Atlantis: Critical Studies in Gender, Culture & Social Justice
Issue 46.1 (2025)
Special Issue: Ruptures, Resistance, Reclamation: Global Feminisms in a Digital Age
Research

Pressing Matters: How AI Irons Out Epistemic Friction and Smooths Over Diversity

by Nicole Ramsoomair

Abstract: This paper explores how Large Language Models (LLMs) foster the homogenization of both style and content and how this contributes to the epistemic marginalization of underrepresented groups. Utilizing standpoint theory, the paper examines how biased datasets in LLMs perpetuate testimonial and hermeneutical injustices and restrict diverse perspectives. The core argument is that LLMs diminish what Jose Medina calls "epistemic friction," which is essential for challenging prevailing worldviews and identifying gaps within standard perspectives, as further articulated by Miranda Fricker (Medina 2013, 25). This reduction fosters echo chambers, diminishes critical engagement, and enhances communicative complacency. AI smooths over communicative disagreements, thereby reducing opportunities for clarification and knowledge generation. The paper emphasizes the need for enhanced critical literacy and human mediation in AI communication to preserve diverse voices. By advocating for critical engagement with AI outputs, this analysis aims to address potential biases and injustices and ensures a more inclusive technological landscape. It underscores the importance of maintaining distinct voices amid rapid technological advancements and calls for greater efforts to preserve the epistemic richness that diverse perspectives bring to society.

Keywords: algorithmic bias; artificial intelligence; echo chambers; epistemic friction; epistemic injustice; standpoint theory

Résumé : Cet article examine la façon dont les grands modèles de langage (GML) favorisent l'homogénéisation du style et du contenu et dont ils contribuent à la marginalisation épistémique des groupes sous-représentés. En s'appuyant sur la théorie du point de vue, l'article explique comment les ensembles de données biaisés des GML perpétuent les injustices testimoniales et herméneutiques et limitent les différents points de vue. L'argument principal est que les GML atténuent ce que Jose Medina appelle la « friction épistémique », qui est essentielle pour remettre en question les visions du monde qui sont prédominantes et déceler les lacunes dans les points de vue courants, comme l'explique Miranda Fricker (Medina 2013, 25). Cette réduction favorise les chambres d'écho, diminue l'engagement critique et renforce la complaisance dans la communication. L'IA concilie les désaccords de communication, réduisant ainsi les possibilités de clarification et de création du savoir. L'article souligne la nécessité d'améliorer la littératie critique et la médiation humaine dans la communication par l'IA afin de préserver la diversité des voix. En préconisant un engagement critique à l'égard des résultats de l'IA, cette analyse vise à lutter contre les préjugés et les injustices potentiels et à garantir un environnement technologique plus inclusif. Elle souligne l'importance de maintenir des voix distinctes dans un contexte où la technologie évolue rapidement et appelle à redoubler d'efforts pour préserver la richesse épistémique que les différents points de vue apportent à la société.

Mots clés : biais algorithmique; intelligence artificielle; chambres d'écho; friction épistémique; injustice épistémique; théorie du point de vue

Author: Nicole Ramsoomair is an Assistant Professor of Philosophy at Dalhousie University, a position she has held since 2021. She specializes in social and political philosophy, feminist philosophy, and applied ethics. She earned her PhD in Philosophy from McGill University in 2019, with a dissertation exploring the conditions of responsibility in cases of radical personality change. Her current research focuses on social responsibility, freedom of speech, and children's rights.

1. Introduction

Often described as “auto-complete on steroids,” Large Language Models (LLMs) are sophisticated computational algorithms designed to analyze, interpret, and synthesize text with human-like comprehension (qtd. in Meuse 2023). LLMs can recognize and replicate syntactic structures and connections in sentences and paragraphs which allows them to efficiently generate organized content, thereby reducing the need for extensive manual edits and facilitating accessible content creation. However, this proficiency may inadvertently dilute the distinctiveness of a user's “voice” and flatten the intricacies of communication. This article argues that the trend toward content homogenization, accelerated by the capabilities of LLMs, exacerbates the epistemic marginalization of underrepresented social identity groups. The limited datasets used for training these systems may normalize expression in ways that undermine individuality and hinder knowledge generation.

