Algorithmic Subjectivities

ERIC P. S. BAUMER, Lehigh University, USA
ALEX S. TAYLOR, University of Edinburgh, Scotland
JED R. BRUBAKER, University of Colorado Boulder, USA
MICKI MCGEE, Fordham University, USA

This paper considers how subjectivities are enlivened in algorithmic systems. We first review related literature to clarify how we see "subjectivities" as emerging through a tangled web of processes and actors. We then offer two case studies exemplifying the emergence of algorithmic subjectivities: one involving computational topic modeling of blogs written by parents with children on the autism spectrum, and one involving algorithmic moderation of social media content. Drawing on these case studies, we then articulate a series of qualities that characterizes algorithmic subjectivities. We also compare and contrast these qualities with a number of related concepts from prior literature to articulate how algorithmic subjectivities constitutes a novel theoretical contribution, as well as how it offers a focal lens for future empirical investigation and for design. In short, this paper points out how certain worlds are being made and/or being made possible via algorithmic systems, and it asks HCI to consider what other worlds might be possible.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models**; **Interaction design theory, concepts and paradigms**.

Additional Key Words and Phrases: Subjectivity, algorithms, reflective HCI

1 INTRODUCTION

Numerous interactive technologies now incorporate sophisticated machine learning (ML) and/or artificial intelligence (AI). Examples range from sentiment analysis of online reviews [199], to social media news feed curation [71], to algorithmically targeted advertising [145, 188]. Many of these technologies are based on automated inferences about attributes of individual users, including age, race, gender, friendship networks, media preferences, political views, religious beliefs, and others [97]. An emerging body of work is examining the kinds of interactions and experiences to which these systems give rise, as well as how to design such systems [5, 10, 16, 36, 48, 61, 71, 72, 80].

This paper offers a theoretical contribution to that growing literature by considering how subjectivities are done through algorithms. Specifically, it proposes **algorithmic subjectivities**, an approach that encourages us to acknowledge, to attend to, and to account for the ways that algorithms, humans, and the broader systems in which they are embedded mutually co-constitute one another, not only as entities but as conceptual categories. This approach takes "human" and "algorithm" not as prefigured concepts, but rather as categories that only come into being through their mutual interconnections and interactions.

Authors' addresses: Eric P. S. Baumer, ericpsb@lehigh.edu, Lehigh University, 113 Research Dr., Bethlehem, PA, USA, 18017; Alex S. Taylor, University of Edinburgh, Design Informatics, Bayes Centre, 47 Potterrow, Edinburgh, Scotland, EH8 9BT, alex.taylor@ed.ac.uk; Jed R. Brubaker, jed.brubaker@colorado.edu, University of Colorado Boulder, UCB 315, Boulder, CO, USA, 80309; Micki McGee, mmcgee@fordham.edu, Fordham University, 441 East Fordham Road, Bronx, NY, USA, 10458.

ACM 1557-7325/2024/4-ART
https://doi.org/10.1145/3660344

This use of the term "*subjectivities*" is aligned with perspectives developed by Foucault [77, 78], by Deleuze and Guattari [58], by Ahmed [4], and by others. It argues that selves, rather than independent and pre-existing, are continuously brought into being by constellations of heterogeneous actors, structures, and processes. Rather than presuppose that subjects precede their subjectivities, we recognize that a much more complex relationship knits together what exactly subjects are and the phenomena of experience. Put simply, rather than argue that humans have experiences, our perspective focuses on the ways in which those experiences come to produce and define what we think of as human.

At first glance, this use of subjectivities may seem a simple acknowledgement of the plurality of subjective experiences people have, or even a post-modern attempt to question reality by claiming that everything is socially constructed. Instead, our intention is to make an important ontological distinction. We aim to draw attention to how subjects come into being through a multitude of subjective experiences. To reiterate, this conceptualization neither presumes people as pre-defined actors (that experience the world), nor does it claim that people do not really exist. Rather, it understands human actors as always arising from the relations in which they are entangled, always subject to (if you will) the constellation of entities and processes through which subjective experiences arise.

Drawing on this and similar work from critical theory [4, 58, 77, 78], from post-humanist and more-than-human scholarship [12, 13, 60, 91, 92], and from critical HCI (human-computer interaction) scholars [14, 15, 17, 185], *this paper offers a primarily theoretical contribution by extending subjectivity theory to help account for the role of algorithmic systems* in "enlivening" a range of subjectivities. Similar to the way Ahmed [3, 4] puts it, that "bodies take shape," the term "enliven" is meant here as connotative of the mutual co-constitution processes described above. In the interweaving of algorithms, the interfaces through which they are encountered, and the subjective experiences thereof, there is no single actor or process responsible for creating or constructing subjectivities. Rather, subjectivities come to into being through the interplays among technical implementation details, affective relations, systems of classification, etc. The knots threaded between subjectivities and the algorithmic are seen here as constitutive of both algorithm and human, not only in terms of individual entities but also in terms of the conceptual categories themselves, i.e., how humans and algorithms are defined and constituted. In light of the prevalence of algorithms, the notion of algorithmic subjectivities helps move beyond the limitations of our conceptualizing people and subjectivity through the lens of "users" [17] to consider how the conditions and experiences of use are shifted in algorithmic contexts.

Consider the following example: The practice of a user "Liking" and then "Hiding" the same post on their Facebook News Feed [71] demonstrates how such subjectivities come about and operate together, sometimes in tension. The algorithms underlying Facebook's feed curation, which determine what is served to the user, and interactional features such as "Like" buttons present the conditions for subjective experience. They afford a narrow range of responses from a user and in turn shape the scope for their experiences. Thus to "Like-and-Hide" is to establish a relationship to content that has been algorithmically ordered and prioritized. Literally giving a post the 'thumbs up' and at the same time demoting it is to affectively respond to an automatically determined output [cf. 36]. Yet, in some way, even if only in a small way, these actions allow one to be algorithmically defined by the contradictions inherent in these decisions. Key here is that the system and user are constitutive of one another, they become specific entities and understood as a distinct assemblage through their inter-relations. Hence our pairing of *algorithmic* and *subjectivities.*

A social media "Like" offers but one example. However, a growing number of different algorithmic techniques operate on a variety of different kinds of data, from text, to images, to sensor data, to networks, to geolocations, to many others. In this paper we turn to examples that focus on language and how the computational processing thereof, often referred to as natural language processing (NLP), enacts subjectivities in very particular ways. Our aim is to use language processing as an avenue by which to start examining algorithmic subjectivities more broadly. Through a series of concrete examples, we consider how paying attention to the kinds of subjectivities enlivened through algorithmic systems can attune us to a more responsive and responsible design. Such a design is sensitive to people's individual interactions with computational systems and services; at the same time, it recognises their complex interplay with the wider sociotechnical structures of classification, automation, and agency that are intertwined with the proliferation of technology (and have the potential to cause damage and harm) [30, 135, 137].

The first of our two case studies concerns the use of a specific form of NLP (topic modeling [24, 25]) to analyze blogs written by parents with children on the autism spectrum. This case study is situated in a historical context about the classification of autism in the Diagnostic and Statistical Manual of Mental Disorders, specifically the fourth and fifth editions (DSM-IV and DSM-5). We use this case to illustrate the complex, often subtle relationships among words/language and subjectivities [cf. 39, 91], as well as how certain kinds of subjectivities are enlivened as part of an assemblage of actors—an assemblage that both creates ways of seeing those actors and structures the types of claims that can be made about them, often with potentially significant consequences. Our second case study expands on this point about the consequences of algorithmic subjectivities using the example of algorithmic content moderation. This case illustrates how language, partly via computational processing, can become interwoven with myriad other facets of algorithmic systems. It also offers opportunities to investigate the complex interconnections and tensions between subjective experiences, such as having one's content auto-moderated, and technical implementation details, such as weights in a feature vector.

Synthesizing across these case studies, we draw on a variety of related prior work to articulate commonalities among the different instances of algorithmic subjectivities. First, algorithms' capacity for *inferring* or making predictions distinguishes these kinds of assemblages from those involving other kinds of computing technologies [cf. 159, 172, 189]. Second, it is the *entangling* of algorithms within such assemblages by which various actors and systems mutually produce one another and by which the exercise of power and authority occurs [cf. 13, 38, 79, 112]. Third, this mutual co-producing means that algorithms play an almost paradoxical role in *humanizing* subjects, i.e., in establishing and enacting definitions of what it means to be human [cf. 37, 92, 106]. While prior work offers useful concepts and theories to make sense of these qualities, it is our highlighting of the consequences arising from their unique combination that constitutes this paper's contribution. Across these qualities, we show that the orientation toward algorithmic subjectivities contributed by this paper elucidates how certain worlds [186, cf.] are being made and/or being made possible via algorithmic systems.

Drawing on this thinking, the paper closes with considerations of how things might be otherwise. Although it makes a primarily theoretical contribution, the ideas presented in this paper may also help inform design. If current algorithmic systems are making certain worlds possible, what other worlds are less readily possible? What other conditions of possibility might we want to enact? What new subjectivities, assemblages of actors, or forms of agency might be possible? What kinds of subjectivities do we want future historians to be able to read from our computational systems?

2 WHAT ARE SUBJECTIVITIES?

Before we present our case studies, this section elaborates on the concept of subjectivities, further introduces the notion of algorithmic subjectivities, and suggests where this notion intersects with interactive systems design and with broader sociotechnical concerns[1].

[1]The term "subjectivity" here refers not to the opposite of "objectivity" but instead draws on theories about subjective experiences and the constitution of the self [e.g., 4, 14, 58, 60]. This definition is distinct from subjectivity as a form of bias, which much work seeks to reduce in the name of fairness or justice [cf. 43, 51, 98, 142].

2.1 The Concept of Subjectivities

While we may experience the world in a variety of ways, it is nevertheless common to see ourselves as bounded individuals, and to understand subjective experiences as 'felt' through interior processes [161]. Consider how, when speaking aloud, the physical movements of one person's vocal cords (inside their body) trigger movements of another person's ear drums (inside their body) [136]. These visceral experiences give the impression of people as discrete individuals, each consisting of an interior that is distinct from the exterior world around them.

This orientation also rings true for how we in HCI view human-computer interactions or user experiences. Consider, for instance, the ways that HCI researchers study online social support [e.g., 54, 118, 132, 187, 192]. Our field's perspective acknowledges the differences in design affordances and subjective experiences between, for instance, receiving a heart emoji in reply to a public Facebook post vs. one sent via Messenger. Yet retained is an idea of the recipient being a unit, a distinct, persistent self, which perceives and then interprets the same world across these different settings.

Using subjectivity theory, Bardzell and Bardzell [14] describe how a reliance on this kind of singular, distinct self runs afoul in the context of HCI and design. Drawing on Foucault and postmodern approaches to subjectivity, they highlight how HCI far too often conflates a user with a coherent singular self. They argue for "subjects of information," contrasting between subject positions—i.e., the structural conditions of use afforded by a system—and subjectivities—i.e., the lived experiences and performances of people in subject positions. Bardzell and Bardzell argue that engaging the concept of the "user" through the lens of subjectivity has utility in that it does not simplistically collapse a person and a user together as equivalent, an argument further elaborated by Baumer and Brubaker [17]. Just as one might maintain multiple email addresses for distinct purposes (e.g., various professional and personal roles), considering each as distinct users (regardless of whether or not they are the same person) has utility in distinguishing between distinct structural and contextual realities—or subject positions. The separate self behind the "user," they argue, is not as important for design as the structural position from which use occurs and the subjective experiences to which they are connected.

If the separate self has limited benefits for the purposes of design, a counter view might be that we are not unitary actors at all. Late stage post-structuralist and post-modern theory has argued that it is the constellation of subjectivities that give rise to subject positions, rather than vice versa. We are continually constituted as differentiated subjects through the *unique conditions* and *relations* we experience. Our selves here are multiple, not purely interior, but always figured or enacted in and through the varied relations we encounter [4, 58, 60]. A constitutive approach to subjectivity casts the *self*, defined by internal subjective experiences, as just one of many ways to understand a human actor and their agencies, and a potentially quite impoverished one at that. As such, the relationalities within these constellations that are productive of subject positions are the focus of our current work. Rather than actors in search of a stage, we are performances resulting from the stage and the assemblages within which that stage resides.

Through this lens, Facebook, TikTok, and email each afford a different set of affinities and accountabilities that allow for different sorts of assemblages and selves to cohere. On social media, for example, the diffuse networks of millions of users and interactions are made possible through subjectivities that are often experienced in terms of simple metrics such as the "Like." While our experiences on these platforms are always more complex than metrics alone [39], we aim here to highlight the interrelations among those metrics, the attendant technical parts of our assemblages, and our experiences on these platforms.

2.2 Defining Algorithmic Subjectivities

It is this relational approach to subjectivities that we want to draw on in introducing and developing algorithmic subjectivities. We offer the term *algorithmic subjectivities* as a tool for thinking [174] that allows us to see how algorithms, in particular, interact within broader systems to create the conditions for subjective experiences and for certain kinds of actors to inhabit and operate in these experiences. Algorithmic subjectivities in this sense flips HCI orthodoxy, allowing us to see not how the user experiences the technology-infused world, but how actors—such as users—come into being through the experiential.

This approach differs from prior work that has considered relationships between subjectivity and algorithms. For instance, Blackwell [23, p. 193] argues that "operational definitions of AI must always be constructed in relation to humans." His concern, though, is primarily in how these human notions of (inter)subjectivity come to shape the assessment of AI systems, particularly the objective functions by which they are designed, rather than the kinds of subjective human experiences that arise around interactions with such systems. Somewhat relatedly, Fisher [74] asks how algorithmic systems provide people with a distinct way of knowing themselves. However, his argument is based on a formulation of subjectivity "as a quasi-transcendental (both utopian and actual) realm of individual freedom and authenticity" [74, p. 1]. Put differently, he focuses more on the formulation of the human as a legitimate (political) subject than on the co-constitution of human and algorithm as conceptual categories. More in line with this paper, Armano et al. [11] are concerned with the conditions of platform capitalism. They consider, for example, how "specific 'modes of feeling', through platforms, become forms of subjectivity, implicit ways of selecting choices and ultimately of looking at the world" [11, p. 3]. However, they spend less attention on the interplay between these "ways [of] looking at the world" and the technical details of the algorithmic components within those platforms. Thus, our approach to defining and articulating algorithmic subjectivities offers a unique perspective on these phenomena.

