

Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis

JESSICA L. FEUSTON, University of Colorado Boulder, USA

JED R. BRUBAKER, University of Colorado Boulder, USA

Large datasets or ‘big data’ corpora are typically the domain of quantitative scholars, who work with computational tools to derive numerical and descriptive insights. However, recent work asks how computational tools and other technologies, such as AI, can support qualitative scholars in developing deep and complex insights from large amounts of data. Addressing this question, Jiang et al. found that qualitative scholars are generally opposed to incorporating AI in their practices of data analysis. In this paper, we provide nuance to these earlier findings, showing that the stage of qualitative analysis matters for how scholars believe AI can and should be used. Through interviews with 15 CSCW and HCI qualitative researchers, we explore how AI can be included throughout different stages of qualitative analysis. We find that qualitative scholars are amenable to working with AI in diverse ways, such as for data exploration and coding, as long as it assists rather than automates their analytic work practice. Based on our analysis, we discuss how incorporating AI into qualitative research can shift some analytic practices, and how designing for human-AI collaboration in qualitative analysis necessitates considering tradeoffs in scale, abstraction, and task delegation.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; **Collaborative and social computing design and evaluation methods**.

Additional Key Words and Phrases: AI, human-AI collaboration, interview, qualitative research, qualitative analysis, user study.

ACM Reference Format:

Jessica L. Feuston and Jed R. Brubaker. 2021. Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 469 (October 2021), 25 pages. <https://doi.org/10.1145/3479856>

1 INTRODUCTION

The quantity of data is growing. While this growth provides new opportunities for understanding the world, it requires that we develop research practices and tools that can keep pace. Diverse fields, including computational social science, data science, and information science, turn to “big data,” using sophisticated tools to interpret large datasets. Supported by the steady increase of computing power, it is now possible to use artificial intelligence (AI) ¹ systems on large datasets to generate new insights. AI excels at providing descriptive overviews, frequency counts, and

¹Our use of “AI” describes automated and intelligent systems and tools [8, 25].

Authors’ addresses: Jessica L. Feuston, University of Colorado Boulder, Department of Information Science, 1045 18th Street, Boulder, Colorado, USA, jessica.feuston@colorado.edu; Jed R. Brubaker, University of Colorado Boulder, Department of Information Science, 1045 18th Street, Boulder, Colorado, USA, jed.brubaker@colorado.edu.



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

© 2021 Copyright held by the owner/author(s).
2573-0142/2021/10-ART469. <https://doi.org/10.1145/3479856>

statistical connections — insights that are typically quantitative in nature. But what about other types of insight?

Prior work argues for complementary and contrasting theoretical and methodological ways of knowing the world [5, 26]. In comparison to quantitative analysis, qualitative approaches to data analysis facilitate deep understanding that reflects the richness, complexity, and, at times, contradictory nature of people and phenomena. However, traditional qualitative approaches do not account for analyzing data that number in the millions. Imagine, for example, the time and labor it could take for a qualitative scholar or team to read and make sense of millions of social media posts — were such a thing possible or expected.

A growing body of research addresses the challenges and opportunities related to making large datasets accessible for qualitative scholarship [3, 12, 17, 25, 33, 35, 43, 49]. Within this prior work, scholars have explored automating analysis (e.g., [32, 49]) and have documented qualitative scholars' high-level concerns about the inclusion of AI in their work (e.g., [25]). Despite concerns, there is a strong potential for human-AI collaboration in qualitative research if the design of AI in this context shifts away from providing automated answers and, instead, focuses on supporting scholars' analytical processes [25]. In this paper, we extend prior work by examining the use of AI at different stages of qualitative analysis, including data exploration and coding. We aim to understand and recommend how AI can support developing qualitative insights from large datasets.

We interviewed 15 Computer Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) scholars who perform qualitative analysis with large corpora. Many of our participants were familiar with using AI to support their qualitative practice. For those who were not, we developed three design scenarios to prompt conversation and reflection. Across our interviews, we found that using AI during different stages of qualitative analysis necessitates understanding and preparing for several trade-offs. Though AI can be beneficial, the use of tools is liable to shift how qualitative work is conducted. We detail 'shifts' to scale, abstraction, and task delegation in our discussion, addressing how we might design AI to support qualitative analysis with them in mind.

We make three primary contributions. The first is a detailed empirical study that examines how CSCW and HCI qualitative scholars use — or could use — AI as part of qualitative analysis. Many of our participants had used or experimented with using AI to support their analytic practices when working with large corpora. Building off of these experiences, as well feedback to the scenarios we presented to our participants, we illustrate how AI can impact different stages of qualitative analysis. The ways that AI can influence the work of human analysts have implications for considering who or what is the primary interpretive lens for analysis as well as when scholars may delegate tasks or accept algorithmic support.

Second, based on these findings, we discuss how human-AI collaboration in qualitative analysis presents three primary shifts to traditional ways of conducting qualitative analysis. These shifts include scalability, abstraction, and delegation, and are largely dependent on the stage of qualitative analysis and the analytic progress of the human scholar. We describe these shifts in conversation with prior work related to human-AI collaboration, attending to the ways that humans and AI can augment one another in the development of qualitative insights. Finally, based on how AI can shift analytic tasks, we suggest design recommendations for human-AI collaboration in qualitative analysis. We argue for designing for collaboration, mitigating the amount of work new algorithmic tools may create for qualitative scholars, and using AI to identify gaps in human analysis.

2 BACKGROUND

Qualitative scholarship involves a particular methodological (e.g., data collection, analysis) commitment to research, where scholars generate insights from rich and situated contexts (e.g., such

as collected from observations, interviews, focus groups). Though qualitative methods may be used within any particular research tradition or paradigm, within CSCW and HCI, we typically see qualitative scholarship in conjunction with interpretivism, which aims to understand people's subjective and co-constructed experiences, perspectives, and values. With respect to data analysis, qualitative approaches include inductive and deductive logics, where insights might be generated from developing patterns from data (inductive) or from checking "existing patterns and theories against data in a top-down fashion" (deductive) [25]. In our work, we focus on inductive approaches to qualitative data analysis.

Scholars take many different approaches to inductive qualitative analysis. Common examples within CSCW and HCI scholarship include the constant comparative method [20], grounded theory methods [10, 21, 47], and thematic analysis [7]. Though each of these methods is distinct, Saldaña summarizes the inductive qualitative coding process in his codes-to-theory model for qualitative analysis [44], which describes how most analytic approaches involve coding data, organizing codes into categories, and comparing categories to "transcend the 'reality' of your data and progress toward the thematic, conceptual, and theoretical" [44]. The codes-to-theory model also illustrates how, as analysis progresses, qualitative researchers move from the real and particular (i.e., observations from the data, codes) to the abstract and general (i.e., categories, themes). We use this model to highlight similarities across different traditions of qualitative analysis and to act as a point of comparison for qualitative data analysis involving algorithmic tools.

Beyond the codes-to-theory model, we also consider how data exploration — in particular, sampling — impacts qualitative analysis. For qualitative scholars, sampling provides a way of entering into a domain and determining what data to collect as well as analyze. In traditional qualitative work, sampling strategies may involve convenience sampling, purposive sampling, theoretical sampling, and snowball sampling [6, 10, 41]. When qualitative scholars work with large corpora, however, their sampling strategies change. For example, scholars may generate a random subsample from the data through an algorithmic tool. Using algorithmic tools to generate subsamples differs from traditional qualitative data sampling approaches and impacts how analysis proceeds. For these reasons, alongside the codes-to-theory model, we consider sampling an integral element of how qualitative scholars perform data analysis.

3 RELATED WORK

To situate our contribution, in this section we describe prior research related to AI assistance in qualitative analysis. Given our empirical focus, we highlight the potential of AI at different stages of analysis and during data exploration. We also review human-AI collaboration scholarship, attending to how AI can act as tools and partners.

3.1 Machine Learning at Different Stages of Qualitative Analysis

A small, though increasing, number of publications examines how AI can support qualitative scholars during data analysis (see [25] for a summary). Rather than summarize, we place prior work in conversation with Saldaña's codes-to-theory model and our included emphasis on data exploration to understand the stages of qualitative analysis that AI systems are being designed to support. Based on our review, we find that human-AI collaboration largely takes place during exploratory phases of qualitative work and certain (often early) stages of coding.

While work at the intersection of AI and qualitative scholarship remains nascent, using AI to assist qualitative scholars in the exploration of large corpora has been one of the largest areas of focus [4, 15, 33]. For example, early work by Muller et al. investigates comparative and hybrid approaches using grounded theory methods and machine learning [33]. In several suggested configurations of these techniques, Muller et al. illustrate how machine learning classification and

‘large-scale pattern analysis’ can inform deep, qualitative dives into specific areas of a dataset. The unique combination of machine learning and grounded theory methods is further explored and articulated through Nelson’s computational grounded theory, a three-step methodological framework combining human interpretive ability with computational pattern detection [35]. Similar work by Rodriguez and Storer, though veering away from grounded theory methods, demonstrates how results produced through topic models were comparable – yet distinct – from a first round of thematic analysis [43].

