiers tried to identify expressions that introduce unattributed sources in journalistic writing (e.g., "according to a source," "a person who wishes to remain anonymous"). After fine-tuning the model with a small number of examples (less than 1% of the entire corpus), the classifier was able to detect these patterns in the rest of the corpus. In comparison, the classifier outperformed hired gig workers, performed similarly to trained research assistants, but it was also able to annotate the entire corpus, a task that was beyond reach for the small group of researchers and their research assistants. The trained BERT classifier did miss a few complex expressions, but unlike humans, it did not suffer from the infamous "fatigue effect," which causes human annotators to make errors due to a lack of focus (Rousson, Gasser, & Seifert, 2002).

LLMs have been applied to a wide range of issues important to cultural sociologists and beyond: e.g., detection of hate speech on social media (Davidson, 2024); identification of the rhetorical style in speeches of U.S. presidential nominees during their campaigns, 1952-2020 (Bonikowski, Luo, & Stuhler, 2022); thematic labeling of a series of texts with high accuracy for immigrant studies (Ren & Bloemraad, 2022); and the spatial variation of religiosity (Jensen et al., 2022).

While the iterative annotation process for fine-tuning a model requires a significant amount of time, labor, and computation, it has at least two benefits. It reduces the need to outsource annotation to research assistants or gig workers. Moreover, training a language model may prompt researchers to refine and improve the definitions of the concepts being classified. For example, when Bonikowski et al. (2022) were training a classifier to detect frames of populism in speeches by U.S. presidential nominees, they discovered that the LLM produced results they had not previously considered yet fit the frames. Because the LLM results improved the analysts' understanding of the concepts being classified they chose to work with the algorithm's suggestions. Following an abductive logic of inquiry, ML procedures can produce new insights, which can then be reapplied to the data for further analysis. LLMs can also be used as a tool for putting analytical categories to the test and may work as an interlocutor for conceptual refinement (Pardo-Guerra & Pahwa, 2022). In sum, in using LLMs sociologists are capable of "extending" (Lundberg, Brand, & Jeon, 2022) and even "augmenting" their expertise (Do et al., 2022).

[8] These and other language- or domain-specific language models are maintained on huggingface.co, which also provides the Transformers Python library (Wolf et al., 2020).

The same supervised methods can be used to train classifiers for both images and audio (Arnold & Tilton, 2019). Mazières, Menezes, & Roth (2021) use a computer vision classifier to identify the appearance of women and men on screen in order to measure gender discrimination in movies. First, they train the classifier to reproduce a few metrics, including the classic Bechdel test, which tracks the proportion of women's faces visible on screen (alone, with a man, etc.). After training the classifier, they apply it to over 3,700 movies, revealing previously unknown gender patterns while confirming others. The same logic can be used to analyze art. Banerjee, Cole, & Ingram (2023) use a convolutional neural network to measure the stylistic distance between specific painters and what is considered the canon at the time, as well as various technologies for summarizing image contents. Increasingly, sound can be used too, transcribing it into text either to transform long interviews or TV shows into transcripts or to study accents and linguistic variations.

Potentials and limitations of LLMs

Although these techniques necessitate complex calculations, their use is not limited to quantitative researchers or those with strong coding skills; many off-the-shelf packages are relatively simple to use. Importantly, these algorithmic tools can be used without following hypothetico-deductive or positivist logics. Indeed, these techniques produce new data or a new sorting of data, which researchers can then investigate based on their own epistemological assumptions. Furthermore, ML applications also provide practical benefits for researchers conducting qualitative interviews: The same models that analyze large text datasets have helped to improve the performance of automatic speech recognition, resulting in significantly higher-quality interview transcriptions and better content extraction for manual coding. To be sure, there are still several challenges to using LLMs for text classification and pattern recognition. For instance, LLMs are still unable to capture certain aspects of language, e.g., irony or sarcasm. They also sometimes fail on tasks that require implicit knowledge. Another issue concerns required resources and high energy use: the models are greedy, and typically need expensive graphical processing units to run.