This perspective is important for at least two reasons. First, it invites us to look beyond the individual experiences of the user and to pay close attention to the conditions and relations that make particular subjectivities possible. In other words, it helps us to examine human-computer interactions well beyond a user's individual, subjective experiences or even the subjective experiences of groups, whether co-located or distributed. In doing so, we find that the things we often conceive of as "the user" and as "interactions" are in fact constituted through heterogeneous entanglements of larger structures—interfaces and algorithms, to be sure, but also systems of norms, value, agency, etc.—that organise and manage the many worlds we experience [17, 177]. Second, it allows us to speculate on the conditions of possibility for doing these worlds differently. This point goes beyond asking counterfactual questions about what other possible worlds might be like [116, 162] [see also 186]. For example, the reduction of human experience to the Facebook "Like" asserts a singular world—a world that casts the subjective self in terms of distinctive social, economic, and political forms of life [see, for example, 57]—at the exclusion of other possible worlds. The perspective advocated here suggests considering how not only the existence of a "Like" but also the technical implementation thereof constrains possibilities by making certain possible worlds less likely. Thinking in terms of algorithmic subjectivities, we suggest, provides a basis for designing algorithmically-infused worlds that might be different, that might create the conditions for more varied and diverse actors being accounted for [93].

2.3 Intersection with Design

For design, the relevance of this view of subjective selves and algorithmic subjectivities ties to what we in HCI see as the frame of analysis, and the scope and scale of what we are intervening in when we design interactive systems. Bardzell and Bardzell [14] make a distinction between subject positions and subjectivities (largely for the sake of clarity). However, algorithmic systems highlight their coupled nature, and it is that coupling to which this paper attends.

Thus, while Bardzell and Bardzell provide useful theoretical grounding, this paper builds on that grounding to focus on how subjectivities emerge and transform, particularly in relation with the design of algorithmic systems. Within such systems, seemingly trivial human-computer interactions, such as trackpad clicks or swiping speeds, can be leveraged and analyzed, feeding back into the design of a system. This coupling of aggregation, analysis, and feedback thus starts to order and to regulate forms of subjective expression and correspondingly to give form to legitimate versions of the subjective self [34, 126]. The importance for HCI of the work that follows, then, is both to show why enlarging the scope and scale of what we study and design for is critical, and to offer examples and concepts for how we might do so.

3 CASE STUDIES

This section traces the doing of algorithmic subjectivities through two case studies. The first case study deals with the application of computational analysis to blogs written by parents with children on the autism spectrum. After a brief review of the historical background on the classification of autism, this case study considers how algorithmic systems involving natural language processing can play a distinct role in enlivening subjectivities. The second case study considers the algorithmic moderation of online content. It gestures toward the ways that algorithmic subjectivities arise from complex assemblages of data—not only linguistic but multimodal and multifaceted—as well as toward some of the consequences of such subjectivities.

Neither of these case studies is intended to advance the overall understanding of their respective application domains, i.e., ASD and content moderation. Rather, they are intended to explore the role of algorithmic systems within the enlivening of subjectivities. Synthesizing across these cases, both in terms of commonalities and in terms of differences, helps enumerate the unique qualities that typify how these situations arise and operate.

3.1 Topic Modeling and Autism (Classification)

This section illustrates the unique roles of algorithmic text processing to enliven certain types of subjectivities within broader systems of classification. In particular, the fraught role of language within the history of classifying and diagnosing autism spectrum disorder (ASD) helps illustrate this paper's focus on the role of natural language processing in enlivening algorithmic subjectivities. After reviewing some of the historical background on the importance of language for the classification of autism, we interrogate the kinds of subjectivities that arise in the context of one specific type of natural language processing algorithm.

3.1.1 Historical Background. Classification procedures and their consequences have long played a significant role in the context of health and medicine, especially in the specific context of mental health. Examples range from the international classification of diseases [30] to the evolving social construction and treatment of "madness" or "insanity" [76], to the reification of "healthy" and "disordered" users in machine learning analysis of social media data [44].

The present case study focuses specifically on the classification and diagnosis of ASD. The Diagnostic and Statistical Manual of Mental Disorders (DSM), produced by the American Psychiatric Association (APA), provides a taxonomy with categories of mental disorders, related criteria for the diagnosis of disorders, and statistics related to these categories. The DSM has undergone several revisions, from the DSM-I (1952) through the DSM-5 (2013). One of these revisions, the DSM-IV (1994) [7], included a typographical error in the diagnostic criteria for autism [105, p. 520]. Specifically, in the DSM-IV, the diagnostic criteria included "severe and pervasive impairment in the development of reciprocal social interaction *or* verbal and nonverbal communication skills, or when stereotyped behavior, interests, and activities are present" [7, p. 77, emphasis added]. As a result, patients who presented *either* impaired development of reciprocal social interaction *or* impaired development of verbal and nonverbal communication skills were diagnosed with Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). This diagnostic category includes Atypical Autism, i.e., "presentations that do not meet the criteria for Autistic Disorder" [7, p. 78]. This typo was amended with the DSM-IV-TR (2000) [8], where the criteria were revised to "severe and pervasive impairment in the development of reciprocal social interaction *associated with* impairment in either verbal or nonverbal communication skills or with the presence of stereotyped behavior, interests, and activities" [8, p. 84, emphasis added]. This change effectively meant that an individual would need to present *both* impaired reciprocal social interaction *and* either impaired communication (verbal or nonverbal) or stereotyped behavior, interests, and activities to receive a diagnosis. Further revisions were conducted in 2007–2012, resulting in the 2013 publication of the DSM-5.

This situation contributed to uncertainty about the autism "epidemic" during the late 1990s. Was there an actually-existing increased prevalence in persons with the constellation of behaviors and habits of mind that had come to be called autism? Or was this a classification error, in part precipitated by a proofreading oversight made along with other adjustments to the DSM-IV-TR? While such questions will likely go unanswered [82], the legitimacy of the diagnostic criteria became a matter of public and professional scrutiny.

These complexities—of language, classification, and repercussions—are not our central focus here. Rather, this context of ontological and epistemological contestation—or, perhaps more aptly, crises—provides necessary background to understand the nature of the present case.

3.1.2 Topic Modeling Subjectivities. This case comes from some of the authors' own experiences conducting research, and was a major factor in drawing our attention to the need for, and theoretical gap around, subjectivities and algorithmic systems. The case centers around a collection of blogs written by parents with children on the autism spectrum. These blogs were identified based on Author M's[2] familiarity with the community, as well as iterative review of the content on each blog. In total, this corpus includes 31,976 documents (i.e., blog posts) with a total of 17,273,079 words (words per document M=540.2, Mdn=430).

Two of the Authors [M and P] have worked on efforts to analyze and to understand the experiences reported in these blogs [anon]. Given the volume of this data set, our primary methods have involved topic modeling [24, 25]. Ostensibly, topic modeling provides a means of identifying the latent themes in a corpus of documents. A theme, or "topic," is represented as a probability distribution over words. For example, one topic might have high probabilities for words such as *human*, *genome*, *dna*, *sequence*, etc., while another might have high probabilities for words such as *computer*, *models*, *information*, *data*, etc. [24]. These probability distributions (i.e., topics) are inferred from a corpus of unlabeled documents; the model is provided no information *a priori* about the topics. The only parameter a human sets for the model is the number of topics, everything else is unsupervised. The model also assigns a topic proportion for each document. Continuing the above example, an article from a bioinformatics journal might be 40% about the first topic above, 30% about the second topic above, and 30% about other topics.

Following on recent work [19, 104, 127, 131, 138, 156], these two Authors [M and P] have applied topic modeling as an interpretive lens on their data. That is, the results of topic modeling were not seen as a map providing a direct representation of the data. Instead, we sought to use them as a compass that could indicate certain directions in which it might be fruitful to look [cf. 70].

In reviewing the resulting topics, we noticed a curious pattern: many topics went beyond being strictly "topical" in nature. Similar to other interpretive uses of topic modeling [157, 179, 180], many of these topics aligned closely with particular discourses, especially specific perspectives or arguments. For example, one of the topics had, as high probability words *spectrum*, *disorder*, *autism*, *diagnosis*, *disorders*, etc. The documents with a high proportion of this topic discuss changes in the DSM-5 diagnostic for ASD, the very changes discussed in the historical background above. Furthermore, most of these documents are fairly critical of the changes being made in the revisions from DSM-IV-TR to DSM-5[3].

Such topics, then, become a lens through which documents are made sense of by human readers. For example, one post in the corpus was assigned as 57.8% about the topic described in the preceding paragraph (*spectrum*,

[2]For blind review, all author names are replaced with pseudonymous letters. In the final version, these pseudonyms will be revised to the authors' family names.

[3]Given the ease with which verbatim phrases can be searched on the Internet, we omit verbatim example quotes from any of these documents to help protect the identities of these bloggers [18].

disorder, *autism*, etc.). The text of this post consists entirely of a copy of the DSM-IV-TR criteria, with the title "So hold on, Does My Kid Actually have Autism?"[4] Such a post could be interpreted in various, potentially contradictory ways: as critical of the diagnostic criteria, as voicing frustration or confusion, as demonstrating the ease with which such criteria could be applied, as asking other parents for help in diagnosing the blogger's child, etc. However, knowing that this post has a high proportion of this topic gives readers an indication that the post (and implicitly its author) takes a critical stance toward the DSM-5 revisions to the autism criteria, *before the reader even begins reading the document*. Put differently, the topic model becomes involved in enlivening certain subjective experiences of and relationalities around the DSM-5 that have been algorithmically calculated.

That said, these algorithmic subjectivities—e.g., of criticality or skepticism toward the DSM-5 autism criteria—are enlivened by a human reading the topic model in a particular way. Within a topic model, each topic is formulated as a probability distribution over words, and each document is formulated as a probability distribution over topics. These kinds of formulations give rise to a particular type of subjectivity, one that differs from the descriptions that would likely be offered by humans mediated by tools other than topic modeling. It is unlikely that either human researchers or the bloggers who hold this stance toward the DSM-5 would describe the stance in terms of calculable probability distributions. Thus, this subjectivity is enlivened through an assemblage of computational manipulations and human interpretations of that computational processing. The orientation of this assemblage likely differs from the manner in which either humans alone or a computational model alone would conceptualize and organize the same data [cf. 37].

To further illustrate the unique qualities of this human-model assemblage, consider how topic modeling, as a computational technique, also affords a variety of inferential possibilities [121]. For example, the model can be used to make predictions about the topics present in a novel document that was not included in the training corpus. This novel document could be another blog post, by either one of the same bloggers or by another author, or it could be some different text (e.g., public statements by elected representatives). The words that occur in this novel document can be used to infer a probability that topic identified from the training corpus is present in the novel document. As another example, the model can be used to computationally identify similarities among different authors. Aggregating the topic probabilities across all the documents written by a given blogger would allow for computing the similarity among any pair of bloggers or even identifying groups of bloggers. Such relationships are based not on any direct interactions among the bloggers but only on latent aspects of their language use. Finally, the topic modeling results could easily be used as input for other predictive analyses. For instance, the occurrence of certain topics in a parent's blog may be predictive of, say, having a child who is non-verbal, or enrolling one's child in a public school, or having received government-supported health assistance. These kinds of inferential possibilities constitute a distinguishing quality of algorithmic subjectivities, as argued further below.

Furthermore, algorithmic subjectivities do not emerge only from individual human-algorithm interactions. Rather, they come into being through becoming entangled within much broader structures and systems. The historical background above circumscribes some of these: clinical diagnoses, insurance companies, government bureaucracies, pedagogical strategies, etc. The incorporation of computational language processing could readily and significantly alter the assemblages, processes, and authorities by which autistic subjects are enacted by various entities.

These alterations also bring potentially significant consequences. One could certainly envision computational tools similar to topic modeling being used as a barometer for public sentiment, or potentially even for diagnosis. Indeed, Keyes [106] analyzes a large corpus of research on AI for autism diagnostics, which uses computer

[4]This title is obfuscated to protect the blogger's identity; see preceding footnote.

vision (e.g., gait analysis), signal processing (e.g., analysis of audio recordings), and other techniques[5]. Having an official autism diagnosis could provide access to (in the US) Medicaid, Supplemental Security Income, private insurance reimbursements for particular services, and other social provisions for care. At the same time, an autism diagnosis might also disqualify a student from funded placement in schools that focus on serving those who are deemed "speech-language impaired" or "emotionally-disturbed." Damaging impacts can occur when such resources are allocated based on official designations, rather than on an individual's educational or social emotional needs [41, 191].

However, the point here is neither the specific diagnosis that a particular individual might algorithmically be assigned, nor whether a particular blogger/parent would be labeled as being critical of certain diagnostic criteria, nor reflections on the social construction of ASD. Rather, this case highlights the particular ways that subjectivities arise around the algorithmic identification of such criticality. Doing so casts criticality of the DSM (revisions) as, perhaps even reduces it to, a statistically identifiable pattern of language use. The DSM critic and the algorithmic system work to co-construct each other in very particular ways. In the case of topic modeling, these ways hinge upon probability distributions over co-occurring word tokens, which then afford numerous inferential possibilities. Yet those inferences, probability distributions, word tokens, etc. are interpreted through the lens of this sociohistorical context, while simultaneously recasting the context as one that can be understood in terms of word tokens, probability distributions, inferences, etc. The examination of these relationships through this case study helps elucidate the unique qualities of algorithmic subjectivities, as described further below.

3.2 Algorithmic Moderation

Content moderation—decisions about what can and cannot be posted in online groups—reveals a complex network of subjectivities that are being extended, negotiated, and infrastructured by the development of algorithmic and hybrid forms of linguistic and textual analysis. Moderation is a key contributor to the success of online communities [109] and has garnered significant research attention [e.g., 45, 46, 64, 73, 81, 88, 101, 102, 117, 129, 181, 182]. Recent news is also rife with cases where the failure to moderate content has been scrutinized (e.g., hate groups, online bullying, or violent extremism) [1]. At first glance, moderation may seem straightforward: content that violates community norms is identified and removed. Upon closer inspection, it is almost always more complex. To make this argument, this section weaves together examples and points across prior literature, both work on content moderation generally and work on algorithmic moderation specifically. This strategy thus complements the preceding case study's use of the authors' own research and experiences, elucidating how concerns around algorithmic subjectivities emerge across a variety of work in the context of content moderation.