Scholars have also developed AI systems for visualizing large corpora (e.g., such as through topic modeling and global keyword frequency) [4, 11]. Systems such as these illustrate a clear need for AI that supports exploration and the development of high-level inferences and interpretive insights. Taking a slightly different approach (i.e., where human-AI collaboration informs exploration), Paredes et al. extend AI-facilitated exploration through their search assistant, which uses a skip-gram variation of word2vec to “search for semantically similar expressions” to human queries [37]. Across all of these examples, we see novel, creative ways of using AI to explore and potentially subsample large corpora. These systems highlight how AI can inspire and inform, rather than prescribe or insist where scholars look or sample. Building from these systems, our work focuses on how qualitative scholars might explore and sample large datasets at different stages of their analytic process.

In addition to the work above, scholars have discussed and developed AI that assists at certain stages of qualitative coding [3, 17, 32, 33, 39, 40, 49]. Systems designed by Marathe and Toyama [32] and by Yan et al. [49] automate elements of coding, focusing on applying codes across a dataset. Taken together, these studies illustrate how automated coding can extend and augment human ability, particularly when systems support a back and forth between human analysts and AI (e.g., such as through the evaluation of AI outcomes). However, a limitation of these systems is their reliance on the standardization of codes, such as through a codebook. As we echo in our findings, for some approaches to qualitative analysis, standardization does not occur until much later in the process of analysis, prompting questions about what role AI might play in the development of codes or when codes are less firm.

Most prior work on qualitative research and AI tends to focus on a single qualitative scholar. Yet qualitative coding is frequently a collaborative endeavor, involving a team of researchers. Prior works by Chen et al. [12], Drouhard et al. [16], and Zade et al. [51] investigate how to support human collaboration during qualitative coding, such as by highlighting ambiguity, inconsistency, and disagreement between human analysts. Similarly, Patton et al. illustrate how their Contextual Social Media Analysis (CASM) approach, which involves collaborative interpretation and coding amongst team members, can be used to train an NLP classifier [38]. These studies demonstrate how AI tools present opportunities to augment human capability. We extend this work through consideration of human-AI collaborations and partnerships, aiming to paint AI as yet another actor within qualitative analysis.

As this collection of literature shows, many scholars have been looking to leverage machine learning techniques for qualitative analysis. Another thread of scholarship argues that machine learning as a practice is fundamentally inductive and, therefore, is compatible with qualitative ways of producing knowledge. Muller et al. compare grounded theory with clustering [33], not to argue for equivalency between these two techniques, but to illustrate how their epistemological orientations to the data – and claims they can make about the data – are similar in their inductive, bottom-up approach (e.g., rather than deductive or top-down). Similar arguments have been articulated in adjacent fields, most notably by Nelson [35, 36], who provides a framework to help us consider the use of machine learning in qualitative analysis and engage in what she calls computational grounded theory. However, though machine learning is epistemologically aligned with qualitative

analysis, it is not qualitative analysis. As Rodriguez and Storer show, there are key differences between how AI and human analysts practice induction and interpret phenomena [43]. Within this context, the focus of our study is to address how qualitative scholars understand the use of machine learning techniques in qualitative research, exploring their existing practices and openness to working with AI.

3.2 Designing for Human-AI Collaboration

Successful human-AI collaboration requires a delicate balance. Collaborations must support humans, such as through task delegation to AI, without displacing them. Lubars and Tan propose a framework of task delegability for human-AI collaboration [30], arguing that we need to consider what AI *should* do, rather than what it *can* do, and proposing four factors for evaluating task delegation to AI: motivation, difficulty, risk, and trust. Building from this framework, Jiang et al. discuss task delegation in the context of qualitative data analysis [25]. In part, they ask, “If qualitative research as a single task should not be delegated, what parts of it are appropriate for AI to assist with?” We address this question by examining how the use of AI impacts different stages of qualitative analysis, including data exploration and coding.

In line with Lubars and Tan [30], prior works by Jiang et al. [25] and Wang et al. [48] argue that algorithmic systems should augment rather than automate human qualitative analytic practices. We draw inspiration from this argument through the consideration of a specific type of human-machine collaboration: human-machine symbiosis. Current conceptualizations of human-machine symbiosis are grounded in Licklider’s “Man-Computer Symbiosis” [28], which imagined close cooperative relationships between humans and computers in order to achieve types of thought and data processing not possible outside the pairing [19]. Recent work by Cho and Rader contends that continuous learning is essential for mutually beneficial human-machine partnerships (e.g., such as where a machine is able to continuously learn from human activity and evaluation) [13, 34]. Similarly, work by Jarrahi illustrates the interplay between the strengths and weaknesses of human scholars and algorithmic systems [24]. Considering these tradeoffs can help us develop algorithmic systems that augment qualitative analysis, rather than automate it or get in the way. Accordingly, we return to human-machine collaboration in our discussion, where we consider how AI may shift qualitative data analysis, and how we can anticipate these shifts and design with them in mind.

4 METHOD

We interviewed CSCW and HCI scholars who conduct qualitative analysis with ‘short texts.’ In our recruitment materials, we described short texts as texts that “included, but are not limited to, content such as tweets, Facebook posts and status updates, Tumblr posts, Reddit posts, newspaper headlines, and text messages.” We intended to speak with individuals who worked with short texts and large corpora, a combination we see as fruitful with respect to the potential inclusion of algorithmic tools. We asked our participants about their practices related to qualitative analysis and talked with them through several ‘machine learning in qualitative analysis’ scenarios, described below.

We recruited participants from our social and professional networks. We also relied on word of mouth and the sharing of our study’s recruitment flyers. Our study received approval from our university’s Institutional Review Board. Interviews lasted approximately 90 minutes and were conducted remotely using Zoom. For their time, our participants were compensated with a \$50 Amazon gift card.

4.1 Participants

Overall, we conducted semi-structured interviews with 15 CSCW and HCI scholars who have experience with qualitative analysis of short texts. In terms of gender, our participants included cis and trans scholars. Our sample was comprised predominantly of women ($n = 10$). We also interviewed two men, two individuals who had “no good answer,” and one individual who described herself as “gender agnostic” (and who used she/her/hers pronouns). Participants ranged in age from 28 to 42 years old ($M = 32.53$). All participants were from the United States.

Our goal was to speak with individuals who had experience and expertise with qualitative methods — analysis, in particular. We spoke with two Ph.D. candidates in their fifth and sixth years of study; one researcher who recently defended her dissertation; four postdoctoral researchers; two industry design (MS) and user experience (Ph.D.) researchers; five assistant professors; and one associate professor. All of the scholars we spoke with perform qualitative work. Our sample included 14 qualitative researchers and one predominantly quantitative researcher, though with extensive experience conducting qualitative analysis. While this quantitative participant met the criteria for our study, they also allowed us to gain a deeper sense of those at the edges of our research population.

Within our community, our sample could be readily identifiable. To protect the anonymity of our participants, we do not provide a mapping of demographic information to participant IDs. Additionally, when we quote participants in the following sections, we remove and modify project-specific information.

4.2 Interview Protocol

We used video for all of our interviews, though only the audio portion was recorded and later transcribed for analysis. During several interviews, participants shared their screens to show us de-identified study materials (e.g., codebooks, qualitative coding software, manuscripts in submission, and published manuscripts). We did not ask participants to share study materials that contained non-anonymized details.

We talked through participants’ de-identified materials during interviews as a way to understand our participants’ practices related to qualitative data analysis. The interview audio often captured the structure and elements of interest (e.g., the structure of a codebook or the formatting and topic of a memo). However, some of the materials shared by participants were distinctly visual in nature (e.g., graphs and other data visualizations). In these instances, we followed up with participants via email and asked them to share these de-identified materials. For example, following her interview, P7 shared a screenshot of a data visualization that had been used to inform qualitative analysis (and that we had discussed during her interview).

Our interview protocol consisted of two phases. In the first phase, we asked participants about a qualitative project where they had worked with short texts. Our interview questions focused on the approach to and process of qualitative analysis. For example, we asked questions concerning the steps people took to develop codes, categories, and themes. For additional context and clarity, we also asked about practices related to data collection and collaboration with other scholars. When asking about the technologies and tools used by participants during their practices of qualitative analysis, we were surprised to find that many of our participants had experience using algorithmic tools in their work. When possible, we asked about qualitative projects involving algorithmic tools to understand the process of analysis as well as the experience of working with algorithmic tools.