When used in sociological research, the rise of generative pre-trained transformer (GPT) models exacerbates these challenges. Such generative models represent a shift away from pattern recognition, as in previous LLMs, and toward the *generation* of free-form text, images, and video. Indeed, the release of ChatGPT and other tools in 2022 brought about a new approach to text classification: zero or few-shot learning. Now, researchers can ask the model to extract information relevant to a given task; it will work as a virtual research assistant for coding and annotating text. Gilardi, Alizadeh & Kubli (2023) demonstrate that using ChatGPT on a set of two dozen tasks yields results of better quality than when using hired crowd workers (also Törnberg, 2023). Furthermore, proprietary generative LLMs are trained on large amounts of online, unknown data, which may result in biased and prejudiced responses. Open-source generative LLMs could provide an alternative route for transparent and accountable training, resulting in less black-boxed research (Spirling, 2023). Thus, generative LLMs have some advantages for sociological research, including improving simulation-based research, acting as research assistants for coding and annotation texts. Nevertheless, they continue to have risks and limitations for research (Bail, 2023; Weidinger et al., 2022). For example, it remains unclear how ethical and replicable research with generative, typically proprietary, and non-replicable LLMs can be conducted (Ollion, Shen, Macanovic, & Chatelain, 2024; Palmer, Smith, & Spirling, 2024). Furthermore, larger societal issues, such as the underpaid behind-the-scenes labor of human annotators and the ecological impact of massive computational and memory energy levels, remain unresolved.

5. Cultural Analysis with Machine Learning

The operationalization of fundamental ideas such as frames, schemas, mental models, and stereotypes has long been a challenge in cultural sociology. Measuring what is often implicit, unsaid, and not even conscious for the actors themselves has always been difficult. Finding a way to measure these unobservable ideas has resulted both in methodological advances but also in a diversity of approaches (e.g., Stoltz & Taylor, 2021).

Stereotypes, Frames, and Schemas in the Age of Machine Learning

For several years, topic models replaced older lexicometric or dictionary-based methods as the preferred technique for extracting culture from text, since this method and the central idea of a formal cultural sociology to find latent patterns of meaning have much in common. When the first tools from the most recent wave of ML became available, a new wave of enthusiasm swept through the field. This was especially true for word embeddings, a method developed in the 2010s (Mikolov, Chen, Corrado, & Dean, 2013) to create a vector representation of every word in a given corpus via a dimensionality reduction process. Word embeddings calculate the proximity and distance between terms and identify associations between these words and specific concepts. The resulting semantic space synthesizes relationships between words, captures those relationships, and allows for a variety of operations on words or concepts. In fact, Stoltz & Taylor (2021, p. 2) refer to word embedding models as "a means of mapping meaning space."

To study bias, frames, and stereotypes, social scientists have used these associations of ideas present in texts. Research builds on the fact that word embeddings mirror stereotypical racial, ethnic, and gender-related biases found in the texts they are trained on (e.g., Brunet, Alkalay-Houlihan, Anderson, & Zemel, 2019; Caliskan et al., 2017; Lewis & Lupyan, 2020). While this is a concern for some research, leading to attempts to "de-bias" word embeddings (Bolukbasi et al., 2016; Gonen & Goldberg, 2019), for others the prevalence of bias presents an opportunity when taking these associations as features of the social world. Rather than distortions in the semantic space, these associations reflect the contours of cultural formations over time (Nelson, 2021). Crucially, these models do not "understand" meaning, nor do these procedures replace interpretation.

The method of word embedding has affinities with important concepts in cultural sociology. Arseniev-Koehler & Foster (2022) suggest that neural word embeddings present ways to analyze schemas, i.e., abstract information and cognitive structures that are internalized across experiences with public culture and foundational for meaning-making (e.g., Leschziner & Brett, 2021; Lizardo, 2017). Word embeddings offer a "crucial step towards a formal model of schema extraction and schematic processing of text data" (Arseniev-Koehler & Foster, 2022, p. 1499) and can "be used to empirically identify the schemas activated and reinforced by public cultural data" (p. 1500). Whether with a focus on schemas, mental models, or more generically to extract "patterns," word embeddings have been used to study a variety of substantive topics, including class culture (Kozlowski et al., 2019), racial connotations in the late 19th century (Nelson, 2021), and the perception of immigrants in US etiquette books (Voyer, Kline, & Danton, 2022).

Part of what makes word embeddings appealing is their capacity to extract latent semantic dimensions, along which sociologically relevant keywords or sentences can be positioned, e.g., to estimate how far apart concepts are in political texts (Rheault & Cochrane, 2020). Yet, word embeddings also have limitations (Rodriguez & Spirling, 2022). Word embeddings' results may be inaccurate and, like other unsupervised methods, provide users with little control over what they

study, though constrained word embeddings address this criticism to some extent (Hurtado Bodell, Arvidsson, & Magnusson, 2019). Word embeddings are unable to consider negation, polysemy of terms, word order, or syntax.[9] They are also unable to determine a speaker's position on a specific topic. For example, in a given corpus, the relationship between immigration and crime can be merely mentioned but also hotly debated. However, since word embeddings are only trained to recognize co-occurrences in language, the model is unlikely to distinguish between these two positions. Supervised classifiers work better at measuring such expressed positions or "stances" (Luo, Card, & Jurafsky, 2020).