On most major platforms, an army of content moderators move through streams of online content that has been flagged by people who object to its presence. These content moderators make judgement calls based on content standards and policies developed by teams that must operationalize the thresholds between acceptable and unacceptable content. These standards, though, are not entirely standardized. For instance, Facebook's public Community Standards explains that, while it can take many forms, bullying is not tolerated "because it prevents people from feeling safe and respected on Facebook" (https://www.facebook.com/communitystandards/bullying). At the same time, the policies also explain that public figures are of a different class (so as to enable public critique), as are 13-18 years olds (for whom protections are heightened). Furthermore, these teams of content moderators are often globally distributed, so as to "follow the sun," as well as contracted, with recent reports describing extreme working conditions where moderators are managed and measured to ensure the contracting agency meets their commitments [133].

[5] This trend is not limited to autism. For instance, machine learning techniques have been applied to the detection of Parkinson's disease using, e.g., speech audio [122]. Similarly, text and other social media data have been used to detect a variety of mental health conditions [44].

The scale of content and limited (human) resources have given rise to hybrid moderation approaches, wherein algorithms flag potentially inappropriate content for human review. Teams of engineers are tasked with developing algorithmic tools to aid moderation, working with still others [83, 160] to determine what types of content—due to volume, severity of the violation, and technical feasibility—should be prioritized. Some types of infractions are easily detected and are delegated to algorithms entirely. Simple rule-based "auto-moderation" may be used by administrators of small communities on sites such as Reddit [107] and Discord [102]. Facebook gave Page admins the ability to automatically filter posts that contained specific profanity as early as 2011 [49]. In situations such as these, algorithmic tools are used to automate some of the human labor of moderating these spaces. These basic moderation techniques function by simplifying the potential meaning (and thus appropriateness) of content and the appropriate response [cf. 55].

Yet many infractions are nuanced, requiring case-by-case decisions that are likely beyond the ability of any automatic tool. Scenarios such as hate speech, photos of breastfeeding mothers, and documentation of violence in conflict zones [1] highlight how "appropriate content" is bound up with context, time, and culture [103]. For instance, Haimson et al. [88] note that content removal, both human and automated, occurs disproportionately often for some populations, including political conservatives, transgender people, and Black people. They argue that moderation practices need to address more directly gray areas—content that is neither obviously acceptable nor obviously objectionable—while pointing out that "there is a vast difference between silencing conservatives' misinformation and hate speech and silencing trans and Black users' personal identity-related content" [88, p. 24]. Even more mundane content presents nuance when deciding what actions should be taken. As Seering et al. [169] found in their study of volunteer moderators, nuance often comes when deciding what level of punishment to apply—from warnings to user bans—when users violate the specific norms or rules of a community. At the same time, a thin and blurry line can emerge between nuance and inconsistency, the latter of which may drive perceptions of unfairness [117] [also see 40].

These complexities are further compounded by varying degrees of transparency and opacity. Responses to the possibility of unknown algorithmic curation in platforms such as Facebook or Twitter often take the form of surprise or opposition [62, 72, 147]. On Yelp, however, responses were even more affectively charged, with reviewers describing the opacity of algorithmic filtering as "sneaky," "deceptive," "misleading" and even "possibly censorship" [73, p. 6]. Thus, we see significant repercussions not only from the presence of algorithmic moderation but also from the specifics of its implementation and interconnection with various relationalities.

At least two points arise from such situations. First, they reveal a potential tension between individual subjective experiences of being moderated and moderators' broader concerns for creating a "safe and respect[ful]" sphere of interaction. Second, it is possible that these moderation systems are in fact being internally consistent but in ways that are not readily legible to the human users whose content is being moderated. Such questions of legibility and interpretability play directly into the kinds of subjectivities enlivened in these systems, as discussed further below.

The unit of moderation is also significant. Does one moderate a specific piece of content? The author of the content? The community in which that content was produced or posted? Here, well intended algorithms can go awry. Leveraging a user's history, while attractive, presents some problems. Content creators on YouTube have reported that, when moderation is applied to one video (e.g., by adding an age restriction to it), their subsequent content receives less traffic, even when that subsequent content includes nothing that warrants moderation. In another example, trolling is better predicted by social context than by a person's history [47]. Bullies, likewise, are often themselves bullied as well [96]. An algorithm designed to identify only the author and target around individual pieces of violating content, however, could only work to enliven mutually exclusive subject positions. Such nuanced multiplicity would be lost.

Alternatively, the unit of moderation may be the community in which certain content was produced. For instance, Reddit will at times quarantine an entire community due to excessive toxicity or questionable (mis)information [150]. However, such quarantining has raised concerns about possible censorship [111, 120]. Thus, moderation beyond the scope of a single piece of content or user may exceed the capacity of algorithms, or at least what we are willing to trust them with, even when working in tandem with teams of human moderators.

Hybrid moderation approaches thus highlight how human moderators and algorithmic moderators are bound up with each other. In this way, content moderation becomes heavily influenced by and reliant on the sociotechnical relations to which algorithms can be responsive. The conditions in which moderation should or might occur must be made legible to an algorithm. Put differently, *moderating content algorithmically requires that numerous different subjectivities are operationalized in particular ways*. The person being bullied, the bully, the moderator, the various audiences witnessing the behavior—each must be conceptualized in computationally codifiable ways. How this is done shifts both the subjectivities themselves and their relationships to each other.

Moderation also gives us another example where the subjectivities at play have consequences. The consequences are different from those arising in the above cases involving the DSM and autism, but they matter nonetheless. Moderation of online platforms is, in effect, moderation of our contemporary public square [151]. The consequences are nothing more than countless viral videos and nothing less than free speech and functioning democracy.

4 THE QUALITIES OF ALGORITHMIC SUBJECTIVITIES

Any meaningful understanding of algorithmic subjectivities requires a processual orientation. An analogy can be drawn here with the relationship between studying the noun *infrastructure* (i.e., treating infrastructure as a distinct entity to be analyzed *per se*) and studying the verb *infrastructuring* (i.e., examining the processes by which infrastructures come to be) [171]. Analogously, we do not posit algorithmic subjectivities as separate entities to be examined in and of themselves. Rather, we are interested in how algorithms and algorithmic systems *do* subjectivities, the processes by which these subjectivities come to be. By synthesizing across the case studies above, this section articulates this paper's contribution in terms of extending subjectivity theory to grapple with the unique qualities that typify the doing of algorithmic subjectivities.

Given the theoretical nature of this contribution, the majority of this section uses prior theoretical literature to articulate these qualities. In many cases, such prior work provides valuable conceptual vocabulary for describing various parts of these qualities and how they operate. At the same time, few if any of these concepts alone help us account for the whole. Put differently, this section shows how the combination of qualities characteristic of subjectivities enlivened within and around algorithmic systems requires the multiplicity of theoretical perspectives marshalled here.

Although this paper focuses primarily on "implications for theory" [68], we also attend to aspects of these processes related with the *design* of algorithmic systems. In contrast with prior work, we offer neither case studies [e.g., 197] that illustrate exactly *what* kind of designs should be implemented, nor guidelines [e.g., 9] prescribing just how such design work should be accomplished. Instead, we provide various conceptual orientations that can help *attune* designers to the unique roles of algorithmic systems within the enlivening of subjectivities.

4.1 Inferring

Inference plays a key role in algorithmic systems. A tuned content moderation system can infer which content is likely to be flagged by users as objectionable. A trained topic model can infer the topics present in a novel document. Numerous potentially sensitive attributes about a social media user can be inferred from seemingly benign information about them [97, 170]. This ability to make inferences has significant consequences, as described in the case studies above. Such inferential predictions, we suggest, are a defining characteristic of how algorithmic subjectivities are done.

Prior work [e.g., 30, 77] has highlighted how the informational activities of labeling and classifying are both epistemic acts of knowing—what type of thing is this data point?—and simultaneously exercises of political power. Put differently, classification places the data point into an existing knowledge structure with its own history, value commitments, political underpinnings, etc. When the thing being classified is a human person, especially when the person is being classified by someone else or something else, the act of classification becomes an exercise of power, insofar as it defines that person in ways that may or may not align with how they define themselves.

However, algorithmic inferences go beyond purely classifying or labeling data points, human or otherwise. Indeed, the subjectivities enlivened through algorithmic systems may not have previously existed as a class or label. Returning to our case studies, a blogger may, by writing in a particular way, indicate implicitly or indirectly that they hold a specific type of critical stance on the autism criteria in the DSM. While it would certainly be possible to label a blogger as being critical of the DSM criteria without topic modeling—based on, say, explicit statements they make—topic modeling works to enliven that subjectivity of criticality in a particular way. This algorithmic enlivening simultaneously casts the subjectivity as manifest through statistical patterns in word choice, suggests that observations of an individual's language use may allow for inferring the degree to which the individual performs that subjectivity, and may overshadow or even exclude other possible interpretations of that same language. In such ways, the interplay of algorithmic systems goes beyond assigning existing categories—anti-vaxxer, troll [47], victim [96], etc. Instead, they work to enliven subjectivities that are seemingly familiar yet simultaneously predicated upon mechanistic inferring [cf. 58].

Those mechanistic inferences are often based not on directly observable data but on a so-called latent feature space [e.g., 123]. In the above ASD case study, each topic (described by a probability distributions over words) is treated as one dimension of such a latent feature space. The words in these probability distributions, due to their semantic meaning, are readily human interpretable. For instance, in the example from the case study above, a topic's high probability words (*spectrum*, *disorder*, *autism*, *diagnosis*, *disorders*, etc.) can be interpreted as providing a brief description of what the topic is about. Similarly, representing a document in terms of these latent topics also provides a human interpretable semantic description about the content of those documents. The dimensions of this space, i.e., the topics themselves, are referred to as "latent" because they are *inferred*, rather than being directly observed in the data.

Although many algorithmic systems employ such latent feature spaces, they are not all as readily interpretable. For instance, large language models (LLMs) [e.g., 63, 123] use latent dimensions that often lack any obvious or semantically meaningful description. Instead, representations of words and documents are transformed into an embedding space, sometimes referred to as a "semantic space," often comprised of a few hundred dimensions. This size contrasts with traditional approaches to document representation, which often have one dimension for every word in the vocabulary (i.e., tens of thousands of dimensions). Thus, although they may use more latent dimensions than is common with topic modeling, these embedding spaces still offer a significant reduction in the number of dimensions used to represent a document. The mappings, from words or documents into these latent representations, are inferred by iteratively optimizing their use in predicting masked tokens (i.e., a single word that has been omitted from a sentence) [63] in massive textual data sets (e.g., web crawls from millions or billions of web pages). The resulting representations often significantly reduce overall sparsity, which in part accounts for these representations enabling improved performance on downstream NLP tasks (sentiment analysis, named entity recognition, text classification, etc.). At the same time, though, the use of such representations significantly reduces the ability for a human to determine what any given latent, i.e., inferred, dimension means.

This difficulty in interpretability of inferences has a number of consequences. For instance, it partially explains why so many different techniques have been developed simply to detect the presence of biases in large language models [e.g., 28, 84, 140, 143], i.e., because the limited interpretability of the semantic representation space makes biases difficult to notice. It also helps account for ways that online advertisements can be targeted to, e.g., cannabis users, even when an advertising platform does not offer cannabis use as an explicit targeting option [29] [see also 125, 170]. Put succinctly, algorithmic systems make inferences based not only upon statistical patterns within observed data but also using transformed representations of those observed data that are themselves inferred and thus often not readily human-interpretable.

At the same time, algorithmic subjectivities are not comprised solely of data points to be analyzed or of models to be manipulated. Consider, for instance, the processes of enlivening autistic subjectivities through assemblages of the DSM, clinical practices, educational institutions, government bureaucracies, etc. One cannot "run" this subjectivity like a model to make predictions about the likelihoods of different outcomes, such as how various treatment options might influence an individual's eventual educational attainment, annual earnings, or life satisfaction. Similarly, one would not be able to predict automatically the goodness-of-fit between this subjectivity and any given person. Thus, algorithmic subjectivities do not, *per se*, make predictions or inferences about the world. However, the capacity for making quantitative predictions—for inferring—becomes a distinguishing characteristic of algorithmic subjectivities when that capacity becomes entangled within broader assemblages of persons, institutions, and structures of meaning, as described further below.

4.1.1 Designing for Inference. These inferential capabilities connect with challenges in designing algorithmic systems [69, 114, 197, 198]. Designers must somehow anticipate not only the results that are algorithmically surfaced during any interaction but also the broader sociotechnical ecosystems in which that content may emerge. Furthermore, the algorithmic models themselves change in response to ever changing data streams.

Two strands of thought are helpful in understanding and designing around such inference. The first comes from work on modernism [159, 189], which is often described as having four key tenets: calculability, efficiency, predictability, and (hierarchical) control [for a concise description, see 35, p. 950]. Inference is intimately connected with each of these: in order to make inferences, bodies must be made calculable; inferential models enable making predictions about a person's actions; etc. Thus, understanding algorithmic subjectivities requires understand how bodies come to be calculated, to be made efficient, to be made predictable, and to be (hierarchically) controlled.

Put differently, seeing algorithmic systems as a modernist enterprise helps guide our analytic and design attentions. Consider two examples, both drawn from the above case study about topic modeling and ASD. On the one hand, topic modeling translates different perspectives (e.g., criticality toward the DSM-5 revisions) into something that can be calculably identified, in this case, by examining statistical patterns of word co-occurrence. These inferences simultaneously enable making predictions about how individual bloggers might behave or react to certain events (e.g., future DSM revisions) [cf. 86, 153]. On the other hand, the use of such inferential predictions enables a host of design possibilities, ranging from tools that individual bloggers could use to reflect upon their own family's journey via analysis of their writing [similar to 32], to systems that posit connections among multiple blogs for various purposes [similar to 20]. Designers can attending to these core tenets of modernism by considering what things are (and what things are not) made calculable, made efficient, predicted, and controlled by different design possibilities. Doing so provides a conceptual language to account for the various ways that algorithmic system design might figure in enlivening different subjectivities.