This article contributes to the evolving discourse on AI by demonstrating the relevance of standpoint theory in epistemology and by drawing on works such as those by Miranda Fricker (2007), Charles Mills (1997), and José Medina (2013). Fricker's analysis of epistemic injustice highlights the communal nature of knowledge creation, enriched by diverse experiences. Conceptual negotiation, involving the exchange and reconciliation of perspectives through dialogue, fosters broader viewpoints and critical reflection. However, the smoothing of communicative edges reduces what Medina identifies as “epistemic friction,” understood as the productive tension arising from the interaction of differing, often conflicting, epistemic viewpoints (Medina 2013, 75). However, with the increasing insularity and exclusivity of knowledge generation, AI replicates biases that favour dominant perspectives. AI-mediated communication exacerbates real-world asymmetries, thereby amplifying the underrepresentation and misrepresentation of marginalized experiences and narrowing the diversity of perspectives represented through these platforms.

Section 2 begins by examining how knowledge generation can reflect power dynamics that marginalize groups through epistemic injustices and biased AI systems, which in turn perpetuate systemic inequalities and restrict access to and participation in collective epistemic resources. Sections three, four, and five of this paper illustrate these effects by showing how AI encourages complacency in communication. The ease of communication offered by LLMs may reduce disagreements and necessary communicative impasses, enhancing echo chambers and reducing opportunities for critical engagement. Despite these potential drawbacks, however, the article concludes by advocating for enhanced critical literacy to address the limitations of LLMs (Tanksley 2024). Such literacy can increase awareness and protect against these diminishing effects, emphasizing the essential role of human mediation. Overall, this analysis aims to stimulate discussion of the epistemic effects of AI innovations and underscore the importance of preserving distinct voices and perspectives amid rapid technological advancements.

2. The Generation of Knowledge

Communal epistemic resources are central to societal progress, yet the processes of knowledge acquisition and distribution often reflect underlying power dynamics, resulting in an uneven allocation of benefits and burdens. Interests shape cognition at all levels, which influences perception, interpretation, classification, and the selection of facts and frameworks. Miranda Fricker (2007) outlines two intertwined processes through which knowledge production accentuates structural inequalities and power imbalances: testimonial injustice and hermeneutical injustice (9, 147).

Testimonial injustice occurs when biases undermine a speaker's credibility, which results in the dismissal of their testimony and the opportunity to contribute to shared knowledge (Fricker 2007, 6). This exclusion precipitates hermeneutical injustice, which manifests when individuals or groups lack the conceptual tools necessary to interpret and articulate their personal experiences, particularly those associated with social harm (Fricker 2007, 147). This injustice stems from gaps in a society's collective interpretative resources, which disproportionately affect marginalized groups.

These groups experience "hermeneutical marginalization," characterized by their exclusion and subordination from participating in the creation and interpretation of collective social meanings (Fricker 2007, 152). Consequently, these groups are disadvantaged in making sense of their own social experiences. This issue is both morally and politically significant; it represents a form of powerlessness, wherein the affected group lacks the necessary interpretive tools to fully understand and communicate their experiences. The lack of language and conceptual tools that accurately reflect the experiences of marginalized individuals within the available hermeneutical resources impedes their ability to articulate their experiences. This gap also hinders others from understanding these experiences, which can lead to misunderstandings or misinterpretations. The dominant epistemic framework lacks the resources to adequately capture these experiences, thus perpetuating a cycle of marginalization. This situation reinforces existing power structures by privileging the dominant group's interpretive frameworks and dismissing or altering the experiences of marginalized individuals to fit these frameworks. As a result, the misrepresentations further entrench marginalization and diminish the epistemic agency of those affected, as their authentic voices and insights are neither acknowledged nor valued within the dominant discourse.

Further, despite widespread recognition of counter-epistemologies, hermeneutical gaps often persist. Linda Martín Alcoff (2007) contends that this ignorance is not simply an accidental oversight that can be corrected by providing additional information; rather, it is structural and resistant to counterclaims and further evidence. Dominant groups, according to Alcoff, "have less interest" in challenging the status quo and "have a positive interest in 'seeing the world wrongly'" (Alcoff 2007, 47). The prevailing social understandings within the hermeneutical resource allow individuals to rationalize their interpretive choices, justifying omissions and emphases that maintain power. This enables individuals to avoid confronting their own ignorance by perceiving the world as complete, even though it represents only one framework among many. Consequently, the exclusionary hermeneutical resource exemplifies how "epistemologies of ignorance" are sustained (Mills 1997, 18).