During the second phase of our interview protocol, we presented participants with three speculative scenarios (visualized as storyboards) that represented different ways algorithmic tools could be integrated into qualitative analysis (Figures 1, 2, and 3). Prior work has demonstrated the benefits

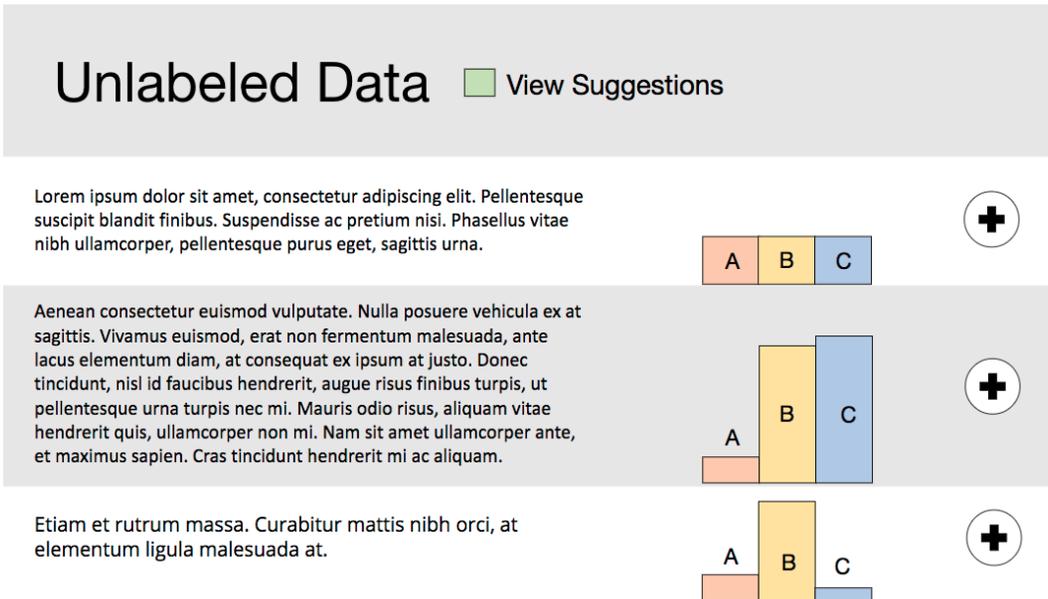


Fig. 1. Algorithmic Recommendations. In this scenario, we invited our participants to discuss the use of an algorithmic tool that recommended how uncoded data could be grouped. A, B, and C are examples of groups to which a text might be assigned. The algorithmic tool suggests how well a text fits into each group. For example, the algorithmic tool is suggesting B for the third text. We described these ‘groups’ — conceptual stand-ins for codes and categories — as being initially developed by a qualitative scholar, such that the algorithmic tool makes recommendations based on work a human has already conducted. In addition to selecting a group to assign a text by clicking on the bar chart, the scenario included the option to generate new groups by selecting the plus (+) sign.

of using scenarios to scaffold conversations about speculative futures and technologies, support participants in articulating their needs, concerns, and perspectives, and act as a reflexive way to understand the very real implications of an imagined technology [9]. We based the scenarios on examples from prior work [4, 25, 33, 49] as well as the expertise of our data visualization collaborators. Our first scenario, Algorithmic Recommendations illustrated how an algorithmic tool could recommend groups (as informed by a human analyst) to assign to a specific text. We used ‘groups’ as a general placeholder for codes, categories, and themes, and let our participants decide the type of segment they wanted to address. This allowed for conversations about how this tool could — or could not — work across different levels of abstraction, including codes and categories. The second scenario, Algorithmic Grouping, invited our participants to discuss the use of an algorithmic tool to automatically generate groups from the data. Finally, in the third scenario, Algorithmic Sampling, we talked with our participants about using algorithmically generated groups explicitly to guide data sampling.

Together, these three scenarios scaffolded talking about how algorithms could be included in qualitative analysis, particularly for researchers unfamiliar with contemporary algorithmic practices (e.g., as they relate to machine learning). Conversations surrounding these scenarios involved questions and concerns, as well as resistance and challenges to the inclusion of these approaches in qualitative analysis. Additionally, our use of scenarios supported participants in extending the storyboards we provided by imagining other uses of algorithmic tools or ways that algorithmic

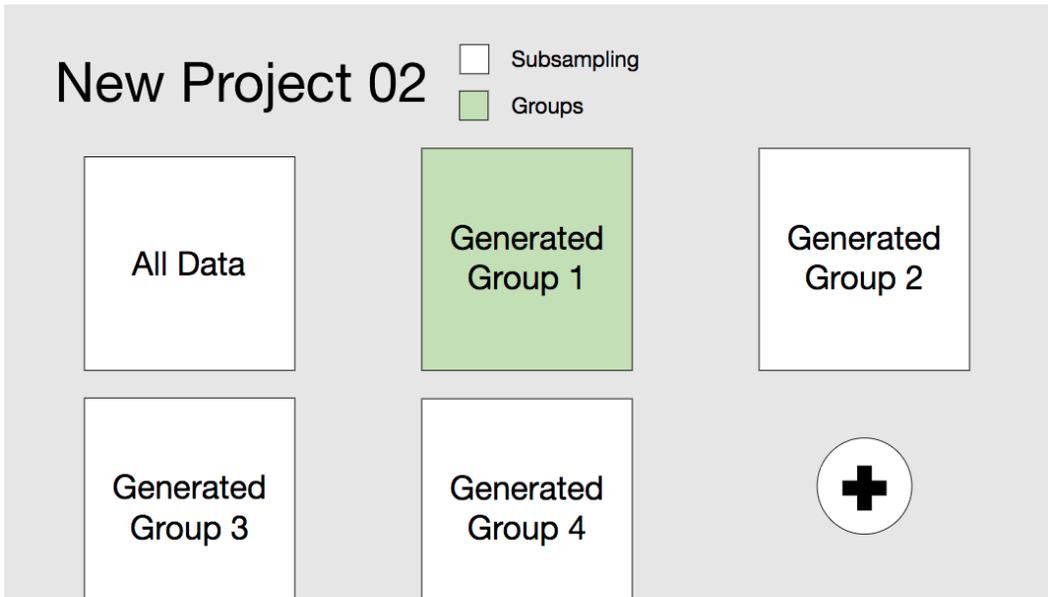


Fig. 2. Algorithmic Grouping. This scenario illustrates how an algorithmic tool could be used to automatically generate groups from a dataset. After starting a new project and uploading data, a participant can select the “Groups” option. The algorithmic tool automatically generates groups from the data (i.e., without being informed by any human analysis). We kept the particular algorithmic process vague to ask our participants about their expectations and needs without biasing them toward a specific algorithmic technique. However, when we needed to make the example more concrete (e.g., for participants unfamiliar with algorithms working in such a way) we talked about the groups as equivalent to latent Dirichlet allocation (LDA) generated topics.

tools could be modified to fit within their analytic practice [9]. Our participants frequently related our three scenarios back to their own qualitative practices and projects, including how our proposed algorithmic tools could look within grounded theory or content analysis approaches, as well as how these tools could have supported – and even gotten in the way – of a recent study.

4.3 Data Analysis

We adopted an inductive approach drawing from grounded theory [10] and thematic analysis [7] to analyze our data. We began to synthesize data during and throughout the interview process (e.g., through analytic memoing, highlighting of transcripts, and discussions between members of the research team), tailoring the direction of our semi-structured interviews to follow-up on initial topics of interest. These topics included, for example, concerns about bias in algorithmic approaches, the potential for algorithms to support the exploration of large corpora, the ability to have control when working with algorithmic tools, and instances where uncommon codes and categories were analytically significant.

Following data collection and transcription of interview audio, the first author engaged in open coding of the transcripts. Through and following discussions with the second author, she developed a preliminary set of categories from the data, including trust in collaborators and in algorithmic systems and ways of gaining familiarity with data, as well as several categories related to using algorithmic tools during qualitative analysis (e.g., to guide sampling, to develop codes, to handle disagreements). After reviewing the data several times, we decided to focus on categories that

generating clusters of words. Semantic network analysis works by clustering words that frequently co-occur [15]. In the tradition of grounded theory, P1 approached the dataset without theory or hypotheses. She explained:

So, we just started with that inductive approach. Without any theoretical framework, just throw it to the computer and then [the computer] just throws back a couple of clusters. And I found it really fascinating that we could outsource... [the start of the] coding process to... the computer.

With clusters in hand, P1 and her collaborator worked to make sense of what each cluster could mean. P1 talked about having to discard clusters, such as those with too few terms or terms that did not cohere. Nevertheless, the algorithmically produced clusters provided a productive starting point. Using the clusters and her initial interpretations of them, P1 returned to the dataset to sample texts and focus her analysis. P1 reflected on how using AI shifted her approach, saying:

[T]he biggest benefit I got is [that] someone told me what the clusters are... That's the first step... It generates a framework and then I go back to the corpus to find the relevant excerpts, the relevant tweets that's linked [to a cluster]. It's a very different approach, rather than me going through all the tweets generating this theme by myself. So, it completely changed the analytical process.

While P1 leveraged algorithmic tools to help develop codes for her large dataset, another participant, P7, detailed a very different situation where she worked with a team of undergraduate research assistants to code for a complex idea: care. Given the subtle ways that care could appear in the data, P7 wanted a way to guide her research assistants, to provide scaffolding as they worked to manually code a large dataset in which care only appeared sporadically.

Envisioning something akin to a highlighter, P7 devised a way to mark areas of the text that warranted attention. She started by manually developing a list of words related to care, then expanded this list of words by running it through a computational linguistic tool based on word2vec that had been trained with her entire dataset. Using this word2vec approach allowed P7 to compile similar words based on the use of language in her dataset. Finally, P7 filtered this initial set of terms into a manageable list through a crowdsourcing task in which she asked people to identify whether or not a word was related to care. Based on the outcome of this task, P7 capitalized care-related words in the dataset that she and her research assistants were coding. P7 explained how this process “was a lot of work to do.”

Despite the multiple layers in P7's process, as well as the diverse influences (i.e., word2vec, a crowdsourcing platform), P7 was adamant that the actual qualitative coding be done by people. She viewed her use of algorithmic tools as a way of helping her team focus their coding efforts, saying:

[T]hat was a way that we programmatically made it easier for our coders to do their job, because some of [the data] are so long, it'd be really difficult to maintain the level of focus that you need to get through that much text without some kind of a pointer to what's of interest.