Although much of cultural sociology has focused on understanding how relationships are structured and develop, word embeddings are, by design, unable to capture the precise relationship between actors. Because of this and echoing earlier research on narrative structure and semantic triplets (Franzosi, 2004), sociological research has begun using dependency parsers (e.g., Chen & Manning, 2014), which identify the syntactic functions that different entities perform within a sentence. In his work on Germany during the 2015 migration crisis, Stuhler (2022) shows how migrants are portrayed in the media, including whether they are depicted as active or passive, whether they are mentioned as subjects with agency, and the emotions associated with each role. By modeling language structure rather than just word correlations, Stuhler's work provides a more accurate understanding of what media reports mean. Other research that conceptualizes language beyond word correlations includes examinations of narrative statements that specify "who does what to whom" (e.g., Ash, Gauthier, & Widmer, 2021; Goldenstein & Poschmann, 2019).

Rapidly Moving Frontiers of Research

At the time of writing, new language models are being released every week, and the promises surrounding the use of ML applications and AI in science and society have reached new heights. Given how quickly the ML landscape evolves, making predictions is rather difficult. Some developments, however, appear to be well under way.

Although it is widely assumed that the field of machine learning-based image recognition, or "computer vision," is more advanced than language processing, sociologists have predominantly studied text instead of images using ML tools. Yet there are many opportunities for sociological research using ML to analyze images.

Research has started to use computer vision for video data analysis to examine movements and interactions using existing video footage (Bernasco et al., 2023) as well as applying additional body keypoint software to images, thus improving the capture of individual bodies across frames (Goldstein, Legewie, & Shiffer-Sebba, 2023). Several software packages are currently available for converting videos into usable, analyzable metrics (Nassauer & Legewie, 2021) which can accurately capture people's walking, gesturing, or interaction patterns, as well as detect individual or group emotions. Computer vision, combined with machine learning-driven video analysis, can shed light on social interactions in specific contexts and cultural settings. Furthermore, given the vast amount of available visual data, insights into situated social settings may also be scaled up. For example, research could focus on issues such as bodily socialization and the role of organizations in shaping movements and social interactions, as well as gendered differences in the use of the body. Contextual

[9] Models of contextual word embeddings address these limitations, however, at the expense of increased complexity (e.g., Arora, May, Zhang, & Ré, 2020).

analyses could also be used to determine whether and how people adjust their movement patterns in response to their social environment. Without a doubt, such research again raises concerns about surveillance and necessitates rigorous ethical and regulatory considerations. Applying facial and image recognition algorithms to large databases of digitized images is an additional method of working with images (van Noord, 2022). In historical sociological studies of culture, such applications of images as data can help to explore the context and find patterns.

In addition to still and moving images, audio is an underutilized data source for sociological research using machine learning. As recordings become more widely available and their algorithmic transformation becomes more reliable and cost-effective, datasets for analysis will gain prominence. Certainly, as with facial recognition and LLMs, automated speech recognition algorithms that convert speech to text require researchers to be aware of encoded bias (Koenecke et al., 2020). With insights from linguistics, sociological research could focus on social persistence, e.g., how persistent language patterns are throughout life, or again on context, e.g., how people adjust their intonations based on their social context and interaction. Furthermore, as these forays into potential future research strands of cultural sociology demonstrate, sociological research using ML will require interdisciplinary collaboration.

7. Conclusion: Generating Theoretical Insights and Methodological Reflexivity

The application of ML algorithms has had a profound impact on cultural sociology. Sociological research is currently in the "golden age of textual analysis," which uses ML algorithms to extract data, identify patterns, and interpret the data. Simultaneously, the use of ML algorithms challenges sociology's understanding of how to work with datasets, generate new theoretical insights, and interpret results.

The chapter suggests that as the toolbox of ML approaches expands, so will the need for methodological reflection on the datasets and algorithms used, analyzed, and interpreted. Research needs to consider the production, curation, and limitation of each dataset by considering each corpus as a product of social practices and decisions (Mützel, 2015a), and thus of a certain data quality (Hurtado Bodell et al., 2022). For sociological research, the growing toolbox also contains data sources other than text, like sounds and images.

