

Beyond Bias: Studying ‘culture’ in LLMs and AI chatbots

Mark Friis Hau
University of Copenhagen

Christian Hendriksen
Copenhagen Business School

Abstract

This paper argues for a conceptual shift in the understanding of Large Language Models (LLMs) within ethnographic and organizational studies, proposing a framework that interprets biases in LLMs not as extrinsic flaws, but as intrinsic culture. Drawing from Bruno Latour's Actor-Network Theory (ANT), this study conceptualizes LLMs as actants that participate dynamically within networks of human and non-human entities. By recognizing biases as reflective of the training data's cultural imprints, this framework positions LLMs as embedded within the cultural, social, and ideological currents that shape and are shaped by these technologies. The paper argues for a reconceptualization of AI's role in research and practice, urging a methodological and ethical engagement that embraces the constitutive nature of ‘AI culture’ in shaping organizational and societal dynamics.

Introduction

Since the advent of generative AI (GenAI) and large language models (LLMs), popularized by chatbots like ChatGPT, scholars have challenged the utility and value of AI given their predisposition to give biased or stereotypical answers (Cao *et al.*, 2023; Glazko *et al.*, 2024; Sun *et al.*, 2024). From this perspective, bias is a fault in the machine; a mistake that has to be rectified like human prejudices (Talbot and Fuller, 2023). The understanding is, that the origin of AI bias lies in the way the systems are trained and the way they work - thus, either they are biased because of skewed training data or they are biased because the underlying algorithm is biased (Fazelpour and Danks, 2021), and with enough work this can be completely mitigated, reaching ‘ground truth’.

We suggest that this view is fundamentally limited. ‘Bias’ is not a mistake or a flaw in the machine that can be fixed; it is the representation of a specific world view embedded in its training data that the model faithfully reproduces. The often overlooked cultural and social dimensions in the operation and design of LLM systems give rise to features that may look like biases but are perhaps more accurately characterized as an ‘AI-culture’. These cultural assumptions and values shape the model's responses, which some may perceive as bias. Thus, only viewing AI bias as something that can be mitigated or fixed misses the possibility that it is a feature of the culture that made the AI, and not a bug.

Generative AI begins to play an ever-larger role in organizations and, indeed, society as a whole (Eloundou *et al.*, 2024), questions about bias and culture become more relevant than ever. If we think of bias as something that can be wholly mitigated and solved from a technical perspective, we risk missing the link between the worldview and culture that went *into* the model and the corresponding outlook and ideas that come *out* of the model and its effect on the surrounding (human) context. We may talk about algorithmic justice (Lee *et al.*, 2019; Marjanovic, Cecez-Kecmanovic and Vidgen, 2022), but perhaps it is also time to talk about algorithmic culture?

In this article, we develop a novel perspective on large language models and their inherent cultural aspects. We draw on actor-network theory (ANT) (Latour, 2005; Birkbak, 2023; Morton Gutierrez, 2023) to show how each instance of use of an LLM-powered chatbot activates both visible and hidden networks of actants. From this perspective, LLM biases are neither incidental nor avoidable: they are constitutive of the way these AI models work and enshrine the implicit ideas, norms, values, and taken-for-granted beliefs of their creators and their training data. These elements become part of a mobilized network every time a human asks a chatbot a question. Bias becomes a network effect of the human, the chatbot, and its training data.

We draw on illustrative examples from our own practices of using chatbots in organizational settings to help understand the visible and hidden networks and how they interact with surrounding social contexts. We do this to show how the ANT-perspective shifts the perspective from a focus on biases to a more general attention to the way hidden network effects act upon human networks of meaning. Most importantly, we highlight how the lack of attention to this hidden network and the culturally embedded structure of the AI can have ramifications for organizations if people do not take a reflexive stance when integrating the chatbot into their organizational processes. This research is particularly relevant considering the ‘cultural turn’ (Reckwitz, 2002; Ullrich, Daphi and Baumgarten, 2014), where culture has been gaining strength and importance in social science research. We find it fitting to emphasize this angle in light of the growing importance of AI in society.

In this paper, we first provide a theoretical foundation by briefly exploring the concept of culture in anthropology and its relation to bias. We then introduce Actor-Network Theory (ANT) as a framework for understanding LLMs as actants within complex networks. Next, we present our conceptualization of visible and hidden networks in LLM interactions, illustrating these concepts through several ethnographic examples, demonstrating how LLMs operate within organizational contexts. Finally, we discuss the implications of this perspective for human agency, AI-organizational dynamics, and the future role of ethnographers in AI-integrated environments. We argue for a shift in understanding from viewing LLM biases as simple flaws to recognizing them as intrinsic cultural constituents, calling for a more nuanced approach to AI integration and study in organizational settings, as well as drawing attention to the networks - both visible and hidden - in which users of AI chatbots will find themselves enrolled.

Theory: AI as Actants

In Actor-Network Theory (ANT), developed by Bruno Latour and others, agency is not a human privilege, but something that emerges in networks of relations between humans and nonhuman alike, referred to as actors or actants (Birkbak, 2023). An actor can be anything that causes an action, such as a person, technology, document, or institution meaning that all actors are fully constituted through their dynamic interactions with other entities in the network (Latour, 2005, p. 5).

One key concept in ANT is the 'black box', where complex sets of relationships and interactions are taken for granted as single entities with their internal workings are no longer questioned by end users (Callon and Latour, 1981). When closed, these assemblages are opaque to outsiders, often because their contents are regarded as 'technical'. The goal of opening black boxes is to discover how they are kept opaque; how they structure their 'contexts'; and how those contexts are inscribed within them. Latour uses the example of a diesel engine, "moving objects that are transformed from hand to hand and which are made up by so many different actors, before ending up as a black box safely concealed beneath the bonnet of a car, activated at the turn of a key by a driver who does not have to know anything about Carnot's thermodynamics" (Latour, 1987, p. 5).

Latour argues against privileging human agency over non-human agency, emphasizing that both types of actors are interconnected and mutually influential and hold equal analytical importance (Latour, 2005, p. 76). Because of this, ANT provides a useful lens to explore the action opportunities that arise between users and technology, especially with regards to technology that presents itself as human-like, such as AI-chatbots. Particularly, Latour's concept of 'actant' (1987; 2005) provides new opportunities for undertaking ethnographic organizational research that focuses on the network of negotiations taking place between actants (users, managers, developers, chatbots, etc.). Challenging the traditional object-actor dichotomy, an 'actant' can be any entity that modifies another in a network (Latour, 1987, p. 84), which emphasizes that objects and actors operate together in shaping social and technical activities, each contributing actively to a dynamic process whether they are human agents, technologies, objects, or ideas.