Charles Mills (1997) describes epistemologies of ignorance as involving a "particular pattern of localized and global cognitive dysfunctions (which are psychologically and socially functional), producing the ironic outcome that whites will, in general, be unable to understand the world they themselves have made" (18). This ignorance is not merely a lack of knowledge but an active production of misunderstanding that serves to uphold existing power structures. By dismissing or invalidating the lived experiences of marginalized groups, the dominant framework perpetuates epistemic injustice. The result is a cyclical reinforcement of ignorance, where the dominant group remains unaware of or indifferent to the systemic inequalities and biases that shape their understanding of the world. Worse, this dominant framework often positions the voices of marginalized individuals to be disregarded or deflected, thereby ensuring that even when these individuals speak out, their perspectives are often ignored or silenced. This dynamic maintains the asymmetry in hermeneutical resources, as the dominant group's interpretive frameworks are privileged and further testimonial and hermeneutical injustice is perpetuated. The epistemic marginalization thus created prevents marginalized groups from contributing meaningfully to the collective understanding of society, which reinforces the power dynamics that keep them in a subordinate position. Consequently, the dominant group's worldview remains unchallenged and unchanged, further entrenching the systemic inequalities that epistemologies of ignorance help sustain.

Marginalized groups face significant challenges to having their experiences acknowledged within broader societal discourse due to epistemic injustices and epistemologies of ignorance. Despite the detrimental effects, feminist standpoint theory suggests that this frustration may also offer epistemological advantages. Standpoint theory exposes the limitations of dominant frameworks that fail to include the experiences of marginalized groups and reveals the artificial and contingent nature of these frameworks. Patricia Hill Collins (1986) describes this as the "outsider within" perspective, which highlights the unique vantage point of those who are both part of and separate from dominant ideologies (514).

By occupying diverse perspectives, marginalized individuals can identify the inherent flaws and biases in dominant epistemic structures due to the generation of "epistemic friction" (Medina 2013, 207). The conflicts between their lived experiences and prevailing frameworks create a "double vision" or "multiplicitous" consciousness. This diverse perspective allows marginalized individuals to see the world from multiple angles and generate a conceptual dissonance that results in epistemic friction between perspectives. Epistemic friction is a critical source of comparative insights that emerges from the interaction of diverse cognitive perspectives and social experiences that challenge and resist one another. The diverse perspective of marginalized groups can illuminate how dominant frameworks obscure or distort reality and can foster critical awareness and drive potential epistemic and social transformation. Therefore, while systemic marginalization perpetuates epistemic injustice, it also equips marginalized individuals with a diverse perspective that can challenge and contest the inadequacies of the prevailing epistemic order, thereby promoting a more inclusive and equitable discourse.

It is important to note that the diverse perspective does not inherently produce truth but rather provides marginalized individuals with heightened awareness of the artificiality in dominant interpretive frameworks which provides an epistemic advantage. Alison Wylie (2012) suggests that "differential access to evidence is rarely an advantage on its own" (347). The epistemic advantage is not automatically accrued by residing in a particular social location; instead, "standpoint theorists often point to a special inferential acuity, a skill at discerning patterns and connections in the available evidence that goes along with sub-dominant status" (Wylie 2012, 347). That is, navigating various cultural contexts results in a diverse perspective that equips marginalized groups—such as people of colour, women, and LGBTQ+ individuals—with a deeper understanding and the ability to identify incongruities between perspectives.

The perspectives of marginalized groups are diverse, shaped by the Intersections of various oppressions, which leads to unique experiences. What these groups share is that their experiences are situated at the periphery of social discourse, creating a dissonance between dominant understandings and their actual lived realities. Privileged individuals, immersed in epistemologies of ignorance, avoid confronting knowledge gaps and overlook the daily realities faced by marginalized groups. Medina (2013) refers to this as "meta-blindness," where individuals inhabit a flattened social reality that prioritizes their own experiences and renders them ignorant of their ignorance (78). For clarity and to contrast this position with those produced by epistemic friction, I will refer to this sort of ignorance as a singular, rather than diverse, perspective.