The chapter also suggests that ML techniques can serve as catalysts for the generation of novel theoretical insights. For example, ML algorithms can be used for exploratory analyses. Unsupervised ML algorithms are especially useful as an inductive data mining strategy for identifying patterns in datasets. In contrast to parametric approaches, which require the functional form to be specified beforehand, ML algorithms can find all possible relationships between variables in the data (Boelaert & Ollion, 2018).[10] Ideally, “universal approximation” could address specification issues, omitted variables, and offer parametric modeling new insights.[11]

[10] Bhatt, Goldberg, and Srivastava (2022) give an illustration of this by using the ML algorithm random forest to solve a classification problem. To uncover how social group boundaries after a merger of companies are maintained or altered, they analyze the language used in 1.5 million internal employee emails. Without prior knowledge about how such boundaries might manifest themselves concretely in email, the random forest classifier allows them to detect subtle cultural distinctions in linguistic styles.

[11] Salganik et al. (2020) demonstrate that it is possible to train different ML classifiers to function on par with the best parametric regressions—even though the latter had been developed and fine-tuned by experienced scholars based on their decade-long experience in this field, while the former had been used by researchers with limited knowledge of the empirical case.

Finding patterns inductively, however, does not suffice. Sociological analysis also requires interpretation. Such an interpretation draws on existing theories while also benefiting from new theoretical insights. Studies in cultural sociology highlight new needs for "methodological bricolage" (Bonikowski & Nelson, 2022), the complex and iterative workflows between qualitative and quantitative analyses as well as between measurement and interpretation that conform to an abductive logic of inquiry (Timmermans & Tavory, 2022). Engaging with social contexts using qualitative research methods, research is able to develop "insight[s] about what to look for in the data and how to theorize what is being observed" (Grigoropoulou & Small, 2022, p. 905). Alternatively, "results from computational workflows" can be used as "prompts shared with informants in ethnographic research;" in turn, insights from informants can be used to design computational analyses (Pardo-Guerra & Pahwa, 2022, p. 1828). Using abductive logics of inquiries between computational analysis and qualitative interpretation, cultural sociology is able "to locate surprising empirical findings" and to discover "innovative and creative theoretical insights" (Brandt & Timmermans, 2021, p. 192).

Furthermore, LLMs help to annotate and classify texts, increasing the precision of subsequent analysis. Results from supervised ML algorithms can also be used to fine-tune sociological concepts because LLMs can detect and suggest previously unconsidered notions.

The use of ML algorithms also challenges basic sociological assumptions about how to validate results. According to DiMaggio, using unsupervised ML algorithms requires "moving outside our comfort zone in accepting interpretive uncertainty and developing robust ways to interpret and validate the results of our models" (2015, p. 2). Similarly, using ML algorithms also questions what constitutes the outcome of a sociological analysis, be it description (Savage, 2009), prediction, or explanation (Hofman et al., 2021; Watts, 2014).

As ML algorithms become more widely used, we believe that researchers will become more aware of the implicit assumptions that underpin them and the datasets used. When sociologists began to use linear regression as their dominant mode of analysis, it significantly impacted how they construed the social world: according to a "general linear reality" (Abbott, 1988).[12] ML algorithms now call for methodological bricolage and present researchers with new interpretative uncertainties. We suspect that the reality of working with ML algorithms, including those methodological flexibilities and uncertainties, will similarly impact how sociologists construe the social world and their own discipline. One result would be to reflect further on the underlying assumptions of the methods and datasets used for cultural analysis. Such an increase in methodological reflexivity can only be beneficial.

Bibliography

Abbott, A. (1988). Transcending General Linear Reality. *Sociological Theory*, *6*, 169-188.

Arnold, T., & Tilton, L. (2019). Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities*, *34*(Supplement_1), i3-i16.

Arora, S., May, A., Zhang, J., & Ré, C. (2020). Contextual embeddings: When are they worth it? *arXiv preprint arXiv:2005.09117*.

Arseniev-Koehler, A., & Foster, J. G. (2022). Machine learning as a model for cultural learning: Teaching an algorithm what it means to be fat. *Sociological Methods & Research*, *51*(4), 1484-1539.

[12] For instance, they started to consider that entities under observation are fixed, that the effects of a variable are monotonous across the data space, or that causality flows in a certain direction only.