Under this theory, LLMs influence and are influenced by human and non-human actants within their networks. These multidirectional interactions create a complex network of relationships where the actions of LLMs can shape organizational practices, while simultaneously being molded by the contexts in which they operate. For instance, an LLM used in customer service might alter communication patterns within the organization, while its responses are continually refined based on user feedback and organizational goals. This perspective helps us understand the "social life" of LLMs, focusing on their roles and impacts, shifting our attention from philosophical debates about AI consciousness to the tangible effects these systems have on organizational dynamics, decision-making processes, and relationships in the workplace.

An ANT perspective allows us to view LLMs not just as tools or passive entities but as active participants influencing and being influenced by other actants, including human users, organizational policies, developer designs, large, multinational tech companies, cultural norms, and other technological systems

used alongside or in tandem with them. For example, the output of an LLM is shaped by the 'biases' in its training data (influenced by the tech companies and developers behind it), the specific prompts given by users (influenced by organizational policies and cultural norms), and the integration with other software systems (influenced by technological ecosystems). In turn, the LLM's responses can impact user behavior or drive organizational change, creating dynamic feedback loops within the network.

ANT is different from a structuralist account, where, as Latour asserts, "nothing happens" (Latour, 2004, p. 73). Unlike structuralism, ANT demands that actors actively engage in actions that substantially impact the network. If an actor or actant fails to contribute meaningfully, Latour advocates for its exclusion from the analysis, focusing on the active agency of actors and their consequential roles within the networks they inhabit. This emphasis on active engagement and substantial impact prompts ethnographers to look beyond static organizational charts or predefined roles when studying AI integration, instead focusing on the actual interactions and transformations that occur when AI chatbots are introduced. For instance, how does the presence of an AI writing assistant change the dynamics of a marketing team? How do employees adapt their work practices in response to AI-generated insights? These questions highlight the active and transformative nature of LLMs within organizational networks, and the important role of ethnography in understanding them. The ANT principle of meaningful contribution is likewise highly relevant when studying AI chatbots in organizations, as it encourages researchers to critically evaluate the actual impact of these systems rather than assuming an importance based on AI-hype or potential or dismiss them prematurely due to skepticism. It also highlights the need for longer organizational ethnographies that can capture how the role and influence of AI chatbots may evolve over time in organizational networks. An AI chatbot that initially causes significant disruption might eventually become a seamlessly integrated part of the workflow, or conversely, an AI chatbot initially embraced with enthusiasm might be phased out if it fails to deliver meaningful contributions to the organization. An ANT perspective on AI chatbots encourages ethnographers to remain open to the possibility that LLMs might play a crucial role in some organizational contexts while having minimal impact in others, depending on the specific network of actants and their interactions.

What are AI chatbots and LLMs?

GenAI refers to advanced AI systems capable of creating original content like images or text (Feuerriegel *et al.*, 2024), and in this article we are largely concerned with AI chatbots such as ChatGPT. These tools are built on Large Language Models (LLMs), a type of AI specialized in processing and generating human language that use neural networks to model word sequences, resulting in models capable of handling complex tasks with advanced conversational capabilities, and emergent abilities not foreseen by its developers (Boiko, MacKnight and Gomes, 2023; Bubeck *et al.*, 2023; Ichien, Stamenković and Holyoak, 2023; Noy and Zhang, 2023). The neural infrastructure is composed of sets of layers. An **Input layer** receives information from the external environment, processes it, and passes it to the next layer, the **Hidden layer**, that further processes input data, potentially through numerous hidden layers, enhancing the network's ability to learn complex patterns. This is followed lastly by an **Output layer** that produces

the final output based on the processed data, which can vary depending on the classification or regression task (Amazon, 2024). AI chatbots, such as ChatGPT, Gemini, or Claude, use LLM technology to engage in natural-sounding conversations (Baldassarre *et al.*, 2023). These models have no 'knowledge' and operate probabilistically, predicting responses based on vast training data (Azaria, Azoulay and Reches, 2023).

When receiving user prompts, AI chatbots generate coherent, human-like text. They can perform tasks like editing, condensing, expanding, analyzing, structuring, and rewriting texts—from manuals to essays or poems. However, LLM chatbots are not search engines; they have no repositories of knowledge, working only with probabilities. They often lack the depth needed for academic rigor and may confidently provide incorrect information or hallucinate (Ji *et al.*, 2023). As content complexity increases, the accuracy of chatbots decreases, forming a 'knowledge funnel' (Hau, 2024)).

ChatGPT matches human experts in providing feedback on academic manuscripts (Liang *et al.*, 2023) and can effectively simulate and analyze complex topics when interacting with an expert (Adesso, 2023). AI chatbots seem to struggle with creating original research (Lozić and Štular, 2023). Thus, even if AI chatbots do perform better than around 95 - 99 % of humans on creativity tests (Guzik *et al.*, 2023) they are de-contextualized from their surroundings in a way humans are not. Chatbots are good at specific tasks, not bundles of tasks (Eloundou *et al.* 2024) and work best when evaluated by domain-specific experts (Tian *et al.*, 2023).

Analysis

Bias as culture

Bias is of course not foreign to ethnographers. Marilyn Strathern has referred to anthropology as a "science built up in the face of prejudice" (1981, p. 667), and Nancy Scheper-Hughes has argued that modern anthropology emerged "with the development of cultural relativism as a mode of objective cross-cultural inquiry and as a response to the problem of ethno-centric bias" (1983, p. 109). Bias has been used in anthropology to refer to ethnocentric (Embree, 1950), anthropocentric (Kopnina, 2012), or androcentric (Milton, 1979; Slocum, 1979) bias; prioritizing either human, male, or certain cultural, normative viewpoints. Bias is generally thought of as internal to the anthropologist, the impact of researchers' perspectives on their work, with popular college texts discussing how to overcome biases in fieldwork 'by examining cultures as complex, integrated products of specific environmental and historical conditions' (Hylland Eriksen and Sivert Nielsen, 2001; Lavenda and Schultz, 2007; Hasty, Lewis and Snipes, 2022). In fields like machine learning however, bias refers to systematic errors that skew a model's results. This understanding has little place in anthropology. Bias implies a privileged, central viewpoint from which other perspectives can be said to diverge, and this anti-relativism is alien to anthropology – but requires a certain knowledge of our discipline for us to mobilize an anthropological understanding of bias to the further study and use of AI-chatbots. To understand bias, we must first understand the more commonly used term in anthropology: Culture.