Second, we can also understand the distinct role that inference plays by drawing on the notion of the scalable subject [172]. Described as a refinement of the data double [87], Stark [172] highlights how digital traces about an individual are used to create mathematical and computational models. These models and their attendant uses represent a unique confluence of work in the psychological sciences and in computer science, one that has significant ramifications for the understanding of, and for the control of, individual persons.

To illustrate these points, Stark draws on a variety of examples, from A/B testing to Facebook's emotion contagion study [108], to mental health tools intended to assist patients with mood and behavior disorders (e.g., Ginger.io). These cases and others, he argues, all involve assuming that relationships between different variables that occur in the aggregate will also apply to those same variables for an individual [citing the "ecological fallacy," 144]. Put differently, scalable subjects are created in part by when observations of past data points (often humans and their activities) are used to draw conclusions about new data points, i.e., to draw inferences about them. It is such inferences that allow for the scalable subject's scalability.

Thus, we suggest, this conceptual apparatus [172] is useful for reasoning about, and perhaps for designing around, the means and consequences of inference. For instance, it is perhaps obvious that the content moderation of posts and individuals—as bully, target, toxic, bystander, etc.—occurs via algorithms making inferences. The conceptual lens of scalability suggests that designers attend to the mechanisms by which inferences are made. For content moderation, a given post would first be projected into a latent feature space, as described above. Then, the post's distance[6] within that feature space could be used to infer how likely the post is to be an instance of, say, bullying or toxicity. The use of this inference mechanism is based on an assumption of scalability: that the manifestation of toxicity results in posts that, when projected into this feature space, become geometrically proximate. Focusing on the mechanisms by which these inferences happen allows designers to consider whether such scalability should hold in this context, or if there might be other means for identifying toxic or bullying content.

While analytically useful for examining, and perhaps for designing around inference, the notion of the scalable subject [172] provides less guidance or insight about individual subjective experience. Indeed, "lost in descriptions of the aggregate are the ways in which individual subjects understand their own scalability" [172, p. 213]. As we have argued here, though, algorithmic subjectivities involve not only individual humans' subjective experiences but also their interplays among multiscalar actors. Thus, understanding algorithmic subjectivities requires complementing the scalable subject with other theoretical devices that account for such entanglements.

4.2 Entangling

Algorithmic subjectivities are enlivened via interactions among and within heterogeneous assemblages of actors and processes. Thus, we do not claim that algorithms enliven subjectivities on their own. A moderation system may flag a charged piece of content; a topic model may assign a poignant topic to a document. Yet such computational procedures do not function as entirely independent, distinct entities, neither analytically nor practically. Rather, the subjectivities in which we are interested arise through algorithms that are enmeshed or entangled [13, 79] in heterogeneous aggregates of entities, actions, and interpretations. Such subjectivities come to be enlivened through the entangling of, for instance, content moderation algorithms within processes that interweave human users, automated systems, policy-forming bodies, low-wage human labor, norms of acceptable interpersonal interaction, etc. Thus, it is not the algorithm *per se* that enlivens algorithmic subjectivities, but how an algorithm operates as part of and comes to be entangled in what are always much broader structures. Indeed, this entanglement is concomitant with the distinctly massive scale and scope at which such subjectivities operate.

Such entangling raises at least three interconnected concerns for which we need to account. All three, in various ways, involve questions of how individual actors interact with one another within these entangled assemblages.

First, *whither agency*? Put differently, if individual actors can no longer be seen as entirely distinct entities, how can we conceive of the agency with which individual actors act? Consider a treatment from Latour [112] of competing claims related to gun regulation. One side claims that "guns kill people," while another side claims that "people kill people; not guns." These two statements offer competing claims about agency (and, thus, responsibility). Latour resolves the apparent tension by arguing that neither statement is entirely accurate. Instead, by entering into relation with one another, the gun and the person holding it combine to become a different kind of actor, "a

[6] Different measures of distance or proximity can be used, including simple Euclidean distance, cosine similarity, spectral metrics, and many others.

citizen-gun, a gun-citizen" [112, p. 32]. In this way, "it is neither people nor guns that kill;" instead, "responsibility for action must be shared among the various actants[7]" [112, p. 34].

Similar logic can be applied to understand agency within algorithmic systems. As an example, stochastic gradient descent (SGD), a common algorithm for training machine learning models, becomes a different kind of actor when applied to data in which harassing or toxic content has been labeled. To be sure, content moderation systems do not kill anyone in the same way that a gun does. However, content moderation systems do play a role in flagging content and banning users. To say that the system (or even the trained model on which the system is based) does the banning or flagging elides the significant complexities involved. The trained model, the SGD algorithm, the training data, the human content moderators on whose labor those training data are based, the programmers who build the system, the (often corporate) organizations who collect these data and oversee these systems—none of these is solely responsible, because agency does not belong to any single one of these entities. Rather, as Latour suggests, responsibility must be distributed among them.

The notion of distributing responsibility raises a second question: given this entangled soup of actors, *how might we go about about defining individual entities*? If responsibility is distributed among the "actants," as Latour [112] calls them, how do we even determine who or what these actants are? Is it reasonable even to consider doing so? One approach to these questions draws on the notion of entanglement from Barad [1[illegible]]. She argues that these entities do not necessarily exist as such prior to their interactions. Rather it is through what Barad calls intra-actions that entities become co-constitutive of one another. As an illustration, quantum physics dictates that it is possible to know either a particle's position *or* its momentum, but not both simultaneously. Barad points out how this account reinforces a duality between world (i.e., the particle) and representation (i.e., measurements of its position, momentum, etc.). Put differently, this problem posits that there exists a particle that has both position and momentum. Instead, Barad suggests that, prior to observation, "the particle simply did not exist in any fixed state" [99, p. 929], but rather existed in a state of indeterminate potentiality. This is not to say that the particle does not exist at all prior to observation, but rather that aspects of its existence were as yet indeterminate. Through the phenomenon of observation, the particle, its position or momentum (but not both!), the measurement apparatus, and the knowing human observer all come to be in a particular stabilized relation. Barad refers to this as an *agential cut*, the moment/process by which indeterminate entities come to be stabilized (even if only temporarily) into mutually constitutive relationships.

Recent work, especially from Frauenberger [79], has suggested that entanglement theories offer a novel generative metaphor for HCI. Entanglement, it is argued, provides a fundamental reconceptualization of the "interaction" in human-computer interaction. "Things and people, as phenomena, mutually constitute each other through their intra-action, i.e., the boundaries between human and machines are not pre-determined, but enacted" [79, p. 9]. Just as the particle is brought into a certain fixed state (and excluded from other states) through its intra-action with a measurement apparatus, a human observer, etc., any given sociotechnical system exists in a state of indeterminacy until entangled in some particular way. Frauenberger [79] illustrates this point using the example of *Flow*, a "hypothetical device [that] displays the ease or anxiety of members of a conversation based on data from a range of sensors" [79, p. 12]. Different means of evaluating this system enact different agential cuts: "an interview study will make Flow a cultural artefact, a controlled user-testing study in the lab will make it a functional tool, and a long-term diary study might make it an artificial sense of people" [79, p. 15].

This kind of thinking helps us account for the complex relationalities within algorithmic systems. The bully, the victim, the toxic content, the onlookers or bystanders—all of these come to be entities with certain properties and relations among them because of the agential cuts made by content moderation systems. Again, this is not to say that, e.g., the bystanders did not exist prior to content being flagged. The bystanders were there, but they came to take on the role and enter into the relations of being bystanders with certain properties (and not other

[7] Latour uses the term "actant" to avoid implicit assumptions of anthropomorphism on the part of any acting entity.

properties they could have had) in part because of the content moderation system's intra-actions with them and with the other actors involved.

If we apply these notions of indeterminacy to algorithmic systems, a third question arises: *how might we conceive of power dynamics and the exercise of authority*? Couldry and Mejias [53] argue that the collection and analysis of data operate in power dynamics that resembles colonialism. Much as colonial governments appropriated land, bodies, and natural resources to maximize profits, technologies companies analogously quantify social interaction into data to extract value from it. Much as capitalism has historically transformed human activity into the commodity of labor, data colonialism transforms the experience of human sociality into the commodity of data. Thus, in this formulation, technology companies who collect and analyze these data become the dominant enactors of data-driven authority.

Similarly, Burrell and Fourcade [38] suggest that the power to make such determinations resides with those whom they call the "coding elite."

> "The coding elite is a nebula of software developers, tech CEOs, investors, and computer science and engineering professors, among others, often circulating effortlessly between these influential roles [....] Most valued in this world are those people who touch and understand computer code. Most powerful are those who own the code and can employ others to deploy it as they see fit" [38, p. 217].

If "code is law" [115], the argument goes, then those who write the code make the laws and, thus, have the power to govern.

However, when considered through the lens of entanglement, the coding elite do not occur as a pre-figured entity. Rather, this group comes to be constituted in this way because of its interconnections within much broader systems, of "start-ups, [...] large firms, government-sponsored research labs, classrooms" etc. [38, p. 217]. Put differently, an entanglement approach suggests that the coding elite do have significant power, but that that power does not occur solely because of their ability to write (and deploy) code, nor because of their ability to collect data and extract value from it [53]. The coding elite come to be the coding elite because of their intra-actions with algorithmic systems, whose design and implementation are a form of intra-action. The coding elite, the bystander(s), the moderation systems themselves—each is mutually co-constitutive of the others and their attendant properties, including power differentials.

Thus, questions of power are not simply a matter of who gets to make the agential cuts. We should not say, for instance, that the coding elite are responsible for creating bystanders (or other subject positions) of online harassment. Instead, we should ask how conditions of possibility are shaped. Recall that, according to Barad [13], before they enter into mutually co-constitutive entanglements, entities reside in a state of indeterminacy. That indeterminacy, though, should not be conceived of as a uniform prior[8], so to speak. Put differently, a given entity is not equally likely to be constituted as, for example, either a bully or a victim in a given situation. Rather, the particulars of the operant sociotechnical assemblages (weighting of feature vectors, platform moderation rules and policies, training data sets, content flagging by individual users, etc.) shape whether certain entities are more or less likely to be co-constituted in certain ways.

Accounting for power dynamics, then, requires examining how current configurations (of bodies, technologies, organizations, etc.) work to make various future configurations more or less likely. For instance, given a particular content moderation system (and its attendant antecedents, as described above), who is more likely to be constituted as the bully, the bystander, the victim, etc.? And what processes operate to increase or decrease these likelihoods? Achieving the accountabilities necessary to address such questions requires, among other things, strategies by which we might use notions of entanglement to inform algorithmic system design.

[8]In Bayesian statistics, a prior describes the probability distribution over possible values of a variable before any observations are taken into account. In a uniform prior, all possible values of the variable are equally likely.

4.2.1 Designing for Entanglement. To return to the thinking we introduced earlier, subjectivities here are not references to internal or mental concepts of the self, the individual; they are not born solely of one's emotional encounters with, say, a topic model, a content moderation system [181], or a Facebook post [cf. 36]. Rather, we use the term "algorithmic subjectivities" to highlight the ways transient or mutable bodies come into being through the affective push and pull, the affective attunements [100], algorithms make possible. For instance, the controversies surrounding the classification of autism in the DSM-5 undoubtedly trigger highly personal subjective responses. As suggested in the topic modeling case study described above, the families of those living with autism are deeply affected by the subtle changes in text across different versions of the DSM.

Our point, however, is that as algorithms assume increasingly prominent roles within these assemblages, a commensurately algorithmic nature arises in the attendant subjectivities that are probable or even possible. The calculative operations of, for instance, the topic model, translating and enumerating the content that people write, provide an affectual register for understanding the DSM controversies. Through their linguistic interactions, the actors across a discussion platform both work to define and come to be defined by the calculative potential of particular topics, as well as by how affectual or subjective responses can be computed across these topics. The attention we give to algorithms in enlivening subjectivities then is due both to the sheer scale of the assemblages they are coming to constitute, and to the power they exert as part of such entanglements [5].

The conceptual and theoretical considerations described above also suggest paths to adapt our design methods. Consider, for instance, personas and scenarios [42, 50, 134, 146]. Typically, a persona is defined in terms of some combination of individual attributes: demographics, professional goals, personal interests, educational background, family and personal connections, etc. As an example, Nielsen [134] describes a set of personas related with electronic health care, one of which is "Gitte":

> "Age 40. General Practitioner. Mother to two kids, age five and seven. Married to an academic. Lives in a larger provincial town. Works in a shared [medical] practice. […] Attends a choir once a week and jogs in a sports club regularly. Is member [sic] of a book club. Has a conservative/minimalist attitude towards technology and professional tools in general. […] She is smart, lean and takes care to be dressed in well-designed clothes, exquisite jewellery, and newly cut hair." [134, pp. 178-179]

Such descriptions implicitly focus the designer's attention on attributes of the individual. Alternatively, we could articulate a persona less in terms of its individual attributes and more in terms of its entanglements. What are the technological, political, cultural, economic, social, and other entities with which the individual in this persona is being entangled? What are the heterogeneous collections of intra-action by which these interweavings occur? What other possibilities existed for this actor prior to their these intra-actions? How is this persona both subject to and constitutive of these entanglements? In this example, how is "Gitte" as an actor mutually constitutive of the practice where she works or the choir in which she sings? Such an approach would likely require us to articulate not only a series of individual personas but also the relationships among them. Similarly, the scope of a scenario shifts from a particular interaction to a particular intra-action, that is, the agential cut(s) by which the entities involved come to take on their state as those particular entities (and not other types of entities) with those particular properties and relationships (and not other kinds of properties and relationships that might have been possible).