Ash, E., Gauthier, G., & Widmer, P. (2021). Text semantics capture political and economic narratives. *Center for Law & Economics Working Paper Series*, *2021*(11), 2108.01720.
Bail, C. A. (2014). The cultural environment: measuring culture with big data. *Theory and Society*, *43*(3-4), 465-482. https://doi.org/10.1007/s11186-014-9216-5
Bail, C. A. (2023). Can Generative AI Improve Social Science? *SocArXiv*. https://doi.org/10.31235/osf.io/rwtzs
Banerjee, M., Cole, B. M., & Ingram, P. (2023). "Distinctive from What? And for Whom?" Deep Learning-Based Product Distinctiveness, Social Structure, and Third-Party Certifications. *Academy of Management Journal*, *66*(4), 1016-1041. https://doi.org/10.5465/amj.2021.0175
Basta, C., Costa-Jussà, M. R., & Casas, N. (2019). Evaluating the underlying gender bias in contextualized word embeddings. *arXiv preprint arXiv:1904.08783*.
Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*. http://dx.doi.org/10.1002/asi.23786
Bearman, P. S. (2015). Big Data and historical social science. *Big Data & Society*(2). http://bds.sagepub.com/content/2/2/2053951715612497.full
Bearman, P. S., & Stovel, K. (2000). Becoming a Nazi: A model for narrative networks. *Poetics*, *27*(2), 69-90.
Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587-604.
Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the dangers of stochastic parrots: Can language models be too big?? FAcct 2021. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency,
Berelson, B. (1952). *Content analysis in communication research*. The Free Press.
Bernasco, W. M., Hoeben, E., Koelma, D., Liebst, L. S., Thomas, J. Appelman, J., . . . Lindegaard, M. R. (2023). Promise Into Practice: Application of Computer Vision in Empirical Research on Social Distancing. *Sociological Methods & Research*, *52*(3), 1239-1287. https://doi.org/10.1177/00491241221099554
Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM*, *55*(4), 77-84.
Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. Srivastava & M. Sahami (Eds.), *Text mining: classification, clustering, and applications* (pp. 71-93). Chapman & Hall.
Blei, D. M., Ng, A. Y., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993-1022.
Boelaert, J., & Ollion, É. (2018). The Great Regression. Machine Learning, Econometrics, and the Future of Quantitative Social Sciences. *Revue Francaise De Sociologie*. https://hal.archives-ouvertes.fr/hal-01841413
Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in neural information processing systems*, *29*. https://doi.org/10.48550/arXiv.1607.06520
Bonikowski, B., Luo, Y., & Stuhler, O. (2022). Politics as Usual? Measuring Populism, Nationalism, and Authoritarianism in US Presidential Campaigns (1952–2020) with Deep Neural Language Models. *Sociological Methods and Research*, *51*(4), 1721-1787.
Bonikowski, B., & Nelson, L. K. (2022). From ends to means: The promise of computational text analysis for theoretically driven sociological research. *Sociological Methods & Research*, *51*(4), 1469-1483.
Bourdieu, P. (1984). *Distinction: a social critique of the judgment of taste*. Havard University Press. (1979)
Brandt, P., & Timmermans, S. (2021). Abductive Logic of Inquiry for Quantitative Research in the Digital Age. *Sociological Science*, *8*, 191–210. https://doi.org/10.15195/v8.a10

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., . . . Askell, A. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Brunet, M.-E., Alkalay-Houlihan, C., Anderson, A., & Zemel, R. (2019). *Understanding the Origins of Bias in Word Embeddings* Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v97/brunet19a.html

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency*, 77-91. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186, Article apr, PubMed ID: 28408601 Publisher: American Association for the Advancement of Science Section: Reports. https://doi.org/10.1126/science.aal4230

Carley, K. M. (1994). Extracting culture through textual analysis. *Poetics*, *22*, 291-312.

Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),

Davidson, T. (2024). Hate Speech Detection and Bias in Supervised Text Classification. In J. P. Pardo-Guerra & C. Borch (Eds.), *The Oxford Handbook of the Sociology of Machine Learning*. Oxford University Press.

Denny, M. J., & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, *26*(2), 168-189. https://doi.org/10.1017/pan.2017.44

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 1-16.

DiMaggio, P. (2011). Cultural Networks. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 286-300). SAGE.

DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*(2). http://bds.sagepub.com/content/2/2/2053951715602908.full

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570-606. https://doi.org/http://dx.doi.org/10.1016/j.poetic.2013.08.004

Do, S., Ollion, É., & Shen, R. (2022). The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. *Sociological Methods & Research*, 1-34.

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, *46*, 61-81.

Erikson, E. (2021). *Trade and Nation. How Companies and Politics Reshaped Economic Thought*. Columbia University Press. https://doi.org/doi:10.7312/erik18434

Evans, J. A., & Aceves, P. (2016). Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, *42*, 21-50. https://doi.org/10.1146/annurev-soc-081715-074206

Fligstein, N., Brundage, J. S., & Schultz, M. (2017). Seeing Like the Fed: Culture, Cognition, and Framing in the Failure to Anticipate the Financial Crisis of 2008. *American Sociological Review*, *82*(5), 879-909

Franzosi, R. (1997). Mobilization and counter-mobilization processes: From the 'red years' (1919-1920) to the 'black years' (1921-1922) in Italy. *Theory and Society*, *26*, 275-304.