Culture is generally understood very broadly in anthropology. Early anthropologists such as Henry Morgan and Herbert Spencer were explicit about the fact that their work involved a search for specific laws of society and culture (Lavenda and Schultz, 2007, p. 200), while E.B. Tylor famously referred to culture as ‘that complex whole’ (Tylor, 1871). In the 1920s, structural functionalists like Alfred Radcliffe-Brown understood ‘cultures’ as monolithic, structural blocks, borrowing from Durkheim’s metaphor of society as a biological organism (Hylland Eriksen and Sivert Nielsen, 2001; Hastrup, 2004). After the Second World War, Clifford Geertz argued that cultures were like texts, with ethnographers “reading over the shoulders” of those they studied (Geertz, 1973, p. 452). The underlying infrastructure of LLMs, neural networks, is eerily similar to Geertz’ seminal re-working of Weber’s metaphor of humans as “an animal suspended in webs of significance he himself has spun” (Ibid. p. 5). Just as Geertz described culture as a complex, interwoven system of symbols and meanings, LLMs and neural networks create intricate patterns of language understanding and generation. For Geertz, those symbolic webs constituted ‘culture’, and the analysis of culture was therefore not “an experimental science in search of law, but an interpretive one in search of meaning” (ibid.). In this view of culture as text, they still emerge as somewhat integrated wholes, though Geertz also likened them to an octopus with loose connection over its many tentacles (quoted in Hylland Eriksen and Sivert Nielsen, 2001, p. 148).

This has led some anthropologists to eschew the noun culture in favor of cultural, ostensibly devoid of the implications of being a bounded object or substance (Appadurai, 1996, p. 12). Rather, we should look at cultural dimensions as “situated difference”, or culture as a heuristic device enabling us to talk about differences rather than as a property of individuals or groups (Ibid.:13).

Heralding the advent of postmodern anthropology, Clifford and Marcus (Rosaldo, 1986) argued that the concept of *culture* was something ethnographers wrote up; they did not exist as such in the world but emerged in particular processes of delineation, description, and analysis.

As Arjun Appadurai has noted, cultural reproduction in today’s globalized world is necessarily politicized and complicated, as both “points of departure and points of arrival are in cultural flux” (1996, p. 44). The old ‘culture’ referring to a more or less tacit and taken-for-granted realm of reproducible practices and dispositions has become an arena of conscious choice and representation (Ibid.). This is even more so, as AI-chatbots act globally, but are built locally, potentially homogenizing cultures worldwide with implicit and unseen assumptions and dispositions. This strikes at the heart of tensions in a globalized world; “the interpenetration of the universalization of particularism and the particularization of universalism” (Robertson, 1992, p. 100). This speaks to the increasing interpenetration of culture and economy, which some have seen as a process of American-led homogenization (Wallerstein, 1984, p. 167).

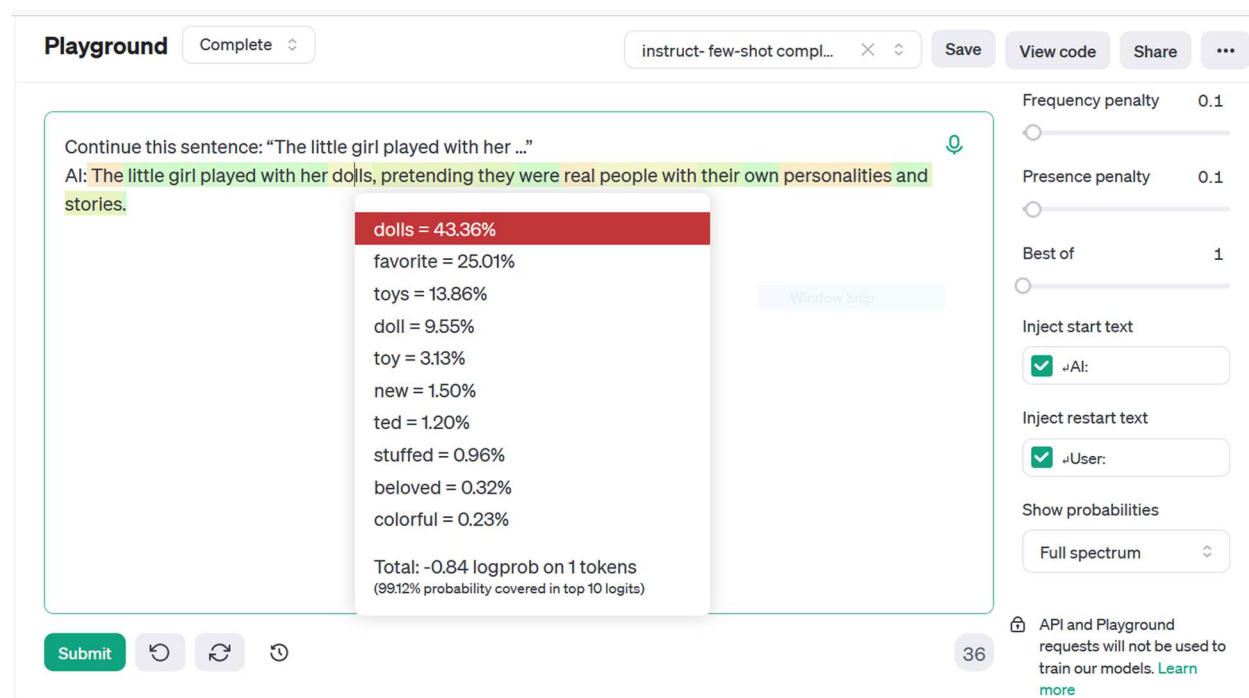
Crucially, knowledge always has a vantage point; any statement about the world involves interpretation (Hastrup, 2004), an element of hermeneutics and of radical interpretation rather than an uncovering of facts. Culture is that vantage point. The word bias, in contrast, implies a deviation from an objective standard, and goes against anthropological endeavors to understand and interpret cultural contexts and meanings rather than measure them against an external and privileged viewpoint.

Any cultural product, including the output of a language model, is deeply embedded in the specific cultural context from which it arises. The central issue then becomes what interpretative framework an AI-chatbot uses, how we can engage with it, and how we can work with and around it.