Going further, the notion of entangling could be used to disrupt the very idea of centering in design. HCI is often described in "user-centered" or "human-centered" terms [17, 75, 152, 177]. The conception of algorithmic subjectivities contributed in this paper offers at least two distinct opportunities for future directions of work. First, designers could resist such centerings from the beginning and instead explore entangling as the conceptual locus of design, or what might be called "entanglement-centered" HCI [cf. 79]. Second, designers could resist the concept of centering altogether. Selecting any given concept or entity—humans, users, entanglement, etc.—as the center of design implicitly de-emphasizes other potential centerings. Instead, the notion of entanglement could provide a path for pursuing what might be referred to as “uncentered” design, one that gives primacy to no single entity or concept. Either of these possibilities resonates with other calls for applying more-than-human centered approaches to design [184, 185].

In saying this, we do not suggest that humans are unimportant. Rather, we suggest that designers explore shifting their focus away from specific attributes of individual entities and toward systemic entanglings. In doing so, the designer’s questions change: from “What tasks are made easier or more difficult by this design?” to “What unique entanglings are made more or less probable by this design?”; from “How useful, pleasurable, confusing, rewarding, etc. are users’ interactions with this design?” to “How could this design play various roles in entangling users within broader sociotechnical structures and systems?”; from “How will this design decision cause users to react?” to “How will this design decision enable or preclude different ways of being?” Again, we are ambivalent as to whether entanglement should lie at the center of design processes or should be one consideration among many in a potentially uncentered design.

In either case, such shifts may help resist the impulse to see humans (or users, or communities, or whatever else is being centered) as prefigured entities. That is, rather than taking the human for granted as an entity around which our design centers, an emphasis on design as entangling draws attention to how the design—both as a rendered technical artifact and as a discursive process—plays a role in co-constituting the humanity of those humans. Such points hint toward another potentially fundamental shift, one that pertains to the “human” in human-computer interaction.

4.3 Humanizing

As described throughout this paper, HCI has developed progressively more nuanced ways of talking about humans and their interactions with computers [e.g., 17, 75, 163, 17[illegible]]. At the same time, as noted above, human-centered approaches implicitly posit “human” as a prefigured category. In contrast, the perspective advanced in this paper suggests attending to the processes by which the categories of human, algorithm, and so on come to be.

A variety of prior work has suggested that the ways a machine “thinks” fundamentally differ from the ways that a human “thinks,” or at least that their capacities for and processes of “thinking” differ in scale and affect [2, 176]. For instance, Burrell [37, p. 6] illustrates how a “neural network [trained to recognize digits in human handwriting] doesn’t, for example, break down handwritten digit recognition into subtasks that are readily intelligible to humans, such as identifying a horizontal bar, a closed oval shape, a diagonal line, etc.”. Rather, the visual features to which the network attends appear entirely alien and nearly unrecognizable for a human viewer (Figure 1). Burrell acknowledges that “there is certainly a kind of opacity in the [largely subconscious] human process of character recognition as well” [37, p. 7]. That said, very few of us, when attempting to decipher a number written by someone with poor handwriting, are likely to say that we look for patterns resembling the amorphous blobs in Figure 1.

The case studies above suggest similar examples. For instance, topic modeling represents relationships among topics as mixtures of probability distributions, but this representation bears little resemblance to the ways a human reader might encounter themes within a corpus of documents [see also 156, 179]. Likewise, content moderation on discussion forums and social media platforms may rely on the coupling of humans and algorithmic processes, but such a coupling can sometimes produce frictions. For machines, the application of standards and policies turns on thresholds. In this way, sand dunes can be flagged as nudity [128], and photojournalism can be flagged as child pornography [178]—the kind of error that is more likely in purely algorithmic moderation than with human moderators. In contrast, humans are compelled to read into and respond to the accounts of people’s lives, such as personal interactions with a user guiding exactly what punishment moderators choose to dole out [169].

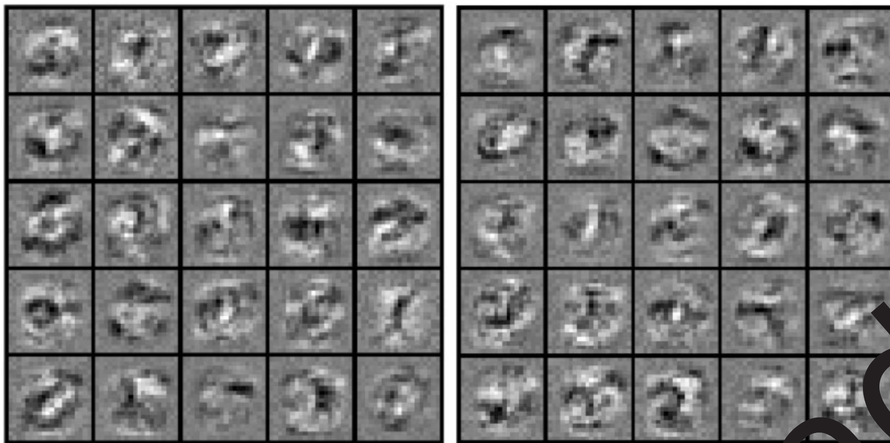

Fig. 1. A visualization of the features used by a neural network trained to recognize handwritten digits [37]. These patterns, Burrell [37] argues, seem to have little resemblance to what humans notice when completing the same task.

At first glance, it may seem like such examples demonstrate two clear categories, human and machine, each with their own styles of thinking. Instead, we suggest that the functioning of algorithmic systems works, almost paradoxically, to define what constitutes the category of human. This defining happens in at least two ways. First, these algorithmic systems both encode and enact definitions of "human." Prior work focuses primarily on how this enactment happens discursively. For instance, Keyes [106] provides a valuable investigation into AI research on diagnosing autism, which often treats social behavior and communication skills, or perhaps an inferred lack thereof, as indicative of autism. Put differently, "autism is treated as oppositional to the traits that 'make' a person a person" [106, p. 14]. Thus, they argue, these systems encode a formulation of autism (and of humanity) in which "autists are portrayed as asocial, fundamentally lacking in the ability to know and understand, and consequently, lacking in agency and personhood" [106, p. 3].

We suggest that similar logics operate both in the technical implementation details and in the practical functioning of algorithmic systems. The Facebook user is defined as much via their interactions with entities on the platform (including friends, organizations, advertisements, etc.) as via the aggregation of their "Likes." Similarly, techniques such as recommender systems encode the user as a mostly-economic actor seeking to maximize selection of objects in a way that aligns with their tastes and preferences [168]. These and other algorithmic systems both construe (and continually reconstrue) what it is to be human.

Second, the very notion of artificial intelligence posits particular types of intelligence as artificial [23, 110, 124, 158]. Put differently, this kind of terminology suggests a distinctly artificial, i.e., algorithmic way of thinking. By labeling it as such, we implicitly define human thinking (and being) as different from, or perhaps even as opposed to, algorithmic thinking (and being). Thus, the manner in which these systems operate works to define what is human via counterexample.

This is not to say that the distinction between algorithms and humans is unnecessary. Put differently, we are not all homogeneous objects [94] or uniform actants within a network [113]. Instead, rather than accepting these as prefigured categories, the position on algorithmic subjectivities advocated here suggests that we—HCI researchers, designers, practitioners, etc.—should attend to the ways that these categories are done, to the processes by which distinctions come to be made between algorithmic and human.

4.3.1 Designing for Human Meaning. In these ways, the functioning of algorithmic systems is deeply, inextricably intermeshed with the human values designed into and interpreted from them [22, 23]. Akin to the cyborg from Haraway [92], the resulting assemblages are a hybrid creation, at once both familiar and strange. They incorporate elements that, on their surface, appear familiar to us—language use, demographics, social connections, etc.—but within algorithmic contexts those elements take on subtly different meanings and significances. For Haraway, and for other accompanying and subsequent work in post-human or more-than-human scholarship [e.g., 31, 100, 175], this hybridity invites a situated reading of meanings, values, (subjective) bodies, and their entanglements. For example, patterns of language use—certain colloquialisms, specific racial slurs, or even particular combinations of emoji—could increase the likelihood that a given piece of content is flagged by a content moderation system [95, 166]. Despite their surface similarity, such features have different meanings for a content moderation classifier and for a human reading (or writing) the content in question. What a human sees as connotative of harassment, an algorithm encodes as weights within a feature vector.

Topic modeling, from the first case study above, offers a prime exemplar. At its core, a topic model is based on counting. The model's probability distributions are fitted to a data set based on patterns in the numbers of times that groups of words co-occur in documents together. To be sure, numbers matter—numbers count, so to speak. At the same time, it can seem almost ridiculous to assert that the most frequently occurring word in a document is also the most important word [141]. Indeed, the single most common word in almost any English document is "the" [56]. This disconnect between frequency and importance contributes to the use of *stop word* lists [149]. These lists include words that occur so frequently as to convey almost no meaning in and of themselves, so they are simply omitted or "stopped out" [119, p. 27] when processing a document. Although such lists are more common and/or more copious with some techniques than with others [119, p. 27], the use of stop words reveals a fundamental divergence: between the frequency counts and probability distributions of topic modeling, and the meanings and significances ascribed by human readers [cf. 155].

Furthermore, computational implementations of topic modeling treat words not as meaningful *per se* but simply as unique tokens. For a topic model, the words "autism" and "autistic" are just as different as the words "apple" and "orange."[9] An occurrence of the word "autism" is only meaningful in so far as it affects the inference of the model's underlying probability distribution. That is, a fundamental disconnect occurs between the computational processing of this language and the human interpretation of it.

To facilitate design processes, it may be productive to foreground such disconnects. For instance, when designing an interactive system built upon topic modeling [e.g., 20, 66], incorporating topic modeling results into early prototypes can help designers, as well as users or co-design participants, interpret the model's results. At the same time, this strategy can give the impression of a system that has some level of human-like understanding about the relationships among words within a topic, such as the "autism" and "autistic" example above. As an alternative, such words could be augmented with the random token ID number that a model assigns to each word, e.g., "token2475-autism" and "token1382-autistic." Doing so offers a simple means to highlight the differential between computational representation and human interpretation, while still enabling productive feedback during design iterations.

On one hand, the specifics of such differentials occur in particular ways for topic modeling. On the other, this same point applies, albeit with different technical details, to other computational approaches for processing natural language.

For instance, during the time that this paper was being finalized for submission, language models informed by distributional semantics [26, 27] and attention mechanisms [183] were developed to address some of the very issues described above. In such large language models (LLMs) and their applications—perhaps most notably

[9] Although techniques such as stemming or lemmatization may enable a model to represent the tokens "autism" and "autistic" as related, doing so often reduces the quality of topic modeling results [167].

ChatGPT [139], but also including compositional word embeddings [123] (e.g., word2vec), bidirectional encoder representations from transformers (BERT) [63], other generative pre-trained transformers (GPT) [148], etc.—the representation of an individual token depends on the context it which it appears, i.e., the other proximate tokens. Furthermore, the tokens in these representations are usually not comprised of natural language words but of subword segments. Returning to the above example, the word “autism” might be represented as two tokens (e.g., “aut” and “ism”), while the word “autistic” might be represented as three tokens (e.g., “aut,” “ist,” and “ic”). Thus, and in contrast to traditional topic modeling approaches [24, 25], such language models often provide very similar representations for such word pairs. This feat is accomplished by leveraging, in addition to subword tokenization, moderately high dimensional so-called semantic spaces, similar to the latent feature spaces described above [63, 123]. In such approaches, an individual word token (or a subword token, or a complete sentence, or an entire document, etc.) can be represented as a vector in that space. The similarity between any two word tokens, subword tokens, sentences, documents, etc. is then based on the cosine of the angle between their vector representations. These representations are derived from training the model on large corpora (usually billions of words from combinations of web crawls, books, and other texts), such that words occurring in similar contexts will have representations in similar regions of the semantic space.

At first glance, this general approach takes clever advantage of the assertion from early Wittgenstein [193] that a word only has meaning within the context of some statement, some logical proposition. Later Wittgenstein [194], however, developed the notion of language-games, asserting that a word (or a complete sentence, or an entire document) derives its meaning from use in the context of some human activity. As a simple example, the single-word sentence “Fly!” has drastically different meanings when uttered by an irritated customer lifting their glass of Chablis, by the pitcher on a baseball diamond, or by a grey wizard dangling from a precipice. This point goes beyond simple polysemy or word sense disambiguation (for which some computational approaches have been developed [e.g., 85, 130, 154]), and beyond the fact that the other surrounding words describing each of these situations will differ. The point is that the processes by which humans actively work to construct and contrast the meanings of these words [cf. 151] likely bears little resemblance to computing the cosines of angles between vector representations based on subword tokens. Yet it is those distances between feature vectors that can make the difference between whether or not a given piece of content is flagged for moderation.

Thus, future work that seeks to advance our understandings about experiences around algorithmic systems must explicitly consider the relationships among the pluralities of meaning in these systems, the computational mechanisms that help give rise to those meanings, and the sources of agency involved with enacting those meanings [15, 65, 165, 195]. As Barad [13, p. 353] puts it, “phenomena—whether lizards, electrons, or humans—exist only as a result of, and as part of, the world’s ongoing intra-activity, its dynamic and contingent differentiation into specific relationalities.” None of these phenomena, entities, etc. can be reductively treated as standing alone or viewed from outside their contexts. Rather, they must be understood as always becoming through their entangled relations. Our analysis here focuses on the specific context of algorithmic systems, and how these complex relations and interplays—among probability distributions, feature vector weights, data point labels, exercises of power, attributions of meaning, etc.—afford a distinct role in co-producing the category of “human.”

Attention to such dynamics is equally important for researchers investigating algorithmic systems and for designers implementing algorithmic systems. As researchers, we should pause before interpreting these systems’ constituent elements in a manner similar to the varying manners in which we humans might interpret those same elements in other, non-algorithmic contexts. For instance, word-based features that are highly informative for a toxicity classifier do not necessarily have the same meanings and connotations in the context of that classifier as when a human reads those same words in a social media post. As designers, we can and arguably should attend to the ways that the internal functionings of the algorithmic systems that we implement work implicitly to define what constitutes humanity. For instance, the features used to curate a social media news feed not only can influence perceptions of how close we are to specific individuals [72], they can also work to reshape how we perceive the constitution, enactment, and performance of human closeness.