Franzosi, R. (2004). *From words to numbers: narrative, data, and social science*. Cambridge University Press.

Fuhse, J., & Mische, A. (2024). Relational Sociology: Networks, Culture and Interaction. In J. McLevey, J. Scott and P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 55–71). Sage.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635-E3644. https://doi.org/10.1073/pnas.1720347115

Gebru, T. , Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H. Iii, H. D., . . . Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86-92.

Ghosh, S., & Caliskan, A. (2023). ChatGPT Perpetuates Gender Bias in Machine Translation and Ignores Non-Gendered Pronouns: Findings across Bengali and Five other Low-Resource Languages. *arXiv preprint arXiv:2305.10510*.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, *7*(2), 2053951720943234.

Goldberg, A. (2015). In defense of forensic social science. *Big Data & Society, 2*(2). http://bds.sagepub.com/content/spbds/2/2/2053951715601145.full.pdf

Goldenstein, J., & Poschmann, P. (2019). Analyzing Meaning in Big Data: Performing a Map Analysis Using Grammatical Parsing and Topic Modeling. *Sociological Methodology*, *49*(1), 83-131. https://doi.org/10.1177/0081175019852762

Goldstein, Y., Legewie, N. M., & Shiffer-Sebba, D. (2023). 3D Social Research: Analysis of Social Interaction Using Computer Vision. *Sociological Methods & Research*, *52*(3), 1201-1238. https://doi.org/10.1177/00491241221147495

Gonen, H., & Goldberg, Y. (2019). Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *arXiv preprint arXiv:1903.03862*. https://doi.org/10.48550/arXiv.1903.03862

GPAI. (2023). *Fairwork AI Ratings 2023: The Workers Behind AI at Sama.* Global Partnership on AI. https://gpai.ai/projects/future-of-work/FoW-Fairwork-AI-Ratings-2023.pdf

Gray, I., & Cointet, J. P. (2023). Multilateralism of the Marginal: How the Least Developed Countries Find Their Voice in International Political Deliberations.

Grigoropoulou, N., & Small, M. L. (2022). The data revolution in social science needs qualitative research. *Nature Human Behaviour*, *6*(7), 904-906.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267-297. https://doi.org/10.1093/pan/mps028

Heiberger, R. H., Munoz-Najar Galvez, S., & McFarland, D. A. (2021). Facets of Specialization and Its Relation to Career Success: An Analysis of U.S. Sociology, 1980 to 2015. *American Sociological Review*, *86*(6), 1164-1192. https://doi.org/10.1177/00031224211056267

Hofman, J. M. Watts, D. J. Athey, S. Garip, F. Griffiths, T. L. Kleinberg, J., . . . Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Hurtado Bodell, M., Arvidsson, M., & Magnusson, M. (2019). Interpretable word embeddings via informative priors. *arXiv preprint arXiv:1909.01459*.

Hurtado Bodell, M., Magnusson, M., & Mützel, S. (2022). From Documents to Data: A Framework for Total Corpus Quality. *Socius*, *8*, 23780231221135523. https://doi.org/10.1177/23780231221135523

Ignatow, G. (2020). *Sociological theory in the digital age*. Routledge.

Ignatow, G., & Mihalcea, R. (2016). *Text Mining: A Guidebook for the Social Sciences*. Sage.

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

Jensen, J. L., Karell, D., Tanigawa-Lau, C., Habash, N., Oudah, M, & Fani, D. (2022). Language Models in Sociological Research: An Application to Classifying Large Administrative Data and Measuring Religiosity. *Sociological Methodology*, *52*(1), 30-52.

Jepperson, R. L., & Swidler, A. (1994). What properties of culture should we measure? *Poetics*, *22*, 359-371.

Kapoor, S., Cantrell, E., Peng, K., Pham, T. H., Bail, C. A., Gundersen, O. E., . . . Malik, M. M. (2023). Reforms: Reporting standards for machine learning based science. *arXiv preprint arXiv:2308.07832*.

Karell, D., & Freedman, M. (2019). Rhetorics of Radicalism. *American Sociological Review*, *84*(4), 726-753. https://doi.org/10.1177/0003122419859519

Karell, D., & Freedman, M. (2020). Sociocultural mechanisms of conflict: Combining topic and stochastic actor-oriented models in an analysis of Afghanistan, 1979–2001. *Poetics*, *78*(October 2019), 101403-101403. https://doi.org/10.1016/j.poetic.2019.101403

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

Koenecke, A., Nam, A. ,Lake, E., Nudell, J., Quartey, M., Mengesha, Z., . . . Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684-7689. https://doi.org/doi:10.1073/pnas.1915768117

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905-949. https://doi.org/10.1177/0003122419877135

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. SAGE.