Our examples with P1 and P7 represent the diversity of how AI can be used in qualitative analysis and illustrate the different stages where AI tools might be used (as well as what the ramifications of use might be). Both P1 and P7 incorporated AI in novel, creative ways to address problems they encountered (e.g., a massive corpus, disrupted focus while coding long texts). Their experiences prompt a number of questions: When during analysis should AI be used, and in what ways? What should the breakdown of responsibilities be in human-AI collaboration for qualitative analysis?

Across the next two sections, we describe how the qualitative scholars in our study used and imagined using AI in their data analysis. Based on our analysis of participant experiences with

AI (like those of P1 and P7) and the feedback participants provided during the scenario portion of our interview, we focus on the roles that AI might play during the exploration and the coding portions of qualitative analysis. Taken together, we found that the stage of qualitative analysis impacts how scholars think about and approach human-AI collaboration, including whether or not – and also how – to use algorithmic tools. We return to this point in depth in our discussion where we consider how AI might shift qualitative work practices and present design recommendations for human-AI collaboration in qualitative analysis.

5.2 Including Algorithmic Tools during Inductive Coding

Our participants highlighted concerns related to the use of AI during the coding stage of inductive qualitative analysis (e.g., when codes are being developed, during the application of codes to data). Many of the individuals we spoke with felt that using AI could unduly influence what they found interesting and how they interpreted their data. However, we found that a number of these concerns were mitigated by reconsidering the particulars of the scenario we presented, such as modifying the point when AI came to be involved.

5.2.1 AI as Instruments for Analysis. Coding is one of the most laborious parts of qualitative analysis. It is also one of the most important when it comes to understanding the data and deciding what is significant or meaningful. When scholars delegate coding work to AI, part of the focus and sensemaking that comes through analysis can be delegated as well. Many of our participants echoed similar concerns to those reported by Jiang et al. [25], particularly when we presented them our Algorithmic Grouping scenario (see Figure 2). In this scenario, an AI system generates groups of data from a large corpus to inform the inductive coding stage of qualitative analysis. This scenario prompted rich discussions related to human-AI collaboration, including at what point in analysis a coding task should be delegated to AI, if ever.

Participants took issue with the speculative scenario that AI could replace them as conduits for analysis during code development. P7 said, “*When I’ve heard about people using QDA software or whatever to kind of come up with codebooks, I’m like, ‘Bullshit.’ My first reaction is like, ‘Bullshit.’ You need that good human brains to do it.*” Similarly, P9 balked at this scenario, saying:

I would be worried that this wouldn’t be very congruent with how I do analysis because I am the instrument, and this is essentially taking me out of the instrumentation. So, then I would have a lot more concerns about whose lens is this really? How is it generating this stuff? Why? What biases are baked into this because I don’t really know. There’s no transparency.

P9’s comment echoes P7’s skepticism and highlights concerns related to control (e.g., “taking me out of the instrumentation”), something we see in prior work [25], as well as transparency [31, 46, 48]. AI that clusters data can act as an instrument for analysis. Through the assembling of groups, algorithmic tools come to be the lens through which meaning is made. As such, when algorithms take on analytic work related to code development (e.g., by generating clusters for a scholar prior to any human coding of the dataset), the analytic role of the human scholar can be displaced. Replacing human analysts with algorithmic ones raises concerns related to the capabilities of AI, the quality of the resulting analysis, and the nature of qualitative work.

Though most participant misgivings involved how AI-generated clusters could be used to inform code development, participants also talked about concerns related to the potential use of AI to organize codes, resulting in categories. P12 imagined a modified version of our Algorithmic Grouping scenario, where the AI generates groups from codes developed and used by human scholars (i.e., rather than groups from data). P12 said she was not interested in having “*an algorithm do the bundling step because to me the bundling step is actually a really important analytical step.*”

Machine learning isn't good enough to be able to... make the same kind of associations that I would." For P12, bundling (i.e., the organization of codes into categories) is an analytically important step. The associations she makes between codes are grounded in her subjectivity, expertise, and methodological approach — elements that AI cannot, and should not, replicate or replace.

Delegating inductive coding related tasks to AI, such as the development of clusters from data or human-produced codes, risks displacing human scholars as the primary analysts and reconfigures the set of actors involved in the instrumentation of qualitative analysis. In our discussion, we build on this thread of delegation to consider how AI can support and augment the work of scholars during inductive coding, potentially entering into partnerships with human analysts.

5.2.2 Reinforcing Past Perspectives during Coding. In the previous section, the scenario we discussed with participants involved AI that produced results from a dataset rather than any type of human involvement, input, or intervention. To understand participants' reactions to an AI tool informed by human practices of analysis, we presented our Algorithmic Recommendations scenario (Figure 1), where AI suggests codes for uncoded texts within a dataset. In this scenario, the codes suggested by AI are developed by human analysts, not computational ones. Though grounded in the practices of human analysts, several participants felt that AI that suggested codes would limit their creativity and analytic potential by infusing their own past biases and assertions into new, uncoded segments of a dataset.

In response to this scenario, P3 said, *"This is the first time I'm worrying. I mean the other things I think are useful tools that I don't see much problems bias wise with. [...] Once you get into suggestions though I start to get nervous."* When asked to describe her use of "bias," P3 said, *"[This AI is] going off the code we've already [developed]. [Using this AI] essentially becomes our past decisions asserting themselves over our present discovery."*

For many qualitative scholars, coding is a looping, flexible, and creative process. P3's concern illustrates how the codes she develops during analysis are not static or necessarily replicable from one portion of the dataset to another. Her analytic perspective, and codes with it, are liable to change and mature as she gains familiarity with her data. P8, similarly, described how his codes frequently change and shift throughout inductive coding — at least, to a point. Though analysis for P3 and P8 follows multiple looping iterations through data, codes must eventually calcify (e.g., such as through publication). Nevertheless, we see with these two participants how codes, by analytic necessity, remain malleable for some time throughout coding. Therefore, AI that suggests or applies codes across a dataset can disrupt the natural flexibility of some approaches to qualitative analysis.

Concerns related to how AI could standardize codes were not relevant for all of our participants. For some scholars in our study, standardization and calcification of codes occurred earlier in the analytic process. Several of our participants, including P5, P6, and P7, commented on similarities between the standardization of codes through AI recommendations and their own qualitative analysis practices (i.e., writing and validating codebooks). In a recent project using qualitative content analysis, P6 and her team spent several weeks validating a codebook such that researchers joining the team could consistently code for phenomena of interest without needing to change the codes. In this example, the standardization of codes is an integral part of analysis. It allows a framework to be established and applied by a number of different actors, potentially algorithmic ones.

The ways that AI can reinforce past perspectives, assertions, and biases during coding is sometimes a problem, sometimes not. Ultimately, it varies, depending on when during coding scholars decide to standardize their codes. This may occur relatively quickly, such as after several weeks of coding subsamples from a dataset. However, it may not happen until a scholar has written and submitted their work for publication. Understanding this variability helps us attend to how the

boundaries between different stages of the code-to-theory model might be inadvertently firmed by AI which, out of necessity, often involves procedures related to standardization (e.g., building specific classifiers). Though we see this type of boundary work in other forms of collaboration, here it helps us foreground important individual differences. As we address in our discussion, AI that supports qualitative analysis cannot be one-size-fits-all [49].

5.2.3 Providing New Perspectives throughout Inductive Coding. In the first section, we highlighted our participants' critiques related to working with AI that clusters data. Recall, our participants' ire was centered on the ways that AI could act as an analytic lens and influence how they developed codes and categories. With respect to timing, this meant that the use of AI occurred prior to scholars developing their own codes from the data. As working with scenarios is inherently flexible, we flipped the script to elicit different attitudes and opinions. We modified our Algorithmic Grouping scenario such that the algorithmic tool generates groups of data *after* a human scholar has already engaged with inductive coding. Through this new context, our participants speculated how they could use AI to provide new perspectives and identify gaps in their own analysis.

Several participants, including P4, described how "*it might be interesting to compare and contrast*" codes derived by human scholars with those produced by AI. In prior work, Muller et al. theorize how AI-generated groups can provide human scholars with new elements or representations of data through which to compare codes and other interpretations [33]. By providing groups or clusters as points of comparison, the initial development of codes is no longer outsourced or delegated to AI. Rather, when scholars have the opportunity to first develop their codes, the role of AI shifts from primary analyst or instrument to collaborator.

When human analysts are presented with AI-generated groups following their own foray into inductive coding, they are provided an opportunity to identify potential gaps in their own analysis. P9 imagined how AI-generated groups could help her "*see other connections that [she] might not have seen before*" – but, she cautioned, only after she has "really good sense of what the data is." P9 made it crystal clear that the primacy of her interpretation and the order of AI involvement were paramount. She does not want her analysis biased or influenced by patterns generated by a machine. Any insights drawn from a comparison between her analysis and work an AI might perform are secondary to her analytic process and, ultimately, should serve to deepen her understanding of the data.