5 IMPLICATIONS AND CONCLUSION

One gut reaction to many of the problems algorithms surface might be to pare back somehow [cf. 21], to reduce the extent to which algorithms entangle with the enlivening of subjectivities. Perhaps we should rethink the distributions of labor, seeking to return the weight of agency to humans in cases similar to those described above? In contrast, more techno-centric solutions suggest improving the transparency of algorithmic processes, building tools that explain how classifications are produced and decisions made [67, 164]. Such suggestions are predicated upon the existence of a neutral, objective, "God's Eye" perspective [90], i.e., the assumption that there exist ways to debias classification schemes [28] and the choices afforded through them [52].

Neither of these solutions do much to take seriously the proliferation of algorithmic processes we have sought to capture here. They presume there are worlds where humans and machines are separable, and decisions can be made based on just the data or facts, detached from situated, human experience. What our cases show, however (alongside a long history of scholarship cited throughout this paper), is that we always already proliferate in worlds that inexorably enmesh human and machine. To define bodies and experiences is never without a politics of what counts as normatively valued. To ban hate-speech and its perpetrators, and to train machines to do so, is to participate in processes of defining hate and how one performs it.

The two-part proposal we wish to make in closing, and one we believe opens up a space for research and for design in HCI, involves attending, first, to the worlds that are being made possible, and, second, to the worlds we might want to make possible, in and through algorithmic subjectivities. Doing so requires more than simply replacing the term "user" with the term "subject" or "subjectivity" [14, 17]. It requires an overall change of orientation, away from seeing the user as given and toward seeing the user, as well as numerous other possible relations, as constantly co-produced in and through design. It is the processes by which this co-producing occurs to which the field of HCI should attend.

That said, this paper provides neither a conclusive endpoint nor precise, prescriptive directives for conducting such work. Instead, the paper offers what we hope constitutes meaningful progress, especially progress upon which others can build—conceptually, empirically, methodologically, theoretically, etc. For instance, although potentially informative, it is unlikely that the list of qualities enumerated here is definitive. Indeed, there exist numerous kinds of algorithmic systems that are not directly addressed in this paper—computer vision algorithms in facial recognition [170, 173], automated task assignment in gig work [6, 190, 196], curation of social media news feeds [71, 72], risk assessment and prediction in criminal justice [33, 48], and many others. Such contexts are absolutely relevant areas in which future work can develop further this paper's core conceptual contributions.

The word "develop" here is crucial. There will be differences among these contexts and others, e.g., in the technical details by which inferences occur, in the specific entanglings among algorithmic inference and systems of classification, or in how the category of "human" is mutually constituted among various actors. Furthermore, such work will need to explore the suitability of and understandings generated with various methodological approaches, including both long-established methods and novel innovations. Similarly, there will be other qualities, beyond the three we have offered, to which future researchers and designers will productively direct their attentions. Regardless, the key—and the first part of our closing proposal—is to ask **what subjectivities are being enlivened in a particular case, and in what ways (or through what relational entanglements) were these subjectivities made possible?**

To be clear, we are not calling for one or more specific studies, or even a series of particular studies. Instead, we are advocating for a larger program of research. This larger program is unlikely to revolve around any single conceptualization of algorithmic subjectivities that remains fixed in the long term. If we claim to offer a finalized, definitive conceptualization, we indirectly undermine the ability for future work to account fully for the continual on-goingness of these entanglements. The perpetually evolving ways that algorithmic systems permeate evermore facets of public and private life mandate commensurate perpetual evolution in the concepts we use to understand these relational arrangements. Thus, we should not expect any individual piece of future work on algorithmic subjectivities to provide a henceforth-definitive account or a fully comprehensive picture. Instead, we should expect such work to articulate how its connection with this larger program works to develop further our understandings in ways that help account for these continual becomings.

Beyond a program of studying algorithmic systems and their subjectivities, we also seek to draw attention to their relations with design. If we are to accept that the design of algorithms and the interfaces that provide access to them are part of doing bodies (i.e., enlivening certain subjectivies), then we must also acknowledge that we are in a position to consider what other kinds of subjectivities we wish to enliven and how. This paper eschews strong injunctions about which kinds of subjectivities might be more or less preferable. Put differently, rather than prescribe answers, we offer conceptual tools that researchers, designers, practitioners, etc. can use to explore such questions. Again, Haraway's cyborg [92] offers an instructive example. The cyborg can legitimately be read as a technocentric imaginary, turning on and amplifying highly masculinized and Western-oriented versions of technoscientific progress and innovation. The question Haraway famously asks, however, is *what other imaginaries might be possible*? What technically and infrastructurally is afforded in the cyborg that opens up the conditions for more equitable and distributed forms of modest, global flourishing?

Analogously, we hope the work we present offers designers a speculative leap that invites asking "what else?" This second part of our closing proposal, then, is to ask **what other relations and attendant subjectivities could be given the chance, here?** How might we approach designing for these relations and subjectivities in ways that are more open or expansive (i.e., that allow for a greater variety of different entanglements among the actors) and that are more generous (i.e., that recognise how all actors have had a say on the outcome)? Put differently, how might design [*a la* 186] open up spaces for envisioning and for enacting other possible worlds [116, 162]? Despret [59, p. 129] views this expansive, more-than-human framing of entangled agencies and subjectivities as "an adventure in the course of which subjectivities overlap, are transformed, actualized and extended." Subjectivity here is never isolated to the individual but always caught up in the give and take among actors of all kinds (human and otherwise). We contend that simply the consideration of, let alone the pursuit of, other imaginaries requires applying similar perspectives to understanding and to designing around algorithmic subjectivities.

For HCI, this leaves us in a place to start imagining those conditions of possibility that might be afforded in the algorithmic systems we build. We thus close with three suggestions for how future work might go about doing so. First, we suggest algorithmic subjectivities as an important area for empirical investigation, as outlined above. Such work should attend not only to lived, subjective experiences, but also to how the intricate entanglements among heterogeneous actors work in concert to make certain subjectivities possible and not others. In other words, the question is not only *what* subjectivities are enlivened by algorithmic systems but also *how* algorithmic systems do subjectivities—what are the processes by which algorithmic systems infer, entangle, and humanize various experiences and entities within broader heterogeneous assemblages? Second, through such empirical work, we can iteratively develop a collective understanding of the roles that various design decisions play. The "design" to which we refer here encompasses not only interfaces or interactions, i.e., the typical locus of inquiry for HCI. It also includes the design of the classificatory schemes on which algorithmic systems are predicated, the mathematical formalisms used to encode algorithmic models, the organizational and/or governmental policies dictating how algorithmic systems can or should be used and by whom, among other things. Such an expansion of how we scope design in HCI is necessary if we wish to take seriously the implications of algorithmic subjectivities. Third, as we have made clear throughout, a central goal should be to envision how things could be otherwise. How might designers make different choices about feature selection, use of latent representation spaces, model (hyper)parameters, tuning and optimization, etc., and how might those choices make certain worlds less likely and other worlds more so? How might we craft algorithmic systems not only to shape the interactions that people have with them but also to influence the imaginaries that make worlds possible? Building an understanding of how the multifaceted aspects of design come to enliven algorithmic subjectivities becomes most impactful when it enables us to see, and perhaps even to enact, these other worlds. We hope this paper will facilitate others who may envision alternative conditions of possibility.

ACKNOWLEDGMENTS

For constructive Feedback on prior drafts, thanks to Jeff Bardzell, Shaowen Bardzell, Michael Ann DeVito, Jessica Fueston, and the anonymous reviewers. This material is based in part on work supported by the NSF under Grant #IIS-1844901.

REFERENCES

[1] Simon Adler. 2018. Post No Evil. https://www.wnycstudios.org/podcasts/radiolab/articles/post-no-evil
[2] Philip E. Agre. 1997. *Computation and Human Experience.* Cambridge University Press, Cambridge.
[3] Sara Ahmed. 2006. *Queer Phenomelogoy: Orientations, Objects, Others.* Duke University Press, Durham, NC.
[4] Sara Ahmed. 2010. *The Promise of Happiness.* Duke University Press, Durham, NC.
[5] Ali Alkhatib and Michael Bernstein. 2019. Street-Level Algorithms: A Theory at the Gaps between Policy and Decisions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, Glasgow, 530. https://doi.org/10.1145/3290605.3300760
[6] Ali Alkhatib, Michael S. Bernstein, and Margaret Levi. 2017. Examining Crowd Work and Gig Work Through The Historical Lens of Piecework. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, San Jose, CA, 4599–4616. https://doi.org/10.1145/3025453.3025974
[7] American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4. ed., 7. print ed.). American Psychiatric Association, Washington, DC.
[8] American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR* (4th ed., text revision ed.). American Psychiatric Association, Washington, DC.
[9] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, Glasgow, Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300233
[10] Mike Ananny. 2011. The Curious Connection between Apps for Gay Men and Sex Offenders. *The Atlantic* (April 2011).
[11] Emiliana Armano, Marco Briziarelli, Joseph Flores, and Elisabetta Risi. 2022. Platforms, Algorithms and Subjectivities: Active Combination and the Extraction of Value Process – An Introductory Essay. In *Digital Platforms and Algorithmic Subjectivities*, Emiliana Armano, Marco Briziarelli, and Elisabetta Risi (Eds.). Vol. 24. University of Westminster Press, 1–18. jstor:j.ctv319wpvm.4
[12] Karen Barad. 2003. Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs* 28, 3 (2003), 801–831. https://doi.org/10.1086/345321
[13] Karen Barad. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning.* Duke University Press, Durham, NC.
[14] Jeffrey Bardzell and Shaowen Bardzell. 2015. The User Reconfigured: On Subjectivities of Information. In *Decennial Aarhus Conference on Human Centered Computing.* Aarhus, Denmark, 133–144. https://doi.org/10.7146/aahcc.v1i1.21298
[15] Shaowen Bardzell. 2010. Feminist HCI: Taking Stock and Outlining an Agenda for Design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, Atlanta, GA, 1301–1310.
[16] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104, 3 (2016), 671–732. https://doi.org/10.15779/Z38BG31
[17] Eric P. S. Baumer and Jed R. Brubaker. 2017. Post-Userism. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, Denver, CO, 6291–6303. https://doi.org/10.1145/3025453.3025740
[18] Eric P. S. Baumer and Micki McGee. 2019. Speaking on Behalf of: Representation, Delegation, and Authority in Computational Text Analysis. In *Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES).* AAAI/ACM, Honolulu, HI, 163–169. https://doi.org/10.1145/3306618.3314292
[19] Eric P. S. 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 (JASIST)* 68, 6 (June 2017), 1397–1410. https://doi.org/10.1002/asi.23786

[20] Eric P. S. 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 (April 2020), 452–480. https://doi.org/10.1080/07370024.2020.1734460

[21] Eric P. S. Baumer and M. Six Silberman. 2011. When the Implication Is Not to Design (Technology). In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Vancouver, BC, Canada, 2271–2274. https://doi.org/10.1145/1978942.1979275

[22] Alan F. Blackwell. 2015. Interacting with an Inferred World: The Challenge of Machine Learning for Humane Computer Interaction. *Aarhus Series on Human Centered Computing* 1, 1 (Oct. 2015), 12–12. https://doi.org/10.7146/aahcc.v1i1.21197

[23] Alan F. Blackwell. 2019. Objective Functions: (In)Humanity and Inequity in Artificial Intelligence. *HAU: Journal of Ethnographic Theory* 9, 1 (March 2019), 137–146. https://doi.org/10.1086/703871

[24] David M. Blei. 2012. Probabilistic Topic Models. *Commun. ACM* 55, 4 (April 2012), 77–84. https://doi.org/10.1145/2133806.2133826

[25] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, Jan (2003), 993–1022.

[26] Gemma Boleda. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics* 6, 1 (2020), 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303

[27] Gemma Boleda and Aurélie Herbelot. 2016. Formal Distributional Semantics: Introduction to the Special Issue. *Computational Linguistics* 42, 4 (Dec. 2016), 619–635. https://doi.org/10.1162/COLI_a_00261

[28] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings. In *Advances in Neural Information Processing Systems (NeurIPS)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.). Curran Associates, Inc., Barcelona, Spain, 4349–4357.

[29] Jacob T. Borodovsky, Lisa A. Marsch, and Alan J. Budney. 2018. Studying Cannabis Use Behaviors With Facebook and Web Surveys: Methods and Insights. *JMIR Public Health and Surveillance* 4, 2 (May 2018), e9408. https://doi.org/10.2196/publichealth.9408

[30] Geoffrey C. Bowker and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences.* MIT Press, Cambridge, MA.

[31] Rosi Braidotti. 2019. *Posthuman Knowledge.* Polity Press, Cambridge.

[32] Stacy M. Branham, Steve H. Harrison, and Tad Hirsch. 2012. Expanding the Design Space for Intimacy: Supporting Mutual Reflection for Local Partners. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS)*. ACM, Newcastle, 220–223. https://doi.org/10.1145/2317956.2317990

[33] Sarah Brayne. 2020. *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford University Press, Oxford, New York.

[34] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness.* Duke University Press, Durham, NC.

[35] Hrönn Brynjarsdóttir, Maria Håkansson, James Pierce, Eric P. S. Baumer, Carl DiSalvo, and Phoebe Sengers. 2012. Sustainably Unpersuaded: How Persuasion Narrows Our Vision of Sustainability. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Austin, TX, 947–956.

[36] Taina Bucher. 2017. The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms. *Information, Communication & Society* 20, 1 (2017), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

[37] Jenna Burrell. 2016. How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3, 1 (Jan. 2016), 2053951715622512. https://doi.org/10.1177/2053951715622512

[38] Jenna Burrell and Marion Fourcade. 2021. The Society of Algorithms. *Annual Review of Sociology* 47, 1 (2021), 213–237. https://doi.org/10.1146/annurev-soc-090820-020800

[39] Judith Butler. 2001. Doing Justice to Someone: Sex Reassignment and Allegories of Transsexuality. *GLQ: A Journal of Lesbian and Gay Studies* 7, 4 (Sept. 2001), 621–636.

[40] Robyn Caplan. 2018. *Content or Context Moderation?* Technical Report. Data & Society.