Homogenization of Content

Standpoint theory posits that knowledge is influenced by individual social positions and emphasizes that marginalized groups can offer unique insights into social structures. It challenges the notion of an objective "view from nowhere," and advocates instead for a "view from everywhere" to fill in hermeneutical gaps where they occur. This approach aims to synthesize multiple viewpoints for a more holistic understanding of reality. However, technological advancements often fail to achieve this epistemic ideal. For instance, while the internet and social media were initially celebrated for democratizing information and enabling grassroots organization, these platforms have been compromised by corporate interests, echo chambers, and polarization. Algorithms that prioritize engagement over truth undermine genuine knowledge democratization. Despite increased access to information, it remains debatable whether this information accurately reflects lived experiences of persons in general. Programs like ChatGPT also fall short in this regard.

Developed using publicly available content from the internet, such as news articles, online forums, books, and digital encyclopedias, the diverse corpus used to train ChatGPT does not necessarily champion varied perspectives; instead, it mirrors and streamlines biases present in the original data. As AI integrates these biases, dominant perspectives—even if biased—may be recognized by the algorithm as statistically significant. Despite efforts to incorporate "guardrails" to prevent content that violates community standards or propagates racism or sexism, and despite techniques such as bias mitigation and continuous oversight, the effectiveness of these measures remains debatable. For instance, research shows that when AI adopts specific personas such as Muhammad Ali, it significantly heightens the risk of perpetuating harmful stereotypes, quickly leading to biased dialogue and offensive viewpoints (Deshpande et al. 2023).

As an "automation of the status quo," it is questionable whether AI can detect these hermeneutical gaps (Fountain 2022, 3). Hermeneutical gaps occur when marginalized groups lack the conceptual tools to interpret and articulate their experiences. These gaps signify more than just data absences; they reflect a deficiency in shared concepts and language essential for making certain experiences intelligible and for generating epistemic friction. AI systems, which learn from existing data and patterns, struggle to represent experiences not yet included in collective knowledge, making comprehensive data representation challenging. Consequently, AI is confined to what is explicit in its training data because it lacks the lived experience to notice gaps, which are unlikely to be represented without explicit prompting. These gaps are often perpetuated in the output of AI models trained on biased data, thereby reinforcing epistemologies of ignorance and representing dominant interests.

Hermeneutical inequality is particularly difficult to identify because interpretive efforts are influenced by interests. This generates "hermeneutical hotspots" in areas where those with power either lack interest in or actively oppose accurate interpretations (Fricker 2007, 152). This disinterest results in, at best, a peripheral recognition of marginalized experiences by those in power, which leads to a lack of necessary context and depth in the articulation of their experiences. Consequently, the perspectives and lived realities of marginalized groups are frequently misrepresented or overlooked in data creation, collection, and interpretation. The data tends to centre on and perhaps exclusively include the experiences of those with social power. These skewed interpretations are then used to frame the experiences of marginalized groups, leading to unrepresentative depictions that justify marginalizing these experiences as outliers, rather than recognizing them as central to the dominant understanding.

The overrepresentation of dominant interests results in representational harms, which involves the unjust distribution of resources and the perpetuation of stereotypes or exclusions (Crawford 2017). Representational harms occur when systems perpetuate the subordination of certain groups based on their identity and lack substantive representation. This bias is readily observed in AI applications that associate "man" with "computer programmer" and "woman" with "homemaker." For example, sentences like ""He is a doctor" are more likely to be generated than "She is a doctor," and sentiment analysis systems often rank sentences containing female noun phrases as indicative of anger more frequently than those with male noun phrases (Sun et al. 2019, 1631). These harms also stem from data collection choices that may reflect traditional classifications. Classification systems, while useful for organizing and understanding complex information, inherently exclude data that does not conform to predefined categories. These systems rely on specific criteria to sort information, which marginalizes data that is ambiguous, overlaps multiple categories, or falls outside the established framework.

These known representational harms raises questions about the training data used in generative AI and which perspectives are prominently represented and which are made invisible. For instance, data from activities like driving in the city or using social media may not represent those using public transit or lacking smartphone access (Fountain 2022). Similarly, digitization often prioritizes collections from well-funded institutions in the Global North, which affects data under-representation (Milligan 2022). The centralization of technological power in a few global locations contributes to data collection tools that can systematically discriminate against marginalized groups, whether intentionally or incidentally.