Other scholars imagined a more middle-of-the-road system, where AI coded with them, rather than for them or waiting until after their coding was complete. P5 was intrigued by the idea of AI that could develop additional codes based on the segments of text she had already coded. She explained, "*If the machine could also suggest, based off of the labeling that I have done so far, what the other labels could potentially be. I think that will be interesting. [...] It's suggesting labels at this point like, 'Okay, so you've done this, this, this, and this. Would you like to...?'*" Rather than work with AI that generates groups from the data, P5 imagined a system that is informed by human analysis and where additional codes are co-constructed between the analytic work a human has performed and suggestions by an algorithmic tool. Here, we see how AI can assist by suggesting new codes, thereby extending work that a scholar has started to perform.

The findings in this section illustrate the importance of considering how and when AI can assist in qualitative coding. As we see above and return to in our discussion, there are opportunities to design AI for a variety of approaches, including those grounded in data and those informed by human analysis. Though individual participants had clear preferences related to the order in which they would include AI in their own qualitative work, these preferences are by no means uniform across the whole. This division points to a clear need to develop and deploy AI assistants

for different ways of performing qualitative analysis — a limitation of this study, though something we hope to see in future work.

5.3 Using AI to Guide and Assist Exploration

In this section, we attend to how the pattern-making prowess of AI can inform sampling in powerful ways, including addressing human limitations (e.g., consider a single scholar endeavoring to read a million tweets) and moving beyond the limitations of a random sample, a common approach for qualitative scholars working with big data. As with any approach or collaboration, there are tradeoffs. Participants cautioned several drawbacks related to using algorithmically generated patterns, including concerns related to the partial perspective of AI as well as AI's tendency to highlight the center (e.g., common co-occurrences) of datasets over the edges.

5.3.1 Supporting Sampling through AI-based Overviews. Our participants talked about numerous strategic approaches for sampling from large corpora, including strategies that involved and excluded AI. Subsampling data influences how analysis proceeds. For example, consider the difference between a random sample and one informed by the interests or earlier analytic work of a scholar (e.g., purposive sampling, theoretical sampling). Our participants discussed examples from their own research as well as our scenarios related to using AI to sample data. We find that using AI for sampling can provide opportunities to leverage insights from the data to inform analysis, but that these sampling strategies are not without constraints and limitations.

P6 talked about using semantic network analysis to strategically subsample from a large dataset of tweets. As we described in the first section, semantic network analysis works by generating clusters of frequently co-occurring terms. Using the terms within each cluster, P6 manually searched their large corpora for texts where these terms occurred. According to P6, working with semantic network analysis facilitated moving “*beyond the limitations of a random sample,*” including lack of representativeness and data missed or excluded (e.g. such that it could lead to “*questions to ask that you don't notice*”).

Using AI overviews to guide sampling shifts where and how abstraction occurs in qualitative analysis. Calling back to the codes-to-theory model, Saldaña illustrates how abstraction results as coding progresses (i.e., from the organization of codes and eventual transcendence of categories). Though many qualitative scholars engage in a back and forth between codes, categories, and themes, moving in non-linear ways between the real and the abstract, traditional approaches to inductive qualitative analysis typically start with the real (i.e., with actual examples of data). When AI overviews, such as through semantic network analysis, enter the picture, scholars may bypass starting with the data and, instead, start with abstract descriptions (e.g., clusters, topics); therefore, moving from the abstract (e.g., clusters) to the real (e.g., data) at the start of an analytic process.

Within our sample, P6's use of semantic network analysis is unique. Many of the other scholars we spoke with used random or chronological (i.e., in the order the data was collected, based on a timestamp) strategies to subsample from large datasets. Though not all of these strategies involved AI, they did involve a variety of computational tools, including Excel and Python. The use of these tools illustrates how data sampling in qualitative analysis of large corpora is frequently delegated to non-human actors. Nevertheless, our participants preferred certain ways of delegating over others. When we presented our participants with our Algorithmic Sampling scenario (see Figure 3), where AI-generated groups are used to identify specific texts, several individuals addressed how this sampling strategy could unduly guide or influence what they found interesting within a dataset. We see here echoes of our previous sections, where participants raised concerns about AI-informed code development.

When used to guide sampling, AI-generated groups and overviews can support scholars in sampling representatively from their dataset, rather than randomly. However, different techniques have different consequences for analysis. We turn to tradeoffs with AI sampling at the end of this section.

5.3.2 Using Human Coding to Inform AI Sampling. Participants discussed how their coding practices and outcomes could be used to inform AI for data sampling (e.g., such as by classifying texts as relevant or not for certain categories). In one example, P11 and her collaborators developed machine learning classifiers to identify examples within a 30,000-text dataset. The classifiers P11 and her collaborators built were informed by the categories that P11 had started to develop through inductive coding. As explained by P11, each category had an associated list of words. These words, as well as semantically similar words (as produced by a computational linguistic tool), were used to develop classifiers to identify relevant texts from the dataset. Without AI, collecting additional samples “*would have taken a long time to do... by hand.*”

However, as with all decisions, there are tradeoffs to consider when using AI. Though P11 implied that AI saved time, in the same breath she mentioned:

[W]e thought we were saving time, but then I don't know if we did save time because we kept iterating and iterating through [the development of the classifiers]. But at least in the future if we have more [data], we could return to this classifier as opposed to building a new one completely from scratch.

Some prior work describes the potential for AI to reduce the time and labor intensity of qualitative analysis. However, P11's example provides another point of view: AI may shift where and how labor occurs within qualitative data analysis.

P14 discussed another drawback. The use of AI informed by human coding (and human selected examples) to sample data introduces challenges to extrapolating beyond examples provided in the training dataset. He said:

[Based on training data], the algorithm is going to think that particular words are associated with this code. And, so, let's say another data point is talking about the same thing, but using different words. The machine is not going to be as likely to think that that code applies even though it very well might.

P11 aimed to mitigate this limitation by using a computational linguistic tool to generate similar words from the list of words-in-categories provided by human scholars. However, this raised questions about what the computational tool considered ‘linguistically similar,’ as these connections were not obvious or clear. P11 grappled with this tradeoff, ultimately deciding it was worthwhile to avoid manually searching the dataset for examples.

In the previous section, we described our participants' concerns related to our Algorithmic Sampling scenario, where data sampling could be informed by algorithmically generated clusters. When modifying this scenario to be informed by human coding (e.g., human coding informs how AI identifies additional data), our participants saw opportunities for AI to support their interests and save them from some of the monotony associated with sampling. P9 explained:

I think if it's true machine learning, where it's figuring out what I'm doing and supporting things I'm already doing, then I'm all for it. As long as I can check that after the fact to make sure it actually did what I wanted, but I don't actually hate that idea. Because, for example, if it's noticing that I'm pulling all these different tweets or whatever that have those same kind of phrases over and over again, then I would love it to just do that for me. That would be great. So, I don't have to keep doing it, especially as things get repetitive.

Algorithmic tools grounded in or that work with human analytic practices, such as coding, invite the potential for human analysts to lead and guide the direction of human-AI collaboration. Rather than suggest directions for sampling, though this may be preferable for some, AI can augment and potentially extend how a scholar is already moving through a dataset. Importantly, when sampling is assisted by AI that is informed by human analysis, scholars, such as P9, may want or expect mechanisms that support evaluating what the AI is doing. In addition to the necessity of transparency within AI systems, in the next section we illustrate how vital it is for qualitative scholars to understand the strengths and limitations of their AI collaborators.

5.3.3 The Limitations of High Frequency Counts. Semantic network analysis has made several appearances throughout our findings (e.g., P1, P6). Other types of AI, including topic modeling [27], use similar approaches, grouping terms based on some element related to frequency (e.g., such as terms within a document). Many approaches to qualitative analysis do not necessitate counting codes or categories or working with quantitative metrics to interpret data. However, when scholars work with AI to sample data, the particular ways a tool works become inscribed in how analysis proceeds, impacting the types of qualitative insights that can eventually be derived. In considering how many AI approaches work, our participants cautioned against becoming too reliant on high frequency counts within the data at the cost of attending to other things.

Several participants, including P2, P8, and P15, addressed how their qualitative analysis included – and at times even focused on – the edge cases within their datasets. P8 described a time that only one piece of data (out of numerous collected from an online forum) included a creative expression (e.g., a work of art) related to the topic of study. This example made its way into subsequent publication as a vignette intended to help capture the nuance and emotional complexity of the domain space. AI that attends to frequent occurrences can miss significant instances where $n = 1$. Examples such as these are important in qualitative scholarship for the depth they provide and the ways they help scholars understand the boundaries of their dataset.

Taking a balanced approach, P2 talked about the importance of gaining a sense of the centers and edges of datasets. She discussed how AI could provide low-frequency counts as well as the overall shape of the dataset, saying:

Low-frequency counts of things would also be so helpful because if you have 800 things, to find the thing that is the least frequent among them is a fucking ask. But if a machine does that for you, it would be cool to actually get a better spread of that, to get a view of the data central frequencies, central tendencies, so central measures, but also the tails and the skews and the shapes of the data, not just accuracy.

While using algorithmic tools to help identify low-frequency counts and the potential ‘edges’ of a dataset is possible, it also raises questions about how these tools define ‘centers’ and ‘edges.’ Likewise, it raises questions about whether the system and human share these definitions.