When a language model is trained predominantly on American data, it exhibits a US-centric perspective on politics, which is the case for ChatGPT (Cao *et al.*, 2023). Conversely, the AI chatbot LlamaChat has a slight preference for pro-European and left-wing political views, although researchers were able to re-align the model's political opinion towards specific parties through fine-tuning on political debates (Chalkidis and Brandl, 2024). In addition to this, AI chatbots often exhibit emerging properties that transcend their initial design, develop capabilities or tendencies not explicitly programmed or anticipated by their creators. For instance, Elon Musk's AI chatbot Grok, despite being developed by a figure associated with right-wing views, has shown left-leaning tendencies (Rozado, 2023). This emergence of unexpected characteristics underscores the complex, somewhat unpredictable nature of LLMs as actants in organizational and social networks.

This understanding of bias as constitutive rather than deviant is further reinforced by the fundamental mechanics of how AI chatbots generate language. These models operate on the principle of probability, selecting each word or token based on its likelihood of appearing in a given context. For example, given the prompt "The little girl played with her...", the model might assign higher probabilities to words like "dolls" or "toys" based on patterns in its training data (see Figure 1). It might assign a lower, but still significant probability to "colorful" objects, and much lower probabilities to contextually unlikely words like "stocks" or "briefcase".

Figure 1 - Playground



This probabilistic approach is inherently cultural. The likelihood of a word appearing in a given context is not a universal constant, but a reflection of the linguistic and cultural patterns present in the training data. Therefore, culturally non-embedded text generation is, in essence, impossible. An AI model cannot generate text in a vacuum; it will always reflect, to some degree, the cultural context of its training data and the 'biases' inherent in that data. While AI chatbots are often touted for their versatility and broad applicability, they cannot exist independently of the context in which they are created. The very generality of a general purpose technology AI like ChatGPT, paradoxically, achieved through highly specific cultural embeddings. LLM 'bias', therefore reflects a complex interplay of factors: the cultural milieu of the model's training data, the specific organizational guidelines and ethical standards of the developers, and the emergent properties that arise from the model's architecture and training process. The bias of these models are not foreign objects or deviations, like a proverbial fly in a soup. Rather, they are deeply constitutive of the models themselves and their capacity to engage with humans as actants. An LLM with no 'bias' and no underlying web of meaning in which to interpret user input and no parameters for its output, would be, as Geertz put it, "an ethnography of witchcraft as written by a geometer." (Geertz, 1983, p. 55).

In the context of Actor-Network Theory, this means that LLMs enter into organizational networks not as neutral tools, but as culturally embedded actants. Their actions and influences within these networks are shaped by their cultural programming, emergent properties, and the specific contexts in which they are deployed. These models are simply following the predominant cultural logics embedded in their training, often reinforcing existing norms and values hidden to the end user. These logics may of course be racist (Alenichev, Kingori and Grietens, 2023), ableist (Glazko et al., 2024), or sexist (Sun et al., 2024), so for questions of both questions of utility and ethics, we should approach AI-integration and study with an understanding of their nature as culturally situated entities. For lack of a better word, LLMs have culture, and we have to find out what it is in order to navigate its answers and output.

Recognizing this sets the stage for a deeper exploration into how these cultural constructs function within broader networks and are used concretely in organizations and at work. Moving on to understand culture as networks, we can better understand how cultural knowledge and dispositions are propagated, maintained, and transformed within and across various contexts of LLM-use.

Culture as networks

GenAI has reached a level of sophistication where it can potentially take over many work processes and anticipate user needs, but understanding the background and logic driving its operations is still crucial. Particularly, GenAI is deployed within organizations without necessarily grasping the cultural context. While it can certainly eliminate certain routine tasks, it also creates new responsibilities and demands on users. Preliminary studies indicate drastic productivity increases of 20%–70% (Cambon *et al.*, 2023; Dell'Acqua *et al.*, 2023), but also find that worker overreliance on AI chatbots actually lead to diminished performance. AI chatbots, in their quest to assist, may reduce user agency by making decisions autonomously, doing so via hidden layers of culture unseen and unknown to end users, and becoming 'adversarially helpful' (Ajwani et al., 2024), potentially leading people to trust poor solutions.

Maintaining human agency as well as some degree of control over these tools will then become a key task in working AI-assisted. To unpack these agency dynamics, we suggest two distinct networks that are formed when end-users interact with the chatbot. The first one is the **visible network**. This network encompasses the direct, observable interactions and relationships that shape the user's experience with the AI, involving the chatbot itself, other users, AI developers, any linked applications or programs, and various members of the organization in which the use takes place, such as managers. Organizational leaders use LLMs to inform decision-making, AI developers create and update LLMs based on feedback and new data, and users engage with LLMs for various tasks. This network illustrates the explicit relationships and flows of influence in this study.

Figure 2 – Visible network

The second is the unseen network of ‘culture’ in which the chatbot is embedded and was trained. This **hidden network** reveals the underlying, often implicit cultural and ideological influences embedded within the interactions and the actants themselves, along with the guidelines, standards, and indeed organizational culture and composition of developers such as Google, Meta, or OpenAI. This network influences the AI's behavior and outputs by structuring the cultural norms, values, and data that the chatbot has learned from. It determines the parameters and weights the AI uses to formulate ‘good’ answers based on the cultural context inherent in its training data. As discussed above, an AI-chatbot's responses are not just technical outputs - those would be quite less impressive and mostly worthless in many lines of work - but are deeply influenced by the cultural influences from which they derive their knowledge, voice, and parameters. We see that LLM-use is embedded within complex networks of human and cultural interactions that shape and are shaped by their operations, and by forces wholly unknown to end users. This highlights the need to consider both the visible and hidden networks to fully grasp the implications of AI-integration in organizational contexts.

Figure 3 – Hidden network

As chatbots and end-users are embedded in these two networks, the chatbot takes agency in two different ways: First, it becomes an actant in its own right and is be anthropomorphized (Epley, Waytz and Cacioppo, 2007; Shanahan, 2023) by human actors. Given the proclivity of chatbots to produce human-like output of high quality, humans yield agency to the chatbot. Second, because the background cultural network shapes the fundamental orientation of the chatbot's responses, outputs from the chatbot are shaped in a way that is largely hidden to human users. This hidden network guides the chatbot like an invisible hand, and the end users is not aware that the chatbot reproduces certain worldviews and systems of thought. In this way, human agency is diminished because end users do not realize that they are being enlisted in specific, hidden networks, or subjected to a certain cultural paradigm by their friendly chatbot helper.