Lee, M., & Martin, J. L. (2015). Coding, counting and cultural cartography. *American Journal of Cultural Sociology*, *3*, 1-33. https://doi.org/10.1057/ajcs.2014.13

Leschziner, V., & Brett, G. (2021). Have Schemas Been Good To Think With? Sociological Forum,

Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, *4*(10), 1021-1028. https://doi.org/10.1038/s41562-020-0918-6

Light, R., & Odden, C. (2017). Managing the Boundaries of Taste: Culture, Valuation, and Computational Social Science. *Social Forces*, *96*(2), 877-908. https://doi.org/10.1093/sf/sox055

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., . . . Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Lizardo, O. (2017). Improving cultural analysis: Considering personal culture in its declarative and nondeclarative modes. *American Sociological Review*, *82*(1), 88-115.

Lundberg, I., Brand, J. E., & Jeon, N. (2022). Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research*, *108*, 102807.

Luo, Y., Card, D., & Jurafsky, D. (2020). Detecting stance in media on global warming. *arXiv preprint arXiv:2010.15149*.

Luo, Y., Gligorić, K., & Jurafsky, D. (2023). Othering and low prestige framing of immigrant cuisines in US restaurant reviews and large language models. *arXiv preprint arXiv:2307.07645*.

Mazières, A., Menezes, T., & Roth, C. (2021). Computational appraisal of gender representativeness in popular movies. *Humanities and Social Sciences Communications*, *8*(1), 1-9.

McFarland, D. A., Lewis, K., & Goldberg, A. (2016). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*, *47*(1), 12-35. https://doi.org/10.1007/s12108-015-9291-8

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *CoRR*, *abs/1301.3781*, 1-12. http://arxiv.org/abs/1301.3781

Miner, A. S., Stewart, S. A., Halley, M. C., Nelson, L. K., & Linos, E. (2023). Formally comparing topic models and human-generated qualitative coding of physician mothers' experiences of workplace discrimination. *Big Data & Society*, *10*(1), 20539517221149106. https://doi.org/10.1177/20539517221149106

Mische, A. (2008). *Partisan Publics. Communication and contention across Brazilian youth activist networks*. Princeton University Press.
Mische, A., & Pattison, P. (2000). Composing a civic arena: Publics, projects, and social settings. *Poetics*, *27*, 163-194.
Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., . . . Gebru, T. (2019). *Model Cards for Model Reporting* Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA. https://doi.org/10.1145/3287560.3287596
Mohr, J. W. (1998). Measuring meaning structures. *Annual Review of Sociology*, *24*, 345-370.
Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDonnell, T. E., . . . Wherry, F. F. (2020). *Measuring Culture*. Columbia University Press.
Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics*, *41*(6), 545-569.
Mohr, J. W., & Duquenne, V. (1997). The duality of culture and practice: poverty relief in New York City, 1888-1917. *Theory and Society*, *26*, 305-356.
Mohr, J. W., Wagner-Pacifici, R., & Breiger, R. L. (2015). Toward a computational hermeneutics. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715613809
Mützel, S. (2015a). Facing Big Data: Making sociology relevant. *Big Data & Society, 2*(2). http://doi.org/10.1177/2053951715599179
Mützel, S. (2015b). Structures of the Tasted: Restaurant Reviews in Berlin Between 1995 and 2012. In A. B. Antal, M. Hutter, & D. Stark (Eds.), *Moments of Valuation: Exploring Sites of Dissonance* (pp. 147-167). Oxford University Press.
Mützel, S. (2022). *Making Sense: Markets from Stories in New Breast Cancer Therapeutics*. Stanford University Press.
Nangia, N., Vania, C., Bhalerao, R., & Bowman, S. R. (2020). CrowS-pairs: A challenge dataset for measuring social biases in masked language models. *arXiv preprint arXiv:2010.00133*.
Nassauer, A., & Legewie, N. M. (2021). Video Data Analysis: A Methodological Frame for a Novel Research Trend. *Sociological Methods & Research*, *50*(1), 135-174.
Nelson, L. K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, *49*(1), 3-42.
Nelson, L. K. (2021). Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century U.S. South. *Poetics*, *88*, 101539. https://doi.org/https://doi.org/10.1016/j.poetic.2021.101539
Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research*, *50*(1), 202–237.
Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown.
Ollion, É., Shen, R., Macanovic, A., & Chatelain, A. (2024). The dangers of using proprietary LLMs for research. *Nature Machine Intelligence*, *6*(1), 4-5. https://doi.org/10.1038/s42256-023-00783-6
Pachucki, M. A., & Breiger, R. L. (2010). Cultural Holes: Beyond Relationality in Social Networks and Culture. *Annual Review of Sociology*, *36*, 205-224.
Palmer, A., Smith, N. A., & Spirling, A. (2024). Using proprietary language models in academic research requires explicit justification. *Nature Computational Science*, *4*(1), 2-3. https://doi.org/10.1038/s43588-023-00585-1
Pardo-Guerra, J. P., & Pahwa, P. (2022). The Extended Computational Case Method: A Framework for Research Design. *Sociological Methods & Research*, *51*(4), 1826-1867. https://doi.org/10.1177/00491241221122616
Perrigo, B. (2023, January 18, 2023). OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. *Time*. https://time.com/6247678/openai-chatgpt-kenya-workers/