Underlying how AI develops patterns is a larger conversation related to the ‘perspective’ or ‘lens’ of a particular tool. AI presents a particular way of viewing data at the expense of other ways of seeing and knowing [14, 23]. Using a metaphor of doors, P15 explained:

It’s like, there’s multiple doors to get into a space and then if you go through this one door, then you see a different view... you see your experiences different. So, I think that’s how I think about it when using [AI] as exploration. Sure, it can be giving you that one view, but it wouldn’t give me necessarily the whole perspective.

Importantly, AI always provides a partial perspective, even when it provides overviews of a dataset. Like others, we argue for careful consideration related to when and how AI might supplement human analysis or come to be delegated tasks during qualitative analysis. In this

section, we saw how AI could simultaneously provide overviews of patterns within datasets while constraining scholar exploration to clusters grounded in frequently co-occurring terms. These limitations raise questions about the utility of algorithmic tools, which we return to in our discussion.

6 DISCUSSION

Our findings illustrate how the stage of qualitative analysis impacts how AI might be used by and alongside human analysts. In the following sections, we extend Jiang et al.'s discussion of human-AI collaboration in qualitative research and start to consider what it might mean to enter into "synergetic partnerships" with AI [24, 25]. Given concerns surrounding displacement — as illustrated by Jiang et al. [25] and echoed here — we draw inspiration from human-machine symbiosis in order to imagine how AI systems might augment and support human cognition and skill throughout qualitative data analysis [24, 25, 48, 50], rather than replacing them.

6.1 Shifts to Qualitative Analysis in Human-AI Collaboration

Shifts in practices of qualitative data analysis that arise from the use of AI may be subtle or extreme. Recall, for example, the differing ways P1 and P7 used AI as part of their analytic process. Through P1's use of semantic network analysis (i.e., to inform the development of codes at the start of inductive coding), the AI became a conduit for analysis, a way of seeing and making sense of the data. P1 worked from results provided by semantic network analysis to inductively generate a framework for viewing her large dataset. Here, the use of AI is a stark departure from traditional qualitative analysis, where human scholars are primarily responsible for instrumentation [29]. P7, in contrast, used AI like a highlighter, calling attention to words within texts that were relevant to the research inquiry (i.e., care). The ways P1 and P7 used AI begin to illustrate the diversity of tools involved in this domain as well as the tasks they can support.

Like with any tool or methodological approach, there are tradeoffs to using AI as part of qualitative analysis. For example, it is not possible, or even reasonable, for P1 to consider reading hundreds of thousands of tweets. Using AI allowed her to look across the entirety of a large corpus and gain an understanding, an overview, of what was going on. However, this benefit is balanced with a limitation: external influence by a tool. When scholars, like P1, decide to work with AI at the start of inductive coding — specifically, to develop codes — they rely on how the AI 'sees' or represents data. This 'algorithmic way of looking' is then carried through the analytic process. Of course, the tradeoffs associated with using AI tools are not always so extreme. P7's use of AI resulted in more moderate tradeoffs. By capitalizing certain words, AI called attention to these areas of the data over others, thus helping P7 and her team attend to examples related to care, but at the cost of potentially missing out on less common articulations of that topic.

Examining how AI can shift practices of qualitative data analysis is integral to understanding how to build and design for AI tools as well as human-AI collaborations and partnerships. In the following sections, we address three areas where we see potential shifts to qualitative analysis: scale, abstraction, and task delegation. We argue for understanding the potential impact of AI during different stages of qualitative analysis, mapping out the strengths and limitations of these tools as a way to imagine and inform potential futures of qualitative work.

6.1.1 Qualitative Analysis at Scale. Qualitative scholars typically work with small amounts of data: tens of interviews or a few hundred survey responses. While this amount of data can still be overwhelming for analysts [25], it's but a drop in the ocean compared with the quantity of data available in big data corpora. We echo others in arguing that AI can support qualitative scholars in increasing the scale of their analysis [4, 25, 49]. However, we share the cautions of Jiang et

al. [25] and the concerns of our participants with respect to the limitations of AI in this context. When aiming to generate insights at a large scale, there can be the risk of leaning on descriptive or numerical insights (e.g., quantitative), rather than ones that help us understand the depth and richness of data (e.g., qualitative). Here we consider what scale means for qualitative analysis, focusing on data sampling and coding.

The diverse shifts to sampling with AI that we see across our findings present potential departures from traditional qualitative sampling strategies, like purposive sampling [42]. Recall, for example, how P6 used terms (within clusters) derived from semantic network analysis to search for examples within their dataset. Using semantic analysis to provide an overview, as P6 did, aligns with Shneiderman's classic visualization mantra: "overview first, zoom and filter, then details on demand" [45]. However, such an approach relies on out-of-the-box computational methods for providing an overview. In contrast, P11 used classifiers grounded in categories she had inductively developed to identify specific types of data within her corpus. P11's use of AI is informed by human analysis and, in some ways, is similar to purposive sampling, where data related to specific topics, concepts, and phenomena are strategically and intentionally collected.

Beyond sampling, AI can extend and augment the ways that qualitative scholars code data. Rather than use terms from AI-developed clusters to sample from datasets, scholars may use them to inductively generate codes and categories that can act as frameworks for analysis or as points of comparison following coding by human analysts. We also saw opportunities for AI to assist human analysts in coding across the entirety of large datasets. We suggest that (partially) automated coding has the most benefits for scholars who standardize and calcify their codes relatively early on during qualitative coding [32, 39, 40, 49] (i.e., 'relatively' when compared with the 'standardization' that, out of necessity, occurs at the very latest during manuscript writing). For scholars whose coding process is not infinitely looping and flexible, AI-assisted coding might reduce the manual, repetitive labor associated with coding and make it possible to apply codes to an entire corpus rather than a subsample [18].

In considering how AI can facilitate increasing the scale of qualitative research, we're left with several open questions related to data sampling and coding. With sampling, we might wonder about the ways that AI overviews call attention to certain topics or areas of the data and not others. Additionally, even when informed by human analysis, such as in P11's example, we might aim to critically interrogate how training classifiers makes some expressions and patterns associated with a topic more visible than others. The use of AI to automate coding raises questions related to applying codes developed within one dataset to another. Here we pause to ask, what is the qualitative benefit? Application of codes across datasets may be more representative of generalizability, which is not necessarily a goal associated with qualitative scholarship. In potentially working with AI at scale to sample and code data, we need to foreground and privilege qualitative ways of knowing the world. The exact ways of doing this are uncertainties that future work can and should explore.

6.1.2 AI Approaches to Abstracting Data. In the codes-to-theory model, abstraction results from coding, where scholars start with data and develop categories and themes as analysis proceeds [44]. Though scholars ultimately move toward higher levels of abstraction, coding commonly involves a back and forth between data, codes, categories, and themes. This back and forth is highly manual, though often supported by computational tools (e.g., Ctrl-F and search, QDA software, Word documents, and spreadsheets) and collaborators [25]. Based on our analysis, we argue that AI can impact when and how abstraction occurs during qualitative data analysis.

AI can provide patterns or overviews (i.e., abstracted groupings) of large corpora. For example, several participants, including P1 and P6, used AI-generated overviews, such as semantic network analysis, prior to or at the start of inductive coding. This early access to abstractions of data differs

from traditional qualitative approaches, which do not account for having an overview of the corpus prior to human analysis. Specifically, though human analysts may traverse between the real and the abstract throughout inductive coding, abstractions of data are not the starting point. As we illustrate, however, working with AI overviews during the early stages of qualitative analysis provides benefits, including opportunities to representatively sample data and inductively generate codes.

Despite benefits, working with abstractions produced by AI is not without limitations. P2 explained how overviews of data produced by AI typically rely on high frequency counts to generate patterns. Privileging frequency is not integral to many qualitative methodologies, and several of our scholars described a vested interest in the edges of their datasets and infrequent-yet-still-significant occurrences. While recognizing how AI privileges certain elements within a dataset is vital, qualitative scholars may not have the training to be aware of or to interrogate the types of properties they are elevating through the use of AI in analysis. In the context of frequency, this creates a risk in which working with AI could mean missing edge cases in a dataset, an oversight that is largely antithetical to how some scholars conduct research. On the other hand, designing AI to support qualitative scholarship could involve developing tools to identify those edge cases, an example we return to in our design recommendations.

6.1.3 Delegation at Different Stages of Qualitative Analysis. Qualitative scholars do not want to delegate the entire ‘task of qualitative analysis’ to AI or be displaced as the primary analysts in their work [25]. While our study echoes much of Jiang et al.’s findings, we build off this prior work by demonstrating how concerns around task delegation and displacement coincide with particular stages and types of qualitative work. We argue that considering the stage of qualitative analysis is key to considering when, how, and whether to introduce algorithmic tools.

Many of the scholars we spoke with were resistant to AI presenting suggestions, clusters, or any type of information at the start of qualitative analysis, including during data exploration and initial coding. Data exploration and coding are practices through which scholars come to know their data. When AI is involved in this work, it creates pressures toward certain ways of seeing or knowing the data (e.g., such as by visualizing the data through certain clusters). Many of our participants felt this type of influence could bias their analysis in unanticipated ways. This suggests that, for many scholars, AI might be better incorporated after humans have started to make analytic progress.