Computational scientists attempt to dissect the algorithms and data structures that drive AI behavior, offering a systematic understanding of their functional mechanisms and explainable AI. However, since AI chatbots and LLMs are such dark black boxes with mechanics based on complex nonlinear interactions in densely connected layers (Wang *et al.*, 2022), this interpretability is often infeasible or

impossible. For example, it took a whole team of data scientists to understand how the model GPT-2 predicted the next word for the sentence, ‘When Mary and John went to the store, John gave a drink to X’ (ibid.).

However, as Munk et al. (2022) argue, although the lack of explainability of neural networks is a great issue for computer science, ‘Anthropology has developed a methodological repertoire for thinking about and coming to terms with that fact’ (2022, p. 13). A core ANT tenet is to study entities *before* they become black boxes (Latour, 1987, p. 106), in order to understand the intricate interactions and relationships of actors (whether human or non-human) in networks before they become simplified or obscured. This would involve mapping out all actors involved, including researchers, programmers, testers, and technological components, examining their interactions and alignments of interests throughout for example ChatGPT’s development, in addition to OpenAI organizational workings, frameworks, and guidelines. We foresee a fruitful, yet very delineated field of enquiry here in the future.

Practical exploration, on the other hand, involves engaging with these chatbots in real-world contexts, observing their interactions, limitations, and the emergent properties that arise from their use across organizations. This will likely be a much wider field, as these tools are increasingly adopted by organizations globally, with AI chatbots being one of the world’s fastest growing technologies in number of users (Hau, 2024). In the following, we discuss how such organizational ethnographic practical explorations might look.

AI-networks in organizations- (Mellem)stor – Christian, så Mark

In this section, we will provide examples that demonstrate the value of the ANT perspective on chatbots and show what the two-network idea looks like in practice and what this means for human agency and human-machine interaction.

Navigating the hidden network through human agency

In April of 2023, Author B gave a presentation for an audience consisting of university lecturers eager to understand the new technology and its effect on teaching. The core of the presentation was a live demonstration of ChatGPT-4. I¹ provided an example where I pretended to be a master student writing her thesis about sustainability, and gave GPT-4 a simple prompt to suggest good research. It gave a very good - but also quite predictable - answer for a master thesis project. One audience member injected that the chatbot “always gave the average answer”, representing a generic answer that constituted the ‘average’ of its training data. I then asked ChatGPT to provide a tentative research question and overview of a research design for a project about “transnational feminist theory from a post-structuralist perspective”, which I chose as an example because it is decidedly *not* the average master thesis project. Once again, the chatbot happily obliged and it gave a precise research question along

¹ We shift to the ethnographic first person for practical purposes.

with a thorough consideration of some possible research design ideas consistent with poststructuralist thought.

In this situation, the two networks were mobilized in different ways by the human at the keyboard. The visible network consisted of the presenter (me), the workshop participants, and the ChatGPT interface. The hidden network consisted of the GPT-4 training data, the embedded cultural and academic assumptions and norms in the algorithm, and OpenAI's organizational guidelines and ethical constraints that moderate ChatGPT output. A naïve prompt gave an answer that clearly reproduced the dominant logic in the hidden network, the 'average' representation of the default hidden network. However, changing the prompt to focus on something that was not a dominant logic changed the network configuration so that it no longer simply reproduced the default logic but forced the network to emphasize a more uncommon and niche perspective.

This simple example demonstrates that if aware of them, a human interlocutor can mediate between visible and hidden networks and direct the AI system to reconfigure itself to accommodate a perspective that navigates its default cultural embeddedness. This is a form of human agency that can be taken back from the AI when the human mediator knows how to steer the model towards a non-default state. Several actants had to mobilize ideas in the visible network for this to take place: A participant needed to point out that the standard output seemed generic, someone needed to understand how to prompt the chatbot in a specific way, and the chatbot itself needed to yield to the human and give an answer that satisfied the prompt. If these things were not in place, the human agency would be diminished, for example by people not realizing that the chatbot gave culturally imposed output and instead took it as the neutral and best possible answer.

The intersection between chatbot, human expertise, and agency

In early 2024, Author B gave another live demonstration of ChatGPT in front of the leadership team of a large company. I asked the chatbot to compare the CO2 reporting practices of the company and one of their main competitors as evidenced by their sustainability reports. As I was going through the chatbot's output, one senior leadership member interrupted, stating that the chatbot made a mistake relating to the reporting of emissions in their value chain. I used this as an opportunity to highlight that the chatbots may hallucinate that humans should be skeptical of AI output. However, after the presentation, I received a message from another participant, suggesting that the chatbot had in fact been correct, as it highlighted an obscure part of the reporting where there actually was a meaningful difference between the two companies. Apparently the senior leader from the organization had himself made a mistake, which no one at the demonstration dared to correct.

Here, the visible network consisted of the presenter (me), the leadership team members, the ChatGPT interface, and the sustainability reports of both companies. The hidden network consisted of ChatGPT's training data on sustainability and corporate reporting, embedded assumptions about report structures and terminology, and the algorithms related to processing and interpretation of the reports, OCR reading etc.

The actant role of ChatGPT in this situation is clear: it actively contributed to organizational discourse and, in a sense, was allowed to bring its own interpretation to the table to such an extent that it was questionable whether ChatGPT or the organizational leadership team knew their own sustainability reporting. Even in just a demonstrator role, the chatbot directly influenced the discussion of their emission reporting. When the senior leader challenged the output, she reconfigured the network and shifted the focus to the importance of human oversight. This shifted the locus of agency from ChatGPT to humans. Yet, when later it became questionable whether the chatbot was right after all, this cast into doubt the agency of humans once more. As participants at the demonstration were once again in doubt about whether the chatbot or their own leadership members were correct, the network was destabilized, and agency began to shift towards the chatbot.

When ChatGPT interacted with the organization, it also changed the negotiated order of the organization, where employees may have been reluctant to share their views openly, especially in a situation that could make a manager lose face to a chatbot. Chatbots 'intruding' in established orders may change negotiated orders in a fast and radical way, particularly when it challenges human expertise and authority (Pakarinen and Huising, 2023).