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., . . . Zettlemoyer, L. (2018). Deep contextualized word representations. . *arXiv preprint arXiv:1802.05365*.
Rawlings, C. M., & Childress, C. (2021). Measure Mohr culture. *Poetics*, *88*, 101611. https://doi.org/https://doi.org/10.1016/j.poetic.2021.101611
Ren, C., & Bloemraad, I. (2022). New Methods and the Study of Vulnerable Groups: Using Machine Learning to Identify Immigrant-Oriented Nonprofit Organizations. *Socius*, *8*, 23780231221076992. https://doi.org/10.1177/23780231221076992
Rheault, L., & Cochrane, C. (2020). Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis*, *28*(1), 112-133.
Roberts, S. T. (2019). *Behind the screen*. Yale University Press.
Rodriguez, P. L., & Spirling, A. (2022). Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, *84*(1), 101-115. https://doi.org/10.1086/715162
Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test–retest reliability of continuous measurements. *Statistics in medicine*, *21*(22), 3431-3446.
Rozado, D. (2020). Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PlosOne*, *15*(4), e0231189. https://doi.org/https://doi.org/10.1371/journal.pone.0231189
Salganik, M. (2017). *Bit by Bit*. Princeton University Press.
Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., . . . McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8398-8403. https://doi.org/10.1073/pnas.1915006117
Savage, M. (2009). Contemporary Sociology and the Challenge of Descriptive Assemblage. *European Journal of Social Theory*, *12*(1), 155-174. https://doi.org/10.1177/1368431008099650
Sen, I., Flöck, F., Weller, K., Weiss, B., & Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *Public Opinion Quarterly*, *85*(S1), 399–422. https://doi.org/https://doi.org/10.1093/poq/nfab018
Shadrova, A. (2021). Topic models do not model topics: epistemological remarks and steps towards best practices. *Journal of Data Mining & Digital Humanities*, *2021*, 1-28. https://jdmdh.episciences.org/
Spirling, A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature*, *616*(7957), 413-413. https://EconPapers.repec.org/RePEc:nat:nature:v:616:y:2023:i:7957:d:10.1038_d41586-023-01295-4
Stoltz, D. S., & Taylor, M. A. (2021). Cultural cartography with word embeddings. *Poetics*, *88*, 101567. https://doi.org/https://doi.org/10.1016/j.poetic.2021.101567
Stuhler, O. (2022). Who Does What to Whom? Making Text Parsers Work for Sociological Inquiry. *Sociological Methods & Research*, *51*(4), 1580-1633.
Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, *30*(3), 167-186.
Timmermans, S., & Tavory, I. (2022). *Data analysis in qualitative research: Theorizing with abductive analysis*. University of Chicago Press.
Törnberg, P. (2023). ChatGPT-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. *arXiv preprint arXiv:2304.06588*.
Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951720919776
Urman, A., & Makhortykh, M. (2023). The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat. https://doi.org/https://doi.org/10.31219/osf.io/q9v8f
van Noord, N. (2022). A survey of computational methods for iconic image analysis. *Digital Scholarship in the Humanities*, *37*(4), 1316-1338. https://doi.org/10.1093/llc/fqac003

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*, 1-11.

Voyer, A., Kline, Z. D., & Danton, M. (2022). Symbols of class: A computational analysis of class distinction-making through etiquette, 1922-2017. *Poetics*, *94*, 101734.

Watts, D. J. (2014). Common Sense and Sociological Explanations. *American Journal of Sociology*, *120*(2), 313-351. https://doi.org/10.1086/678271

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., . . . Kasirzadeh, A. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency,

White, H. C. (1992). *Identity and Control: A Structural Theory of Social Action*. Princeton University Press.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., . . . Schwartz, O. (2018). *AI Now Report 2018*. AI Now Institute at New York University New York.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., . . . Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language* Processing: System Demonstrations, 38-45.