When we consider how AI might look during later stages of qualitative analysis, such as after categories have been developed, we see opportunities for AI to assist without displacing scholars or potentially stepping on their toes. AI might be involved as a tool, collaborator, or partner, rather than the primary instrument for analysis. For example, AI might assist by automating work associated with coding, by suggesting new codes (based on the work scholars have performed), and by identifying gaps in human analysis. In these examples, we see how AI can highlight different ways of looking at data, similar to the work a human collaborator might perform.

We caution, however, that our suggestions are inspirational rather than prescriptive. Several of our participants provided counterpoints to the above considerations of timing. For example, P1 and P6 found that working with AI at the start of qualitative analysis helped them gain a better sense of what was going on within their dataset, which influenced their analytic focus. As with our other sections, our intent is to illustrate the complexity and diversity associated with using AI in qualitative analysis. Ultimately, this complexity indicates a need for flexible, responsive systems. Just as there is no one particular way to perform inductive qualitative analysis, there is no one-size-fits-all for designing AI.

6.2 Designing for Human-Machine Collaboration in Qualitative Analysis

We find that all stages of qualitative analysis are amenable to algorithmic support. What remains an open question is the types of support AI should provide across these stages, and how this support impacts the interpretive process. In the following sections, we detail three high-level considerations for technologists interested in building algorithmic systems for qualitative data analysis: the nature of the interaction, the balance between labor and insight, and the ways data and concepts are privileged (or not) during analysis.

6.2.1 From Displacement to Collaboration. Rather than delegate qualitative analysis to AI, qualitative scholars should remain the primary analytic lens. Accordingly, it is important to consider the types of collaboration that future systems could enable. As we have discussed, participants largely found the idea of working with AI during sampling and coding inappropriate until after they had developed their own interests and insights. However, once participants started analysis, they were open to AI providing results, such as clusters, as points of comparison for their own codes and categories. Given these findings, we caution against any AI that acts as an “eager” collaborator and that might displace the scholar as the primary conduit of analysis. Eager AI is particularly risky given the diverse paths that qualitative scholars take as they move from data to abstract claims. Acknowledging the myriad ways that AI can support qualitative analysis requires abandoning the wish to design for any particular ‘pipeline’ or process. Scholars want algorithmic tools that fit their particular practices. Incorporating AI when the analyst is ready is one way to address this need.

In most cases, the ideal human-AI collaborations for our participants would be deeply interactive, producing opportunities to design symbiotic partnerships with AI. For example, using active learning techniques means that AI can tailor results to work a human has performed [13]. We see potential in developing algorithmic tools that continuously learn from the work that qualitative analysts perform. Partnering with algorithmic tools in this way would most likely involve the adoption of active learning approaches to AI, where the work an algorithmic tool conducts is grounded in what a human scholar has already performed (e.g., coding). Ideal tools for qualitative research would not only require initial human input, but also human evaluation of what an algorithmic tool has done. As we address in the next section, however, we need to be cognizant of introducing new types of work into qualitative analysis, which is already labor-intensive.

6.2.2 Extra Work vs. Extra Insight. Prior work frequently describes how the inclusion of AI into qualitative analysis can reduce the time and labor of traditional qualitative work. After all, the use of AI in qualitative analysis can expand the scale at which analysis occurs. A machine synthesizing a million tweets is certainly a much quicker and more reasonable task than if assigned to or carried out by a human analyst. Nevertheless, our findings also show how using AI can create labor for qualitative analysts, some of which may be unanticipated. Several participants, including P1, P8, and P11, were surprised by the labor associated with the algorithmic approaches they used — whether this involved the necessary tuning and evaluation of machine learning classifiers, or the time needed to search through a corpus for specific data related to machine-generated clusters. The unexpected labor associated with algorithmic approaches indicates a need to modify the expectations of qualitative scholars using algorithmic approaches and to design transparent AI tools [46].

Qualitative analysis is a slow, labor-intensive, and often conceptually messy process. It would be a misstep to expect AI to “solve” these challenges. Analysts who include, or who are thinking about including, AI in some part of their practice may do so with preconceived notions that these tools will make their work somehow quicker or easier by transforming how they approach qualitative analysis (e.g., including AI may enable them to work at larger scales of data, ask new questions,

and solicit different perspectives). When using AI during qualitative analysis still results in a long and challenging process, as it did for several of our participants, scholars might feel misled by their vision of human-AI collaboration. Part of designing AI for use in qualitative analysis involves training human analysts to work with these tools. Understanding what AI does and how it works sets up realistic expectations and supports qualitative scholars in appropriately accounting for the time associated with using AI.

Though AI may not drastically reduce the time associated with qualitative analysis, tools should avoid generating unnecessary labor. One way to think about designing to avoid unnecessary labor on behalf of the scholar involves designing for transparency. One type of transparency we address here relates to linking the abstract results of AI with the real, particular examples from the data. For example, imagine a semantic network analysis algorithm that provides clusters of frequently co-occurring terms as well as examples of data that could belong in each cluster. Rather than search through the entire dataset for specific texts with those terms, examples would be provided. The human analyst could use their energy to evaluate the examples and work to train and refine the clusters as needed, rather than searching the dataset. This back and forth between machine abstractions and real data, though still manual and involving the analyst, becomes a little less time-intensive and better supported.

6.2.3 Focusing the Analytic Lens. Finally, the ways that AI synthesizes or abstracts data tend to privilege certain representations and data over others. Across our interviews we heard how algorithmic tools privilege high-frequency counts and examples included in training data. However, qualitative scholars are not necessarily looking for commonalities or frequencies within a dataset or examples exactly similar to those they have already coded. Though there is a risk that algorithmic approaches may focus on the wrong information or elements of data, there are ways that algorithmic tools can augment and extend analysis (rather than exactly replicate it), such as by highlighting gaps in qualitative analysis performed by human scholars.

Several of our scholars discussed the importance of edge cases and one-off examples within their work. Less common occurrences help scholars map the conceptual space and often provide valuable counterexamples to normative trends within a dataset. Using algorithmic tools that focus on high-frequency counts may mean inadvertently ignoring less common or less machine recognizable patterns (e.g., such as an important ‘pattern of one’ example). Technologists developing AI for qualitative analysis could use algorithmic approaches to highlight the edges, infrequent patterns, or “rare anomalies” [1, 2]. For example, work by Ahmed and Ullah details a novel approach to clustering that identifies infrequent patterns in smart healthcare environments [2].

For qualitative scholars working with algorithmic tools, including a variety of AI techniques within a system can facilitate critical and reflective questioning of the dataset and analytic process. For example, providing the spread of data (e.g., an overview of what is frequent and what is not), rather than simply highlighting algorithmically determined centers, supports qualitative scholars in examining the impact that their own subjectivity has during analysis (e.g., where they have and have not focused during analysis). By viewing points of comparison, scholars may be able to ask questions about the differences between their categories and those developed by AI and about the relationships between AI-generated clusters or groups.

Technical strategies for doing all of the above can be found in an active learning paradigm. In that paradigm, diversity sampling can identify data points that the model does not understand [22], and uncertainty sampling can identify data points about which the model is unclear how to classify (i.e., on the border between two different classes) [52]. However, we want to highlight that the ways qualitative scholars think about their data do not align with how machines think about data. For example, qualitative scholars rarely think about their data as a scatter plot with a decision

boundary between two classes. One of the more striking aspects for us as scholars in this space has been the significant gap between how qualitative scholars and those in AI discuss and think about data. This suggests that there are important challenges that fields like information visualization and human-centered AI need to address when it comes to bridging that gap.

6.3 Limitations

This study has several limitations. First, we conducted interviews exclusively with CSCW and HCI qualitative scholars. Within CSCW and HCI, there is a focus on machines (and interactions with them) that makes total unfamiliarity with machine learning unlikely. Indeed, all of our participants were familiar with machine learning and how it could be used in qualitative analysis. As our findings illustrate, some of our participants had even incorporated machine learning into their own work. The familiarity with machine learning and AI that our participants demonstrated may not be shared in disciplines that are less computer forward. Future work examining how to design AI for qualitative analysis should recruit from a more diverse representation of academic fields, including those not so closely tied to computing.

Second, in this study we were interested in speaking with participants who worked with short texts, such as those collected from social media platforms. We intentionally excluded qualitative work with longer texts (e.g., interviews) from our recruitment call. Longer texts are typically associated with smaller corpora. For example, scholars conducting interviews in CSCW and HCI tend to work with approximately 15 to 30 transcripts — nowhere near the hundreds or thousands of texts in large datasets. As such, for this study, we prioritized short texts (in large corpora) as a way of working with scholars who had experience analyzing datasets they had not been able to read in their entirety. Given this limitation of our study, we cannot speak to how qualitative scholars might consider the adoption of machine learning in the analysis of large texts.

Third, the scenarios we presented to our participants were grounded in the early stages of inductive qualitative analysis, including sampling and coding. Accordingly, this is where conversations with our participants largely remained. There were very few instances in our data where participants described using AI to support the comparison of categories and eventual development of themes. Looking across related work in this domain, we see this as a limitation associated with how we as scholars are developing AI for qualitative data analysis. Future work warrants investigating AI in the context of thematic development.

Finally, while our scenarios helped us attend to the different stages of qualitative analysis, they are devoid of the broader context of how qualitative analysis is performed. This means that they do not provide a clear sense of how using AI at one stage of qualitative analysis cascades or impacts others. To this end, future work that deploys AI throughout entire qualitative analysis projects could help scholars understand and examine how AI influences the ways that qualitative analysis is performed as well as how insights and abstract claims are generated. Using AI with particular scholars and projects could also provide opportunities to explore how to customize AI for the particular work that individuals are doing.