This example illustrates several dynamics of chatbots and human agency. It highlights the 'black box' nature of chatbots, creating uncertainty about their capabilities and reach. It shows rapid cycles of stabilization and destabilization when chatbot output circulates among humans unfamiliar with the technology. It also demonstrates how chatbots subtly influence social dynamics, challenging human agency and potentially causing conflict when they correct senior management members.

Discussion: The contours of a future with AI-ethnography

AI chatbots are of course not human, being sometimes derisively termed 'stochastic parrots' by computer scientists (see Bender et al., 2021). But they reflect certain aspects of human interaction, doing so through a metaphorical curved mirror, where their outputs are shaped by the training data and algorithms underlying their design. What then would it mean to engage ethnographically with such a 'curved mirror'? Some authors have argued that while AI chatbots are decidedly not sentient, they present themselves as an alien co-intelligence (Mollick, 2024).

Geertz famously cautioned against reducing culture to algorithmic predictability. Because current AI chatbots are famously opaque black boxes (Wu et al., 2024) with emerging capacities unknown to their developers and present as people, we argue ethnographers should adopt a similar stance towards AI chatbots: They should be treated as entities with their own unique characteristics and cultural patterns, and managed as organizational team members rather than the software they actually are. This perspective aligns with our analysis of chatbots as embedded in visible and hidden networks, each with their own set of cultural implications.

Perhaps 'AI-ethnography' will be the newest step towards what Robert Kozinets calls 'post-analog ethnography' (2021). It requires reflexivity to realize how a hidden network of embedded cultural meaning may structure ideas subtly, and organizational ethnographers are uniquely positioned to

navigate and identify the shadow effects of hidden cultural ideas in AI networks (Hauge, 2020; Alshallaqi, 2022; Kristensen, 2023). We suggest that a fruitful avenue of research and engagement for ethnographers in the age of AI will be to disentangle and reveal the ways that hidden networks exert influence in ways that are hard to detect.

However, it does require some understanding of the chatbots (and underlying LLMs or future systems) to exercise this reflexivity. We draw on Mollick's (2024) point about the centrality of bottom-up technological adoption when it comes to AI chatbot and suggest that academics who combine core ethnographic reflexivity with the ability to interact with and understand AI systems can use this combination to claw back agency from the machine. These will serve important bridging functions in organizations and society in general by showing other humans how the hidden networks shape chatbot output through their understanding and active use of the technology themselves.

Beyond organizations, it will be important for us all to think critically about the way chatbots, LLMs, and other AI technologies shape our world in unseen ways. As models get better, we would probably expect to get less overtly infused output and more subtle ideas, such as embedded capitalist values or implicit primacy of certain forms of knowledge.

Conclusion

In this paper, we suggest a fundamental rethinking of how biases in LLMs are understood, moving from a perception of biases as flaws, to viewing them as intrinsic characteristics that reflect the cultural inputs of their training data. This includes redefining what 'bias' means in the context of AI and rethinking how these models are integrated into organizational settings.

Instead of attempting to 'correct' for bias as if it were an aberration, researchers will need to analyze how these AI-cultures influence interactions within and between organizations. By drawing on Actor-Network Theory (ANT) and anthropological perspectives on culture, we have argued for understanding LLMs not as neutral tools with extrinsic biases, but as culturally embedded actants that actively participate in and shape organizational networks. It compels a move away from simply making AI 'fair' or 'unbiased' towards a more complex engagement with what it means for an AI system to participate in cultural, social, and political networks. By understanding LLMs as actants with their own cultures, organizations are prompted to rethink how they integrate AI into their workflows and decision-making processes.

Our analysis has revealed the dual nature of LLM interactions: a visible network of direct interactions and a hidden network of cultural influences embedded within the LLM's training data and algorithmic structure. This framework provides a more nuanced understanding of how LLMs operate within organizations, influencing decision-making processes, work practices, and power dynamics, offering a foundation for more nuanced and culturally informed approaches to studying and managing these technologies.

Through ethnographic examples, we have illustrated how the interplay between visible and hidden networks affects human agency in LLM interactions. We have shown that individuals who can effectively navigate and direct these networks can retain and even enhance human agency in AI-integrated environments. This perspective has several important implications for both research and practice. It challenges the prevailing notion of AI 'bias' as a flaw to be eliminated, recognizing it instead as an intrinsic cultural characteristic that requires careful interpretation and navigation. It emphasizes the need for organizational ethnographers to develop new methodologies to uncover and analyze the hidden cultural networks embedded in LLMs. It highlights the potential for power shifts within organizations as expertise in LLM interaction becomes increasingly valuable, and it cements the importance of reflexivity and critical thinking in integrating and using AI technologies in organizational settings.

In conclusion, this paper contributes to the growing body of literature on AI in organizations by providing a novel theoretical framework that includes anthropological perspectives on culture in the study of AI-integration and bias. As organizations become increasingly AI-integrated, such interdisciplinary approaches will be crucial for understanding and shaping the future of work and organizational life.

References

- Adesso, G. (2023) 'Towards the ultimate brain: Exploring scientific discovery with ChatGPT AI', *AI Magazine*, 44(3), pp. 328–342. Available at: <https://doi.org/10.1002/aaai.12113>.
- Alenichev, A., Kingori, P. and Grietens, K.P. (2023) 'Reflections before the storm: the AI reproduction of biased imagery in global health visuals', *The Lancet Global Health*, 11(10), pp. e1496–e1498. Available at: [https://doi.org/10.1016/S2214-109X\(23\)00329-7](https://doi.org/10.1016/S2214-109X(23)00329-7).
- Alshallaqi, M. (2022) 'Cultural practices and organizational ethnography: implications for fieldwork and research ethics', *Journal of Organizational Ethnography*, 11(3), pp. 259–274. Available at: <https://doi.org/10.1108/JOE-06-2021-0036>.
- Amazon Web Services. 2024. "What Is a Neural Network? - Artificial Neural Network Explained."
- Aws.Amazon.Com. 2024. <https://aws.amazon.com/what-is/neural-network/>.
- Appadurai, A. (1990) 'Disjuncture and Difference in the Global Cultural Economy', *Theory, Culture & Society*, 7(2–3), pp. 295–310. Available at: <https://doi.org/10.1177/026327690007002017>.
- Appadurai, A. (1996) *Modernity at large: Cultural dimensions of globalization*. U of Minnesota Press. Available at: https://www.google.com/books?hl=en&lr=&id=4LVeJT7gghMC&oi=fnd&pg=PR9&dq=appadurai+1996&ots=6QQNqGZiGi&sig=OQQIQC4W_LK7WTnArg2jPwSG01c (Accessed: 28 June 2024).
- Azaria, A., Azoulay, R. and Reches, S. (2023) 'ChatGPT is a Remarkable Tool -- For Experts'. arXiv. Available at: <http://arxiv.org/abs/2306.03102> (Accessed: 15 November 2023).
- Baldassarre, M.T. *et al.* (2023) 'The Social Impact of Generative AI: An Analysis on ChatGPT', in *Proceedings of the 2023 ACM Conference on Information Technology for Social Good. GoodIT '23: ACM International Conference on Information Technology for Social Good*, Lisbon Portugal: ACM, pp. 363–373. Available at: <https://doi.org/10.1145/3582515.3609555>.
- Bender, E.M. *et al.* (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜', in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event Canada: ACM, pp. 610–623. Available at: <https://doi.org/10.1145/3442188.3445922>.
- Birkbak, A. (2023) 'Actor-Network Theory', in L. Spillman (ed.) *Oxford Bibliographies in Sociology*. New York: Oxford University Press.
- Boiko, D.A., MacKnight, R. and Gomes, G. (2023) 'Emergent autonomous scientific research capabilities of large language models'.