[41] Sarah J. Carrington, Rachel G. Hunt, Jarymke Maljaars, Ann Le Couteur, Judith Gould, Lorna Wing, Ilse Noens, Ina Van Berckelaer-Onnes, and Susan R. Leekam. 2014. DSM-5 Autism Spectrum Disorder: In Search of Essential Behaviours for Diagnosis. *Research in Autism Spectrum Disorders* 8, 6 (June 2014), 701–715. https://doi.org/10.1016/j.rasd.2014.03.017

[42] John M. Carroll. 2000. *Making Use: Scenario-Based Design of Human-Computer Interactions.* The MIT Press, Cambridge, MA.

[43] Simon Caton and Christian Haas. 2023. Fairness in Machine Learning: A Survey. *Comput. Surveys* (Aug. 2023). https://doi.org/10.1145/3616865

[44] Stevie Chancellor, Eric P. S. Baumer, and Munmun De Choudhury. 2019. Who Is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 147:1–147:32. https://doi.org/10.1145/3359249

[45] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #Thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. ACM, San Francisco, CA, 1201–1213. https://doi.org/10.1145/2818048.2819963

[46] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit. *ACM Transactions on Computer-Human Interaction* 29, 4 (March 2022), 29:1–29:26. https://doi.org/10.1145/3490499

[47] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. ACM, Portland, OR, 1217–1230. https://doi.org/10.1145/2998181.2998213

[48] Angèle Christin. 2017. Algorithms in Practice: Comparing Web Journalism and Criminal Justice. *Big Data & Society* 4, 2 (Dec. 2017), 2053951717718855. https://doi.org/10.1177/2053951717718855

[49] Josh Constine. 2011. Facebook Adds Keyword Moderation and Profanity Blocklists to Pages. https://www.adweek.com/digital/keyword-moderation-profanity-blocklist/.

[50] Geoff Cooper and John Bowers. 1995. Representing the User: Notes on the Disciplinary Rhetoric of HCI. In *The Social and Interactional Dimensions of Human-Computer Interfaces*, Peter J. Thomas (Ed.). Cambridge University Press, Cambridge, 48–66.

[51] Sam Corbett-Davies, Johann D. Gaebler, Hamed Nilforoshan, Ravi Shroff, and Sharad Goel. 2023. The Measure and Mismeasure of Fairness. *Journal of Machine Learning Research* 24, 312 (2023), 1–117.

[52] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv* (July 2018).

[53] Nick Couldry and Ulises A. Mejias. 2021. *The Costs of Connection – How Data Is Colonizing Human Life and Appropriating It for Capitalism.* Stanford University Press, Redwood City, CA.

[54] Constantinos K. Coursaris and Ming Liu. 2009. An Analysis of Social Support Exchanges in Online HIV/AIDS Self-Help Groups. *Computers in Human Behavior* 25, 4 (July 2009), 911–918. https://doi.org/10.1016/j.chb.2009.03.006

[55] Kate Crawford and Tarleton Gillespie. 2016. What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint. *New Media & Society* 18, 3 (March 2016), 410–428. https://doi.org/10.1177/1461444814543163

[56] Mark Davies. 2008. Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

[57] Tisha Dejmanee. 2013. Bodies of Technology: Performative Flesh, Pleasure and Subversion in Cyberspace. *Gender Questions* 1, 1 (Jan. 2013), 3–17.

[58] Gilles Deleuze and Félix Guattari. 1987. *A Thousand Plateaus: Capitalism and Schizophrenia.* University of Minnesota Press, Minneapolis. Translation of: Mille plateaux, v. 2 of Capitalisme et schizophrénie A companion volume to Anti-Oedipus : capitalism and schizophrenia Includes index.

[59] Vinciane Despret. 2008. The Becomings of Subjectivity in Animal Worlds. *Subjectivity* 23, 1 (July 2008), 123–139. https://doi.org/10.1057/sub.2008.15

[60] Vinciane Despret. 2013. From Secret Agents to Interagency. *History and Theory* 52, 4 (2013), 29–44. https://doi.org/10.1111/hith.10686

[61] Michael Ann DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Montréal, QC, 120:1–120:12. https://doi.org/10.1145/3173574.3173694

[62] Michael Ann DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Denver, Colorado, USA, 3163–3174. https://doi.org/10.1145/3025453.3025659

[63] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*. ACL, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[64] Thiago Dias Oliva, Dennys Marcelo Antonialli, and Alessandra Gomes. 2021. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture* 25, 2 (April 2021), 700–732. https://doi.org/10.1007/s12119-020-09790-w

[65] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism.* MIT Press, Cambridge, MA.

[66] Karthik Dinakar, Jackie Chen, Henry Lieberman, Rosalind Picard, and Robert Filbin. 2015. Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI)*. ACM, Atlanta, GA, 417–426. https://doi.org/10.1145/2678025.2701395

[67] F. K. Došilović, M. Brčić, and N. Hlupić. 2018. Explainable Artificial Intelligence: A Survey. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 0210–0215. https://doi.org/10.23919/MIPRO.2018.8400040

[68] Paul Dourish. 2006. Implications for Design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Montréal, QC, 541–550. https://doi.org/10.1145/1124772.1124855

[69] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning As a Design Material. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Denver, CO, 278–288. https://doi.org/10.1145/3025453.3025739

[70] Dmitry Epstein, Merrill C. Roth, and Eric P. S. Baumer. 2014. It's the Definition, Stupid! Framing of Online Privacy in the Internet Governance Forum Debates. *Journal of Information Policy* 4 (2014), 144. https://doi.org/10.5325/jinfopoli.4.2014.0144 jstor:10.5325/jinfopoli.4.2014.0144

[71] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "Like" It, Then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, San Jose, CA, 2371–2382. https://doi.org/10.1145/2858036.2858494

[72] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning About Invisible Algorithms in News Feeds. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Seoul, 153–162. https://doi.org/10.1145/2702123.2702556

[73] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Association for Computing Machinery, Glasgow, 1–14. https://doi.org/10.1145/3290605.3300724

[74] Eran Fisher. 2022. *Algorithms and Subjectivity: The Subversion of Critical Knowledge.* Routledge, London. https://doi.org/10.4324/9781003196563

[75] Jodi Forlizzi. 2018. Moving Beyond User-centered Design. *Interactions* 25, 5 (Aug. 2018), 22–23. https://doi.org/10.1145/3239558

[76] Michel Foucault. 1965. *Madness and Civilization* (1988 ed.). Vintage Books, New York.

[77] Michel Foucault. 1977. *Discipline and Punish: The Birth of the Prison.* Vintage Books, New York.

[78] Michel Foucault. 1988. Technologies of the Self. Lectures at the University of Vermont, October 1982. In *Technologies of the Self*, Luther H. Martin, Huck Gutman, and Patrick H. Hutton (Eds.). University of Massachusetts Press, 16–49.

[79] Christopher Frauenberger. 2019. Entanglement HCI The Next Wave? *ACM Transactions on Computer-Human Interaction* 27, 1 (Nov. 2019), 2:1–2:27. https://doi.org/10.1145/3364998

[80] Tarleton Gillespie. 2013. The Relevance of Algorithms. In *Media Technologies*, Tarleton Gillespie, Pablo Bockzkowski, and Kirsten Foot (Eds.). MIT Press, Cambridge, MA, 167–193.

[81] Tarleton Gillespie. 2020. Content Moderation, AI, and the Question of Scale. *Big Data & Society* 7, 2 (July 2020), 2053951720943234. https://doi.org/10.1177/2053951720943234

[82] Temple Grandin and Richard Panek. 2013. *The Autistic Brain: Thinking Across the Spectrum.* Houghton Mifflin Harcourt.

[83] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work.* Houghton Mifflin Harcourt, Boston.

[84] Wei Guo and Aylin Caliskan. 2021. Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. ACM, Virtual Event, 122–133. https://doi.org/10.1145/3461702.3462536

[85] Weiwei Guo and Mona Diab. 2011. Semantic Topic Models: Combining Word Distributional Statistics and Dictionary Definitions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (EMNLP)*. ACL, Edinburgh, Scotland, UK., 552–561.

[86] Rahul Kumar Gupta, Ritu Agarwalla, Bukya Hemanth Naik, Joythish Reddy Evuri, Apil Thapa, and Thoudam Doren Singh. 2022. Prediction of Research Trends Using LDA Based Topic Modeling. *Global Transitions Proceedings* 3, 1 (June 2022), 298–304. https://doi.org/10.1016/j.gltp.2022.03.015

[87] Kevin D. Haggerty and Richard V. Ericson. 2000. The Surveillant Assemblage. *The British Journal of Sociology* 51, 4 (2000), 605–622. https://doi.org/10.1080/00071310020015280

[88] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 466:1–466:35. https://doi.org/10.1145/3479610

[89] Blake Hallinan and Jed R. Brubaker. 2021. Living With Everyday Evaluations on Social Media Platforms. *International Journal of Communication* 15, 0 (March 2021), 19.

[90] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599. https://doi.org/10.2307/3178066 jstor:3178066

[91] Donna J. Haraway. 1991. A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In *Simians, Cybors, and Women: The Reinvention of Nature*. Routledge, New York, 149–182.

[92] Donna J. Haraway. 1991. *Simians, Cyborgs, and, Women: The Reinvention of Nature.* Routledge, New York.

[93] Donna J. Haraway. 2013. SF: Science Fiction, Speculative Fabulation, String Figures, So Far. *Ada: A Journal of Gender, New Media, and Technology* 3 (Nov. 2013).

[94] Graham Harman. 2018. *Object-Oriented Ontology: A New Theory of Everything.* Pelican Books, London.

[95] Camille Harris, Matan Halevy, Ayanna Howard, Amy Bruckman, and Diyi Yang. 2022. Exploring the Role of Grammar and Word Choice in Bias Toward African American English (AAE) in Hate Speech Classification. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. ACM, Seoul, Republic of Korea, 789–798. https://doi.org/10.1145/3531146.3533144

[96] Denise L. Haynie, Tonja Nansel, Patricia Eitel, Aria Davis Crump, Keith Saylor, Kai Yu, and Bruce Simons-Morton. 2001. Bullies, Victims, and Bully/Victims:: Distinct Groups of At-Risk Youth. *The Journal of Early Adolescence* 21, 1 (Feb. 2001), 29–49. https://doi.org/10.1177/0272431601021001002

[97] Joanne Hinds and Adam N. Joinson. 2018. What Demographic Attributes Do Our Digital Footprints Reveal? A Systematic Review. *PLOS ONE* 13, 11 (Nov. 2018), e0207112. https://doi.org/10.1371/journal.pone.0207112

[98] Anna Lauren Hoffmann. 2019. Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse. *Information, Communication & Society* 22, 7 (June 2019), 900–915. https://doi.org/10.1080/1369118X.2019.1573912

[99] Gregory Hollin, Isla Forsyth, Eva Giraud, and Tracey Potts. 2017. (Dis)Entangling Barad: Materialisms and Ethics. *Social Studies of Science* 47, 6 (2017), 918–941.

[100] Carla Hustak and Natasha Myers. 2012. Involutionary Momentum: Affective Ecologies and the Sciences of Plant/Insect Encounters. *differences* 23, 3 (Dec. 2012), 74–118. https://doi.org/10.1215/10407391-1892907

[101] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does Transparency in Moderation Really Matter? User Behavior After Content Removal Explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 150:1–150:27. https://doi.org/10.1145/3359252

[102] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-based Online Communities on Discord. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 55:1–55:23. https://doi.org/10.1145/3359157

[103] Jialun Aaron Jiang, Morgan Klaus Scheuerman, Casey Fiesler, and Jed R. Brubaker. 2021. Understanding International Perceptions of the Severity of Harmful Content Online. *PLOS ONE* 16, 8 (Aug. 2021), e0256762. https://doi.org/10.1371/journal.pone.0256762

[104] Matthew L. Jockers. 2013. *Macroanalysis: Digital Methods & Literary History.* University of Illinois Press, Chicago.

[105] Christ Plauché Johnson and Scott M. Myers. 2008. CHAPTER 15 - Autism Spectrum Disorders. In *Developmental-Behavioral Pediatrics*, Mark L. Wolraich, Dennis D. Drotar, Paul H. Dworkin, and Ellen C. Perrin (Eds.). Mosby, Philadelphia, 519–577. https://doi.org/10.1016/B978-0-323-04025-9.50018-0

[106] Os Keyes. 2020. Automating Autism: Disability, Discourse, and Artificial Intelligence. *The Journal of Sociotechnical Critique* 1, 1 (Dec. 2020). https://doi.org/10.25779/89bj-j396

[107] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* ACM, San Jose, CA, 1152–1156. https://doi.org/10.1145/2858036.2858356

[108] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences (PNAS)* 111, 24 (2014), 8788–8790. https://doi.org/10.1073/pnas.1320040111

[109] Robert E. Kraut and Paul Resnick. 2011. *Building Successful Online Communities: Evidence-Based Social Design.* Massachusetts Institute of Technology, Cambridge, MA.

[110] Ray Kurzweil. 1990. *The Age of Intelligent Machines.* MIT Press, Cambridge, MA.

[111] landoflobsters. 2018. Revamping the Quarantine Function.

[112] Bruno Latour. 1994. On Technical Mediation: Philosophy, Sociology, Genealogy. *Common Knowledge* 3, 2 (1994), 29–64.

[113] Bruno Latour. 1999. On Recalling Ant. *The Sociological Review* 47, 1_suppl (May 1999), 15–25. https://doi.org/10.1111/j.1467-954X.1999.tb03480.x

[114] Min Kyung Lee, Alexandros Psomas, Ariel D. Procaccia, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, and Siheon Lee. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–35. https://doi.org/10.1145/3359283

[115] Lawrence Lessig. 2000. Code Is Law. https://harvardmagazine.com/2000/01/code-is-law-html.

[116] David Lewis. 1978. Truth in Fiction. *American Philosophical Quarterly* 15, 1 (1978), 37–46.

[117] Renkai Ma and Yubo Kou. 2022. "I'm Not Sure What Difference Is between Their Content and Mine, Other than the Person Itself": A Study of Fairness Perception of Content Moderation on YouTube. *Proceedings of the ACM on Human-Computer Interaction* (Oct. 2022). https://doi.org/10.1145/3555150

[118] Julie Maitland and Matthew Chalmers. 2011. Designing for Peer Involvement in Weight Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI).* ACM, Vancouver, BC, 315–324. https://doi.org/10.1145/1978942.1978988

[119] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval.* Cambridge University Press, Cambridge.