7 CONCLUSION

The amount of data currently available and used in research is staggering — and it's only growing. AI is commonly used to assist in the analysis of large corpora; however, the use of these tools is frequently associated with quantitative ways of knowing. There is a clear need to make large corpora accessible to qualitative scholarship, to complement descriptive and numerical ways of knowing the world with rich and nuanced interpretations. In this study, we examined how AI might come to be used at different stages of qualitative data analysis. We focus on data exploration and coding, describing the diverse ways in which our participants have used — or imagined using

– AI at different steps of their analytic process. While some participants prioritized the primacy of their analysis and the order of AI involvement, others preferred intimate partnerships with these non-human actors. These differences point to the need for flexibility and customization in the design of AI for qualitative analysis. We echo other scholars when we argue that tools in this context cannot be one-size-fits-all. Further, we extend prior work by detailing three ways that using AI systems may result in shifts to practices of qualitative analysis: scale, abstraction, and delegation. Designing for these shifts means recognizing and working with the strengths and limitations of AI, as well as human analysts, to create algorithmic systems that can support and augment the work of qualitative scholars.

ACKNOWLEDGMENTS

We thank our participants, CSCW and HCI researchers who let us ask about their analytic practices and imagine AI futures in qualitative research. We also thank Danielle Albers Szafir, Casey Fiesler, Matt-Heun Hong, and Kandrea Wade for their feedback and support throughout the entirety of this study. This work was supported by National Science Foundation Award #1764089.

REFERENCES

- [1] Mohiuddin Ahmed. 2019. Intelligent big data summarization for rare anomaly detection. *IEEE Access* 7 (2019), 68669–68677.
- [2] Mohiuddin Ahmed and Abu SSM Barkat Ullah. 2018. Infrequent pattern mining in smart healthcare environment using data summarization. *The Journal of Supercomputing* 74, 10 (2018), 5041–5059.
- [3] Eric PS Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1397–1410.
- [4] Eric PS Baumer, Drew Siedel, Lena McDonnell, Jiayun Zhong, Patricia Sittikul, and Micki McGee. 2020. Topicalizer: reframing core concepts in machine learning visualization by co-designing for interpretivist scholarship. *Human-Computer Interaction* 35, 5-6 (2020), 452–480.
- [5] Anol Bhattacharjee. 2012. Social science research: Principles, methods, and practices. (2012).
- [6] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1 (2016), 1–115.
- [7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [8] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24.
- [9] John M Carrol. 1999. Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 11–pp.
- [10] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
- [11] Nan-Chen Chen, Michael Brooks, Rafal Kocielnik, Sungsoo Ray Hong, Jeff Smith, Sanny Lin, Zening Qu, and Cecilia Aragon. 2017. Lariat: A visual analytics tool for social media researchers to explore Twitter datasets. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- [12] Nan-Chen Chen, Margaret Drouhard, Rafal Kocielnik, Jina Suh, and Cecilia R Aragon. 2018. Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 2 (2018), 1–20.
- [13] Janghee Cho and Emilee Rader. 2020. The Role of Conversational Grounding in Supporting Symbiosis Between People and Digital Assistants. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.
- [14] Joseph E Davis. 2019. Toward the Elimination of Subjectivity: From Francis Bacon to AI. *Social Research: An International Quarterly* 86, 4 (2019), 845–869.
- [15] Philipp Drieger. 2013. Semantic network analysis as a method for visual text analytics. *Procedia-social and behavioral sciences* 79 (2013), 4–17.
- [16] Margaret Drouhard, Nan-Chen Chen, Jina Suh, Rafal Kocielnik, Vanessa Pena-Araya, Keting Cen, Xiangyi Zheng, and Cecilia R Aragon. 2017. Aeonium: Visual analytics to support collaborative qualitative coding. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 220–229.

- [17] Alicia Eads, Alexandra Schofield, Fauna Mahootian, David Mimno, and Rens Wilderom. 2021. Separating the wheat from the chaff: A topic and keyword-based procedure for identifying research-relevant text*. *Poetics* 86 (2021), 101527.
- [18] Casey Fiesler, Jialun Jiang, Joshua McCann, Kyle Frye, and Jed Brubaker. 2018. Reddit rules! characterizing an ecosystem of governance. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.
- [19] Alina Gerber, Patrick Derckx, Daniel A Döppner, and Detlef Schoder. 2020. Conceptualization of the Human-Machine Symbiosis—A Literature Review. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
- [20] Barney G Glaser. 1965. The constant comparative method of qualitative analysis. *Social problems* 12, 4 (1965), 436–445.
- [21] Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
- [22] Zhiqiang Gong, Ping Zhong, and Weidong Hu. 2019. Diversity in machine learning. *IEEE Access* 7 (2019), 64323–64350.
- [23] Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies* 14, 3 (1988), 575–599.
- [24] Mohammad Hossein Jarrahi. 2018. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons* 61, 4 (2018), 577–586.
- [25] Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R Brubaker. 2021. Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis. *arXiv preprint arXiv:2102.03702* (2021).
- [26] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.
- [27] Loet Leydesdorff and Adina Nerghe. 2017. Co-word maps and topic modeling: A comparison using small and medium-sized corpora (N< 1,000). *Journal of the Association for Information Science and Technology* 68, 4 (2017), 1024–1035.
- [28] Joseph CR Licklider. 1960. Man-computer symbiosis. *IRE transactions on human factors in electronics* 1 (1960), 4–11.
- [29] John Lofland and Lyn H Lofland. 1971. Analyzing social settings. (1971).
- [30] Brian Lubars and Chenhao Tan. 2019. Ask not what AI can do, but what AI should do: Towards a framework of task delegability. *arXiv preprint arXiv:1902.03245* (2019).
- [31] Maximilian Mackeprang, Claudia Müller-Birn, and Maximilian Timo Stauss. 2019. Discovering the Sweet Spot of Human-Computer Configurations: A Case Study in Information Extraction. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
- [32] Megh Marathe and Kentaro Toyama. 2018. Semi-automated coding for qualitative research: A user-centered inquiry and initial prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [33] Michael Muller, Shion Guha, Eric PS Baumer, David Mimno, and N Sadat Shami. 2016. Machine learning and grounded theory method: convergence, divergence, and combination. In *Proceedings of the 19th international conference on supporting group work*. 3–8.
- [34] Katashi Nagao. 2019. Symbiosis between humans and artificial intelligence. In *Artificial Intelligence Accelerates Human Learning*. Springer, 135–151.
- [35] Laura K Nelson. 2020. Computational grounded theory: A methodological framework. *Sociological Methods & Research* 49, 1 (2020), 3–42.
- [36] Laura K Nelson. 2021. Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century US South. *Poetics* (2021), 101539.
- [37] Pablo Paredes, Ana Rufino Ferreira, Cory Schillaci, Gene Yoo, Pierre Karashchuk, Dennis Xing, Coye Cheshire, and John Canny. 2017. Inquire: Large-scale early insight discovery for qualitative research. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1562–1575.
- [38] Desmond U Patton, William R Frey, Kyle A McGregor, Fei-Tzin Lee, Kathleen McKeown, and Emanuel Moss. 2020. Contextual analysis of social media: The promise and challenge of eliciting context in social media posts with natural language processing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 337–342.
- [39] Tim Rietz and Alexander Maedche. 2020. Towards the Design of an Interactive Machine Learning System for Qualitative Coding. In *Proceedings of the Forty-First International Conference on Information Systems*.
- [40] Tim Rietz, Peyman Toreini, and Alexander Maedche. 2020. Cody: An Interactive Machine Learning System for Qualitative Coding. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 90–92.
- [41] Oliver C Robinson. 2014. Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology* 11, 1 (2014), 25–41.
- [42] Rebecca S Robinson. 2014. Purposive sampling. *Encyclopedia of quality of life and well-being research* 6 (2014), 5243–5245.
- [43] Maria Y Rodriguez and Heather Storer. 2020. A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services* 38, 1 (2020), 54–86.
- [44] Johnny Saldaña. 2009. *The coding manual for qualitative researchers*. SAGE Publications.

- [45] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 364–371.
- [46] Aaron Springer. 2019. Enabling Effective Transparency: Towards User-Centric Intelligent Systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 543–544.
- [47] Anselm Strauss and Juliet M Corbin. 1997. *Grounded theory in practice*. Sage.
- [48] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists’ perceptions of automated AI. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.
- [49] Jasy Liew Suet Yan, Nancy McCracken, Shichun Zhou, and Kevin Crowston. 2014. Optimizing features in active machine learning for complex qualitative content analysis. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*. 44–48.
- [50] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable ai: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [51] Himanshu Zade, Margaret Drouhard, Bonnie Chinh, Lu Gan, and Cecilia Aragon. 2018. Conceptualizing disagreement in qualitative coding. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [52] Jingbo Zhu, Huizhen Wang, Benjamin K Tsou, and Matthew Ma. 2009. Active learning with sampling by uncertainty and density for data annotations. *IEEE Transactions on audio, speech, and language processing* 18, 6 (2009), 1323–1331.

Received April 2021; revised July 2021; accepted July 2021