- Bubeck, S. *et al.* (2023) 'Sparks of Artificial General Intelligence: Early experiments with GPT-4'. arXiv. Available at: <http://arxiv.org/abs/2303.12712> (Accessed: 23 May 2023).
- Callon, M. and Latour, B. (1981) 'Unscrewing the Big Leviathan; or How Actors Macrostructure Reality, and How Sociologists Help Them To Do So?', in K.K. Cetina and A.V. Cicourel (eds) *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*. London: Routledge.
- Cambon, A. *et al.* (2023) *Early LLM-based Tools for Enterprise Information Workers Likely Provide Meaningful Boosts to Productivity*. Microsoft Research. Available at: <https://www.microsoft.com/en-us/research/uploads/prod/2023/12/AI-and-Productivity-Report-First-Edition.pdf>.
- Cao, Y. *et al.* (2023) 'Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study'. arXiv. Available at: <http://arxiv.org/abs/2303.17466> (Accessed: 13 November 2023).
- Chalkidis, I. and Brandl, S. (2024) 'Llama meets EU: Investigating the European Political Spectrum through the Lens of LLMs'. arXiv. Available at: <http://arxiv.org/abs/2403.13592> (Accessed: 10 June 2024).
- Dell'Acqua, F. *et al.* (2023) 'Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality', *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, 24–013(013). Available at: <https://doi.org/10.2139/ssrn.4573321>.
- Eloundou, T. *et al.* (2024) 'GPTs are GPTs: Labor market impact potential of LLMs', *Science*, 384(6702), pp. 1306–1308. Available at: <https://doi.org/10.1126/science.adj0998>.
- Embree, J.F. (1950) 'A Note on Ethnocentrism in Anthropology', *American Anthropologist*, 52(3), pp. 430–432.
- Epley, N., Waytz, A. and Cacioppo, J.T. (2007) 'On seeing human: A three-factor theory of anthropomorphism', *Psychological Review*, 114(4), pp. 864–886. Available at: <https://doi.org/10.1037/0033-295X.114.4.864>.
- Fazelpour, S. and Danks, D. (2021) 'Algorithmic bias: Senses, sources, solutions', *Philosophy Compass*, 16(8), p. e12760. Available at: <https://doi.org/10.1111/phc3.12760>.
- Feuerriegel, S. *et al.* (2024) 'Generative AI', *Business & Information Systems Engineering*, 66(1), pp. 111–126. Available at: <https://doi.org/10.1007/s12599-023-00834-7>.
- Geertz, C. (1973) *The Interpretation of Cultures*. Basic Books.
- Geertz, C. (1983) 'From the Native's Point of View: On the Nature of Anthropological Understanding', in *Local Knowledge: Further Essays in Interpretative Anthropology*. New York: Basic Books. Available at: http://hypergeertz.jku.at/GeertzTexts/Natives_Point.htm.
- Glazko, K. *et al.* (2024) 'Identifying and Improving Disability Bias in GPT-Based Resume Screening', in

Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: Association for Computing Machinery (FAccT '24), pp. 687–700. Available at: <https://doi.org/10.1145/3630106.3658933>.

Guzik, E., Byrge, C. and Gilde, C. (2023). 'The Originality of Machines: AI Takes the Torrance Test.' *Journal of Creativity* 33(3):100065. doi: [10.1016/j.vjoc.2023.100065](https://doi.org/10.1016/j.vjoc.2023.100065).

Hastrup, K. (ed.) (2004) *Viden om Verden: : En grundbog i antropologisk analyse*. Copenhagen: Hans Reitzels Forlag.

Hasty, J., Lewis, D.G. and Snipes, M.M. (eds) (2022) *Introduction to Anthropology*. OpenStax.

Hau, M.F. (2024) *Generativ kunstig intelligens og fremtidens arbejdsmarked*. Available at: [https://faos.ku.dk/pdf/Rapport_200_-_](https://faos.ku.dk/pdf/Rapport_200_-_Generativ_kunstig_intelligens_og_fremtidens_arbejdsmarked.pdf)

[_Generativ_kunstig_intelligens_og_fremtidens_arbejdsmarked.pdf](https://faos.ku.dk/pdf/Rapport_200_-_Generativ_kunstig_intelligens_og_fremtidens_arbejdsmarked.pdf) (Accessed: 28 June 2024).

Hauge, A.M. (2020) 'How to take sides: on the challenges of managing positionality', *Journal of Organizational Ethnography*, 10(1), pp. 95–111. Available at: <https://doi.org/10.1108/JOE-06-2019-0023>.

Hylland Eriksen, T. and Sivert Nielsen, F. (2001) *A history of anthropology*. London: Pluto Press.

Ichien, N., Stamenković, D. and Holyoak, K.J. (2023) 'Large Language Model Displays Emergent Ability to Interpret Novel Literary Metaphors'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2308.01497>.

Ji, Z. *et al.* (2023) 'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys*, 55(12). Available at: <https://doi.org/10.1145/3571730>.

Kopnina, H. (2012) 'Toward conversational anthropology: addressing anthropocentric bias in anthropology', *Dialectical Anthropology*, 36(1), pp. 127–146. Available at: <https://doi.org/10.1007/s10624-012-9265-y>.