Representational harm is particularly dangerous because it shapes our perception of reality and upholds pernicious epistemologies of ignorance. These harmful representations limit hermeneutical resources and conceptual vocabulary, creating further asymmetric burdens that lead to continued marginalization and reinforce bias. The result is a "runaway feedback loop," wherein biases embedded in large datasets used for training AI algorithms perpetuate and amplify historical and societal prejudices. This, in turns, makes the biased representation appear as an accurate reflection of reality (Gebru 2020, 256). In the tech industry, marginalized individuals often face hostile environments, diminished recognition, and limited advancement opportunities. These conditions foster false beliefs that these individuals lack the necessary skills or aptitude for their positions. Such misconceptions are then reinforced in the data used by hiring algorithms, thereby further excluding underrepresented groups.

For instance, if an AI system is trained on biased hiring data that underrepresents women in certain roles, it may continue to recommend fewer women for those roles and thus perpetuate the cycle of exclusion. Automated hiring tools, such as those used by Amazon, have exhibited negative biases against women; resumés that reference gender-specific activities or institutions are penalized (Sun et al., 2019). These representational harms replicated in the data reinforce systemic disadvantages for marginalized groups by maintaining existing power structures and inequalities. Hermeneutical gaps prevent these groups from effectively communicating their experiences and advocating for changes in AI systems. These gaps significantly impact AI and are impacted by AI by limiting the interpretive resources available to marginalized groups and entrenching the dominant discourse.

3. Flattened Voice

Standpoint diversity is crucial for AI to foster meaningful and inclusive public discourse, rather than reinforce existing biases and exclude minority perspectives. AI struggles to comprehend nuanced and less statistically prevalent contexts essential for generative epistemic friction. Without diverse contexts to interpret words and phrases, conversations risk misrepresentation or defaulting to the status quo. This homogenization results not only in content reflecting the status quo but also in dialogues and word choices replicating prevailing norms and biases, potentially sidelining minority standpoints and eliminating chance miscommunications. LLMs may not capture the unique idioms, dialects, and styles of various standpoints. This leads to a homogenized voice that fails to represent the richness of diverse experiences and results in flattened communication as the technology becomes more ubiquitous.

AI systems are typically designed to produce consistent and standardized outputs that can suppress diverse standpoints as the AI aims to generate broadly acceptable and non-controversial responses. Writing for Vox, Signal Samuel observes that “they have a tendency to talk in a bland, conformist, Wikipedia-esque way” (Samuel 2023, para. 17). The voice reflected in the output is crucial. Unlike previous language processing models built on expert-created rules or trained on constrained datasets specific to grammar and spelling tasks, current generations of LLMs utilize advanced analytical techniques. The transformer architecture, with multiple layers of self-attention mechanisms, enables the model to understand patterns, context, and relationships between words across extensive text sequences. For example, when predicting a missing word in a sentence, the model considers the relationships and dependencies between all the words in that sentence. This capability allows for a profound transformation in the output. In consequence, using a transformer model as an editing tool might lead to significant but subtle changes in salience, tone, meaning, and connotations, thereby altering the user's original voice to conform to patterns derived from broad, statistically dominant internet sources.

The non-literal and idiosyncratic aspects of communication present significant challenges for technologically mediated exchanges, often marked by representational biases. If the training data primarily includes content created by and for a specific demographic, the AI's voice and style will mirror those biases, regardless of user diversity. The ability of AI to maintain the depth and complexity of human interaction is compromised by its reliance on statistical patterns. This results in the loss of the rich, diverse tapestry of human dialogue, reducing it to a uniform, flattened exchange. Critical engagement arising from miscommunication, where participants navigate and negotiate meaning, is essential for robust and dynamic conversations. When AI fails to replicate this, it diminishes the potential for genuine understanding and the exchange of innovative ideas. The epistemic friction that often arises in human interactions, fiction that pushes individuals to engage critically with differing perspectives, is lost in AI-mediated exchanges.

Much of communication relies on shared frameworks to understand the flow of ideas beyond the literal meaning of words. According to H. P. Grice's (1975) theory of conversational implicature, conversations have implicit goals guiding their flow and content. Grice posits that speakers often imply additional meanings beyond the literal content of their utterances. In everyday conversations, speakers contribute cooperatively through maxims that ensure smooth communication. Implicatures occur when these maxims are flouted. However, with the assumed cooperation, listeners may still infer indirect meanings from context rather than explicit statements. For instance, if John asks Mary if she will attend Paul's birthday party and Mary replies, “I have a lot of work to catch up on,” the implication is that Mary cannot attend, even though she did not state this directly. This understanding relies on shared knowledge between conversational participants to fill in the conversational gaps. These implicatures provide an example of communication barriers that result in epistemic friction and impede the smooth exchange of knowledge and information. Points of miscommunication prompt active interpretation and negotiation, motivating interlocutors to generate additional meanings. This encourages them to critically engage with each other's implicit assumptions and the underlying norms guiding their communication, which ultimately improves mutual understanding. When confronted with uncertain or novel situations, humans often intuitively respond with "I don't know" and seek clarification. AI systems, however, might gloss over these subtle points of miscommunication and defer to data trained on dominant interpretations and understandings. These systems may miss subtle miscommunications and fail to seek clarification or ask follow-up questions. Instead, users are often provided with plausible-sounding and confident answers that can be deeply factually incorrect or under-representative.