[120] Lucas Matney. 2019. Reddit Quarantines Its Biggest Headache.

[121] Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit.

[122] Jie Mei, Christian Desrosiers, and Johannes Frasnelli. 2021. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. *Frontiers in Aging Neuroscience* 13 (2021). https://doi.org/10.3389/fnagi.2021.633752

[123] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Advances in Neural Information Processing Systems (NeurIPS)*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., Lake Tahoe, NV, 3111–3119.

[124] Marvin Minsky (Ed.). 1968. *Semantic Information Processing.* MIT Press, Cambridge, MA.

[125] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society* 3, 2 (Dec. 2016). https://doi.org/10.1177/2053951716679679

[126] Torin Monahan. 2017. Regulating Belonging: Surveillance, Inequality, and the Cultural Production of Abjection. *Journal of Cultural Economy* 10, 2 (March 2017), 191–206. https://doi.org/10.1080/17530350.2016.1273843

[127] Michael Muller, Shion Guha, Eric P. S. Baumer, David Mimno, and N. Sadat Shami. 2016. Machine Learning and Grounded Theory Method: Convergence, Divergence, and Combination. In *Proceedings of the ACM Conference on Supporting Group Work (GROUP)*. Sanibel Island, FL.

[128] Margi Murphy. 2017. Artificial Intelligence Will Detect Child Abuse Images to Save Police from Trauma. *The Telegraph* (Dec. 2017).

[129] Sarah Myers West. 2018. Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. https://doi.org/10.1177/1461444818773059

[130] Roberto Navigli. 2009. Word Sense Disambiguation: A Survey. *Comput. Surveys* 41, 2 (Feb. 2009), 10:1–10:69. https://doi.org/10.1145/1459352.1459355

[131] Laura K. Nelson. 2017. Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research* (Nov. 2017), 3–42. https://doi.org/10.1177/0049124117729703

[132] Mark W. Newman, Debra Lauterbach, Sean A. Munson, Paul Resnick, and Margaret E. Morris. 2011. It's Not That I Don't Have Problems, i'm Just Not Putting Them on Facebook: Challenges and Opportunities in Using Online Social Networks for Health. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*. ACM, Hangzhou, China, 341–350. https://doi.org/10.1145/1958824.1958876

[133] Casey Newton. 2019. The Secret Lives of Facebook Moderators in America. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.

[134] Lene Nielsen. 2004. *Engaging Personas and Narrative Scenarios.* Ph. D. Dissertation. Department of Informatics, Copenhagen Business School, Copenhagen, Denmark.

[135] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press, New York.

[136] Walter J. Ong. 1982. *Orality and Literay: The Technologizing of the Word* (2002 ed.). Routledge, New York.

[137] Cathy O'Niel. 2016. *Weapons of Math Destruction.* Crown, New York.

[138] Yotam Ophir, Dror Walter, and Eleanor R Marchant. 2020. A Collaborative Way of Knowing: Bridging Computational Communication Research and Grounded Theory Ethnography. *Journal of Communication* 70, 3 (June 2020), 447–472. https://doi.org/10.1093/joc/jqaa013

[139] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training Language Models to Follow Instructions with Human Feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*. New Orleans, LA.

[140] Orestis Papakyriakopoulos, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. 2020. Bias in Word Embeddings. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*. ACM, Barcelona, Spain, 446–457. https://doi.org/10.1145/3351095.3372843

[141] Samir Passi and Steven Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. ACM, Portland, OR, 2436–2447. https://doi.org/10.1145/2998181.2998331

[142] Dana Pessach and Erez Shmueli. 2022. A Review on Fairness in Machine Learning. *Comput. Surveys* 55, 3 (Feb. 2022), 51:1–51:44. https://doi.org/10.1145/3494672

[143] Davor Petreski and Ibrahim C. Hashim. 2022. Word Embeddings Are Biased. But Whose Bias Are They Reflecting? *AI & Society* (May 2022). https://doi.org/10.1007/s00146-022-01443-w

[144] S. Piantadosi, D. P. Byar, and S. B. Green. 1988. The Ecological Fallacy. *American Journal of Epidemiology* 127, 5 (May 1988), 893–904. https://doi.org/10.1093/oxfordjournals.aje.a114892

[145] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *Proceedings of the USENIX}Security Symposium.* USENIX Association, Vancouver, BC, 935–951.

[146] John Pruitt and Jonathan Grudin. 2003. Personas: Practice and Theory. In *Proceedings of the ACM Conference on Designing for User Experiences (DUX)*. ACM, San Francisco, CA, 1–15. https://doi.org/10.1145/997078.997089

[147] Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Seoul, 173–182. https://doi.org/10.1145/2702123.2702174

[148] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. *Improving Language Understanding by Generative Pre-Training.* Technical Report. OpenAI, San Francisco, CA.

[149] Anand Rajaraman and Jeffrey David Ullman. 2011. Data Mining. In *Mining of Massive Datasets.* Cambridge University Press, Cambridge, 1–17. https://doi.org/10.1017/CBO9781139058452.002

[150] Reddit. [n. d.]. Quarantined Subreddits. https://reddit.zendesk.com/hc/en-us/articles/360043069012-Quarantined-Subreddits.
[151] Michael J. Reddy. 1979. The Conduit Metaphor: A Case of Frame Conflict in Our Language about Language. In *Metaphor and Thought*, Andrew Ortony (Ed.). Cambridge University Press, Cambridge, 164–201.
[152] Johan Redström. 2006. Towards User Design? On the Shift from Object to User as the Subject of Design. *Design Studies* 27, 2 (March 2006), 123–139. https://doi.org/10.1016/j.destud.2005.06.001
[153] Jason Ren, Russell Kunes, and Finale Doshi-Velez. 2020. Prediction Focused Topic Models via Feature Selection. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 4420–4429.
[154] Philip Resnik. 1997. Selectional Preference and Sense Disambiguation. In *Tagging Text with Lexical Semantics: Why, What, and How?* ACL.
[155] Lisa Rhody. 2014. The Story of Stopwords: Topic Modeling an Ekphrastic Tradition. In *Proceedings of the Digital Humanities Conference*. ADHO, Lausanne.
[156] Lisa Marie Rhody. 2013. Topic Modeling and Figurative Language. *Journal of Digital Humanities* 2, 1 (April 2013).
[157] Lisa Marie Rhody. 2017. Beyond Darwinian Distance: Situating Distant Reading in a Feminist *Ut Pictura Poesis* Tradition. *PMLA* 132, 3 (May 2017), 659–667. https://doi.org/10.1632/pmla.2017.132.3.659
[158] Elaine Rich and Kevin Knight. 1991. *Artificial Intelligence*. McGraw Hill, New York.
[159] George Ritzer. 1993. *The McDonaldization of American Society: An Investigation into the Changing Character of Contemporary Social Life*. Pine Forge Press, Newbury Park,CA.
[160] Sarah T. Roberts. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press. https://doi.org/10.2307/j.ctvhrcz0v
[161] Nikolas Rose. 2010. *Inventing Our Selves: Psychology, Power, and Personhood*. Cambridge University Press, Cambridge.
[162] Marie-Laure Ryan. 2013. Possible Worlds. In *The Living Handbook of Narratology*, Peter Hühn, John Pier, Wolf Schmid, and Jörg Schönert (Eds.). Hamburg University, Hamburg, Germany.
[163] Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. 2010. Intermediated Technology Use in Developing Communities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Atlanta, GA, 2583–2592. https://doi.org/10.1145/1753326.1753718
[164] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. 2017. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. (Aug. 2017).
[165] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging Identity Through Gender, Race, and Class. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Denver, CO, 5412–5427. https://doi.org/10.1145/3025453.3025766
[166] Ari Schlesinger, Kenton P. O'Hara, and Alex S. Taylor. 2018. Let's Talk About Race: Identity, Chatbots, and AI. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Montréal, QC, 315:1–315:14. https://doi.org/10.1145/3173574.3173889
[167] Alexandra Schofield and David Mimno. 2016. Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics* 4 (July 2016), 287–300.
[168] Nick Seaver. 2022. *Computing Taste: Algorithms and the Makers of Music Recommendation*. University of Chicago Press, Chicago, IL.
[169] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator Engagement and Community Development in the Age of Algorithms. *New Media & Society* 21, 7 (July 2019), 1417–1443. https://doi.org/10.1177/1461444818821316
[170] Patrick Skeba and Eric P. S. Baumer. 2020. Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 101:1–101:22. https://doi.org/10.1145/3415172
[171] Susan Leigh Star and Geoffrey C. Bowker. 2002. How to Infrastructure. In *Handbook of New Media: Social Shaping and Consequences of ICTs*, L. Lievrouw and S. Livingstone (Eds.). SAGE, 55 City Road, 151–162. https://doi.org/10.4135/9781848608245
[172] Luke Stark. 2018. Algorithmic Psychometrics and the Scalable Subject. *Social Studies of Science* 48, 2 (April 2018), 204–231. https://doi.org/10.1177/0306312718772094
[173] Luke Stark. 2019. Facial Recognition Is the Plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (April 2019), 50–55. https://doi.org/10.1145/3313129
[174] Isabelle Stengers. 2013. Introductory Notes on an Ecology of Practices. *Cultural Studies Review* 11, 1 (Aug. 2013), 183–196. https://doi.org/10.5130/csr.v11i1.3459
[175] Isabelle Stengers. 2017. Thinking Life: The Problem Has Changed. In *Posthumous Life: Theorizing Beyond the Posthuman*. Columbia University Press, New York, 325–338. https://doi.org/10.7312/wein17214-015
[176] Alex S. Taylor. 2009. Machine Intelligence. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Boston, MA, 2109–2118.
[177] Alex S. Taylor. 2015. After Interaction. *interactions* 22, 5 (Aug. 2015), 48–53. https://doi.org/10.1145/2809888
[178] J.K. Trotter. 2016. Facebook Admits Pulitzer-Winning Photograph Is Not Child Pornography. https://gizmodo.com/facebook-admits-pulitzer-winning-photograph-is-not-chil-1786441732.
[179] Ted Underwood. 2012. What Kinds of "Topics" Does Topic Modeling Actually Produce?

[180] Ted Underwood. 2014. Theorizing Research Practices We Forgot to Theorize Twenty Years Ago. *Representations* 127, 1 (2014), 64–72.

[181] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 167:1–167:22. https://doi.org/10.1145/3415238

[182] Kristen Vaccaro, Ziang Xiao, Kevin Hamilton, and Karrie Karahalios. 2021. Contestability For Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 318:1–318:28. https://doi.org/10.1145/3476059

[183] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS, Vol. 30)*. Curran Associates, Inc.

[184] Ron Wakkary. 2020. A Posthuman Theory for Knowing Design. *International Journal of Design* 14, 3 (2020), 12.

[185] Ron Wakkary. 2021. *Things We Could Design: For More Than Human-Centered Worlds.* The MIT Press, Cambridge, MA.

[186] Ron Wakkary, William Odom, Sabrina Hauser, Garnet Hertz, and Henry Lin. 2015. Material Speculation: Actual Artifacts for Critical Inquiry. In *Decennial Aarhus Conference on Human Centered Computing*. https://doi.org/10.7146/aahcc.v1i1.21299

[187] Joseph B. Walther and Shawn Boyd. 2022. Attraction to Computer-Mediated Social Support. In *Communication Technology and Society: Audience Adoptino and Uses*, C. A. Lin and D. Atkins (Eds.). Hampton PRess, Cresskill, NJ, 153–188.

[188] Chi Wang, Rajat Raina, David Fong, Ding Zhou, Jiawei Han, and Greg Badros. 2011. Learning Relevance from Heterogeneous Social Network and Its Application in Online Targeting. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. ACM, Beijing, China, 655–664. https://doi.org/10.1145/2009916.2010000

[189] Max Weber. 1920 (2003). *The Protestant Ethic and the Spirit of Capitalism.* Dover, Mineola, NY.

[190] Martin Wiener, W. Alec Cram, and Alexander Benlian. 2023. Algorithmic Control and Gig Workers: A Legitimacy Perspective of Uber Drivers. *European Journal of Information Systems* 32, 3 (May 2023), 485–507. https://doi.org/10.1080/0960085X.2021.1977729

[191] Lorna Wing, Judith Gould, and Christopher Gillberg. 2011. Autism Spectrum Disorders in the DSM-V: Better or Worse than the DSM-IV? *Research in Developmental Disabilities* 32, 2 (March 2011), 768–773. https://doi.org/10.1016/j.ridd.2010.11.003

[192] Andrew Winzelberg. 1997. The Analysis of an Electronic Support Group for Individuals with Eating Disorders. *Computers in Human Behavior* 13, 3 (Aug. 1997), 393–407. https://doi.org/10.1016/S0747-5632(97)00012-2

[193] Ludwig Wittgenstein. 1922. *Tractatus Logico-Philosophicus.* Routledge & Kegan Paul, London.

[194] Ludwig Wittgenstein. 1953. *Philosophical Investigations.* Blackwell, Oxford.

[195] Marisol Wong-Villacres, Arkadeep Kumar, Aditya Vishwanath, Naveena Karusala, Betsy DiSalvo, and Neha Kumar. 2018. Designing for Intersections. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS) (DIS)*. ACM, Hong Kong, 45–58. https://doi.org/10.1145/3196709.3196794

[196] Alex J Wood, Mark Graham, Vili Lehdonvirta, and Isis Hjorth. 2019. Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society* 33, 1 (Feb. 2019), 56–75. https://doi.org/10.1177/0950017018785616

[197] Qian Yang, Justin Cranshaw, Saleema Amershi, Shamsi T. Iqbal, and Jaime Teevan. 2019. Sketching NLP: A Case Study of Exploring the Right Things To Design with Language Intelligence. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Glasgow, 185:1–185:12. https://doi.org/10.1145/3290605.3300415

[198] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS)*. ACM, Hong Kong, 585–596. https://doi.org/10.1145/3196709.3196730

[199] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. 2011. Review Spotlight: A User Interface for Summarizing User-Generated Reviews Using Adjective-Noun Word Pairs. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Vancouver, BC, 1541–1550.

Received 7 June 2023; revised 6 March 2024; accepted 7 March 2024