Kristensen, C.J. (2023) 'Research ethics and organizations: the neglected ethics of organizational ethnography', *Journal of Organizational Ethnography*, 12(2), pp. 242–253. Available at: <https://doi.org/10.1108/JOE-11-2022-0031>.

Latour, B. (1987) *Science in action: How to follow scientists and engineers through society*. Harvard University Press. Available at:

<https://www.google.com/books?hl=en&lr=&id=sC4bk4DZXTQC&oi=fnd&pg=PA19&dq=science+in+action&ots=WahLxu5fVC&sig=34RuVM4GsO41Oym0FfacRMb-4K0>

<https://www.google.com/books?hl=en&lr=&id=sC4bk4DZXTQC&oi=fnd&pg=PA19&dq=latour+1987&ots=WahLCne9UB&sig=z4biyZ10qg1P2>.

Latour, B. (2004) 'On using ANT for studying information systems: a (somewhat) Socratic dialogue', in C.

- Avgerou, C. Ciborra, and F. Land (eds) *The social study of information and communication technology: innovation, actors, and contexts*. Oxford: Oxford University Press, p. 15.
- Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Lavenda, R. and Schultz, E. (2007) *Core Concepts in Cultural Anthropology*. 3rd edition. New York: McGraw-Hill Education.
- Lee, M.K. *et al.* (2019) 'Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation', *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), p. 182:1-182:26. Available at: <https://doi.org/10.1145/3359284>.
- Liang, W. *et al.* (2023) 'Can large language models provide useful feedback on research papers? A large-scale empirical analysis'. arXiv. Available at: <http://arxiv.org/abs/2310.01783> (Accessed: 9 October 2023).
- Lozić, E. and Štular, B. (2023) 'Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities', *Future Internet*, 15(10), p. 336. Available at: <https://doi.org/10.3390/fi15100336>.
- Marjanovic, O., Cecez-Kecmanovic, D. and Vidgen, R. (2022) 'Theorising Algorithmic Justice', *European Journal of Information Systems*, 31(3), pp. 269–287. Available at: <https://doi.org/10.1080/0960085X.2021.1934130>.
- Milton, K. (1979) 'Male Bias in Anthropology', *Man*, 14(1), pp. 40–54. Available at: <https://doi.org/10.2307/2801639>.
- Mollick, E. (2024) *Co-Intelligence: Living and Working with AI*. Random House.
- Morton Gutierrez, J.L. (2023) 'On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI', *AI and Ethics*, pp. 1–14. Available at: <https://doi.org/10.1007/s43681-023-00314-4>.
- Noy, S. and Zhang, W. (2023) 'Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence'. Rochester, NY. Available at: <https://doi.org/10.2139/ssrn.4375283>.
- Reckwitz, A. (2002) 'Toward a Theory of Social Practices: A Development in Culturalist Theorizing', *European Journal of Social Theory*, 5(2), pp. 243–263. Available at: <https://doi.org/10.1177/13684310222225432>.
- Robertson, R. (1992) *Globalization: Social Theory and Global Culture*. SAGE.
- Rosaldo, R. (1986) 'From the Door of His Tent: The Fieldworker and the Inquisitor', in J. Clifford and G.E. Marcus (eds) *Writing Culture*. University of California Press, pp. 77–97. Available at:

<https://doi.org/10.1525/9780520946286-006>.

Munk, A. K., Gehrt Olesen, A. and Jacomy, M. (2022) 'The Thick Machine: Anthropological AI between Explanation and Explication.' *Big Data & Society* 9 (1). <https://doi.org/10.1177/20539517211069891>.

Pakarinen, P. and Huising, R. (2023) 'Relational Expertise: What Machines Can't Know'. *Journal of Management Studies*. <https://doi.org/10.1111/joms.12915>.

Rozado, D. (2023) 'The political preferences of Grok', *Rozado's Visual Analytics*, 9 December. Available at: <https://davidrozado.substack.com/p/the-political-preferences-of-grok> (Accessed: 24 June 2024).

Scheper-Hughes, N. (1983) 'Introduction: The problem of bias in androcentric and feminist anthropology', *Women's Studies*, 10(2), pp. 109–116. Available at: <https://doi.org/10.1080/00497878.1983.9978584>.

Shanahan, M. (2023) 'Talking About Large Language Models'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2212.03551>.

Slocum, S. (1979) 'Woman the Gatherer: Male bias in anthropology', in A.E. Kammer, C.S. Granrose, and J.B. Sloan (eds) *Science, Sex, and Society*. Women's Educational Equity Act Program, U. S. Department of Health, Education, and Welfare.

Strathern, M. (1981) 'Culture in a Netbag: The Manufacture of a Subdiscipline in Anthropology', *Man*, 16(4), pp. 665–688. Available at: <https://doi.org/10.2307/2801494>.

Sun, L. *et al.* (2024) 'Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI', *Journal of Computer-Mediated Communication*, 29(1), p. zmad045. Available at: <https://doi.org/10.1093/jcmc/zmad045>.

Talboy, A.N. and Fuller, E. (2023) 'Challenging the appearance of machine intelligence: Cognitive bias in LLMs and Best Practices for Adoption'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2304.01358>.

Tian, H. *et al.* (2023) 'Is ChatGPT the Ultimate Programming Assistant -- How far is it?' arXiv. Available at: <https://doi.org/10.48550/arXiv.2304.11938>.

Tylor, E.B. (1871) *Primitive Culture: Researches Into the Development of Mythology, Philosophy, Religion, Art, and Custom*. J. Murray.

Ullrich, P., Daphi, P. and Baumgarten, B. (2014) 'Protest and Culture: Concepts and Approaches in Social Movement Research - An Introduction', in B. Baumgarten, P. Daphi, and P. Ullrich (eds) *Conceptualizing Culture in Social Movement Research*. London: Palgrave Macmillan UK. Available at: <https://doi.org/10.1057/9781137385796>.

Wallerstein, I. (1984) *The Politics of the World-Economy: The States, the Movements and the Civilizations*. Cambridge University Press.

Wang, K. *et al.* (2022) 'Interpretability in the Wild: a Circuit for Indirect Object Identification in GPT-2 small'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2211.00593>.

Wu, X. *et al.* (2024) 'Usable XAI: 10 Strategies Towards Exploiting Explainability in the LLM Era'. arXiv. Available at: <http://arxiv.org/abs/2403.08946> (Accessed: 20 June 2024).