Further, as interpersonal communication becomes more ubiquitous, we can foresee scenarios where email exchanges consist solely of one-click responses that strip away nuances. Such interactions may proceed without substantive idea exchange, with friction interpreted purely through statistical probability. This shift toward brevity can expedite decision-making but may lead to communicative complacency, where individuals are less likely to clarify ambiguities or correct misinterpretations. Without miscommunications leading to clarifications, responses become mere products of statistical patterns rather than meaningful dialogue. Through disagreement, errors, and moments of unintelligibility, individuals are prompted to step back and resolve communication issues. In contrast, AI-generated texts and translations inherently lack sensitivity to hermeneutical gaps or an awareness that intelligibility may reach its limits in each context. AI-mediated answers tend to "hallucinate" and present falsities under the guise of an authoritative tone (Metz 2023). Consequently, epistemic friction gives way to seamless communication.

Some have already experienced this homogenization and loss of voice as a result. According to Halcyon M. Lawrence, representational biases that favour English within these datasets echo historical linguistic imperialism. She observes, "For millennia, civilizations have effectively leveraged language to subjugate, even erase, the culture of other civilizations" (Lawrence 2021, 474). Lawrence underscores that English continues to predominate online informational spaces, comprising fifty-one percent of web pages as of November 2017. This dominance leads to the vast underrepresentation of other languages, thereby suppressing diverse voices. While LLMs have often performed well in translation tasks, the voice of the output tends to be heavily influenced by Western English-language use. Content related to economically disadvantaged countries is underrepresented in the training data, leading to less accurate predictions and occasionally resulting in the omission or neglect of these regions and their nuanced dialects in the models' outcomes.

Gebru further illustrates these biases with a recent incident wherein a Palestinian's Arabic post saying "good morning" was erroneously translated by Facebook Translate as "hurt them" in English and "attack them" in Hebrew (Gebru 2020, 264). Platforms created by major technology firms like Google and Facebook are frequently geared towards translations between English and other Western languages. This orientation reflects the linguistic preferences of researchers and the concentration of funding, particularly in places like Silicon Valley. Consequently, there is a distinct bias towards resolving translation issues between languages like English and French, while languages like Arabic are neglected. Had the field of language translation been more inclusive of Arabic-speaking populations and other underrepresented languages, it is conceivable that such an error might not have occurred.

It is also important to recognize that the predominance of Western English in the voice of the output diminishes the diversity of linguistic styles to which users are exposed. Lawrence emphasizes the overrepresentation of this voice, arguing that it negates specific benefits associated with hearing linguistic diversity, such as foreign-accented speech. These advantages include not only enhanced comprehension of various styles but also transformed attitudes towards speakers with accents, owing to increased familiarization. With a growing exposure to outputs generated by such models, a subtle normalization of the dominant dialect occurs, reinforcing the perception of foreignness among accented speakers. This lack of representation in linguistic diversity confirms "the prevailing misconception that accents are not only undesirable but unintelligible for use in speech technologies" (Lawrence 2021, 491). Connected to the absence of diversity in the output is the system's capacity to establish a standard that may implicitly marginalize various linguistic expressions. This can perpetuate stereotypes against such expressions, which further exacerbates the exclusion of those whose dialects do not neatly conform. The dominance of Western English in AI and LLM outputs perpetuates historical patterns of linguistic imperialism, marginalizes non-Western languages, and reduces exposure to linguistic diversity.

4. Communicative Complacency and Echo Chambers

It might be argued that failures, such as mistranslations, are productive as they highlight blind spots in the data, can prompt greater attention to standpoint diversity, and can potentially motivate efforts to address these deficiencies. While AI technology is still in its infancy, its ability to capture unique voices shows promise. With sufficient time and input, AI may overcome its tendency to default to dominant language patterns through more deliberate example-level instruction engineering. This approach trains AI using specific, context-rich examples designed to teach it how to interpret and generate language with greater nuance. By clarifying contextual meanings, enhancing pattern recognition, and fostering pragmatic understanding, example-level engineering helps AI better grasp subtleties such as tone, intent, and cultural diversity. By incorporating diverse perspectives into its training, this method enables conversational agents to more accurately reflect often-overlooked viewpoints. The current "flattened" voice of AI may be less a limitation of its potential and more a symptom of its developmental immaturity.

An example of AI's evolving capability in mimicking human communication is "grief tech," which allows individuals to capture their essence in a chatbot that interacts with loved ones after their death (Fitzhugh-Craig 2023). “Grief bots,” or “ghostbots,” simulate conversations with deceased individuals using extensive data from their digital footprints. For instance, Michael Bommer, facing terminal colon cancer, created an interactive AI version of himself to support his loved ones after his passing (Kelly 2024). He recorded 300 sentences to capture his voice nuances and 150 stories detailing his life experiences and principles. These provided the AI with content to construct personalized responses that reflect Bommer's knowledge and personality. This ensures his digital avatar offers guidance and reassurance aligned with his approach to life.

However, despite the potential for replicating a unique voice and idiosyncratic contexts, significant concerns remain. While AI may eventually master maintaining an individual's conversational style and perspective, its increasing prevalence reduces opportunities for meaningful unmediated communication. There are immediate concerns that this emerging technology, driven by for-profit interests, may prioritize increasing engagement over fostering genuine communication, which is particularly troubling in the context of grief tech. Additionally, as AI-generated suggestions and communication become more integrated into daily activities, there are concerns that increased and more personal interaction with AI could significantly alter how we communicate, potentially reducing the depth and complexity of our interactions. Notably, misunderstandings and mistranslations do not prompt the system to halt and investigate its errors; the system simply produces another text that irons out these communicative wrinkles, thereby ensuring epistemic friction is avoided without being addressed. The experience of dissonance between one's conception of the world and the way it is being framed is significant. Such dissonance is crucial because moments of miscommunication reveal failures in shared understanding, underscoring that shared meaning cannot always be assumed or dismissed.

Employing LLMs, even for creative writing or brainstorming, can guide users toward specific responses that conform to pre-existing norms. Consequently, certain avenues of thought may become underrepresented or neglected altogether, an outcome that would direct thought and normalize particular viewpoints. Should AI-generated content become normalized, it might enter its own data training and perpetuate a closed feedback loop where it draws upon itself, continually feeding and enlarging its own biases like an informational ouroboros.

The self-perpetuating characteristics of LLMs can inadvertently reinforce what Thi Nguyen terms "epistemic bubbles" (Nguyen 2020, 141). Nguyen defines an epistemic bubble as an informational context where certain perspectives are consistently underrepresented or disregarded. These bubbles often emerge from the natural dynamics of social alignment and community formation. Social networks, acting as channels for disseminating information, further amplify these shared beliefs and create an epistemic filter that resists differing opinions, reinforces confidence in confirmatory information, and perpetuates homogeneity within discourse. This can inadvertently lend credibility to potentially harmful content by maintaining and exacerbating existing biases and a notable lack of opportunities for friction.

A similar reduction in opportunities for friction occurs when current AI technologies are employed for informational searches. The subtleties and context essential to research are frequently overlooked or stripped away, often presented without counter-information unless specifically requested. The resulting responses are typically tailored, influenced by algorithms and corporate interests, and delivered without adequate context, nuance, or source attribution. This lack of critical elements facilitates their unchallenged acceptance. For example, the landing page of conventional search engines like Google is often inundated with content created for advertising purposes, the company's primary commercial interests. Cory Doctorow (2023) likens the function of search results to a "payola" system where top visibility is granted to the highest bidder. The organization of websites within search results constructs a perceived hierarchy, wherein higher-ranked sites may be regarded as more credible and relevant. This preferential positioning drives user engagement, further enhancing a website's statistical relevance and reinforcing its prioritization by the algorithm. Currently, platforms such as Google have assigned this top position to their AI-generated overviews, further centralizing user engagement and consolidating information access within algorithmically curated content.

The complexity of this issue is amplified by the algorithms' opaque nature which leaves users with little information about content selection and prioritization. This lack of transparency exacerbates the difficulty in understanding the mechanisms that govern information production and can lead to confusion about how content is selected and prioritized. Users may not be unaware of how much the information has already been mediated by the time ChatGPT answers them with an authoritative voice. Acting more as content curators than providers of a comprehensive overview, informational searches create an illusion of exhaustive research. The outcomes, prompted by a user's selection of terms, foster a belief that these results are unbiased answers to their queries rather than products significantly influenced by corporate agendas and algorithmic biases.

Information obtained through search engines often becomes detached from its original context and the underlying biases of the medium. This constructs an appearance of neutrality and objectivity, further reinforced by the "Wikipedia-esque" tone (Samuel 2023, para. 17). This guided navigation contributes to what Eli Pariser (2012) describes as a "filter bubble" that stems from personalized technology's filtering function. Algorithms process vast amounts of user data to furnish hyper-targeted content, aligning selections with user habits and analogous profiles. By doing so, they guide decisions, often sideline personal evaluations, and create a decision consensus through repeated analogous choices. These selections are then logged and prioritized in future recommendations to establish a self-sustaining feedback loop that buttresses pre-existing beliefs. Informational searches rely on queries supplied by the user, and fine-tuning these inquiries can result in highly skewed results, just as it can generate accepted facts. Likewise, LLMs further enable targeted searches that conceal and minimize encounters with counterevidence that might otherwise create dissonance or friction with one's held beliefs.

These systems often present information with markers of credibility, especially when the output aligns with a user's existing beliefs and is curated through chosen prompts or when a specific perspective is normalized within the user's informational ecosystem. This reinforces the acceptance of potentially biased or uncited information. With the help of LLMs, users may not simply ignore websites containing countervailing information—these sites may not appear at all, thus lending further credibility to the skewed belief.

Safiya Umoja Noble (2018) highlights the potential harm in disseminating misinformation through limited search engine results. Noble underscores the critical role of online platforms in fostering harmful ideologies, as illustrated by the case of Dylann Roof. In 2015, Roof, a self-identified white supremacist and neo-Nazi, executed a tragic shooting at the Emanuel African Methodist Episcopal Church in South Carolina. His online manifesto disclosed that his motivations were shaped by internet searches centered on "black on white crime" that lead him towards a restricted, aggressive, and racially prejudiced perspective. These search outcomes glaringly lacked counterarguments, anti-racist resources, or comprehensive understandings of groups such as the Council of Conservative Citizens, notorious for their anti-Black, anti-immigrant, anti-LGBTQ, and anti-Muslim positions.

Additionally, the search term “black on white crime” often failed to provide links to authoritative discussions on race or credible sources that outline the historical dynamics of racial relations in the US. This limitation in search results can result in "credibility excess," where certain sources are disproportionately credited with reliability and authority (Fricker 2007, 17). As users often do not explore beyond the first few results, these initial sources can unduly influence their understanding and opinions, even when the information is biased, incomplete, or misleading. Consequently, the perceived credibility of these sources is artificially inflated by their prominence in search results, which can contribute to the spread of potentially harmful ideologies and misinformation.

The decontextualized information offered by ChatGPT operates similarly to a targeted keyword search. Ian Milligan likens the use of keyword searches to gleaning information with almost “surgical” precision (Milligan 2022, 27). While this technological advancement facilitates knowledge acquisition and enables users to efficiently sift through vast amounts of data, it forfeits the valuable skill of skimming. As Milligan observes, “If a user simply relies on keyword searches within historical collections, they would be unaware of what they are missing. The absence of search results might be interpreted as a complete lack of relevant information, when in fact it could be an inaccurate representation of reality” (Milligan 2022, 24). ChatGPT functions analogously, as the answers provided are restricted to those explicitly requested, and notably with less contextual information than might be obtained from a keyword search that at least directs to citations and further research opportunities. The monological responses supplied by ChatGPT, particularly without source information, diminish the likelihood of encountering opposing views and areas of contention.

Users receive data congruent with their accepted social understanding, often finding endorsement within online communities. In today's pervasive social media usage, once private or selectively shared views are now widely distributed. Such dissemination often garners endorsement from familiar community peers whose views might otherwise have been overlooked. This familiarity enhances and normalizes these opinions, integrating them into individual worldviews. Pervasive social media usage exposes users to increased consensus, unjustifiably boosts confidence, and facilitates widespread acceptance of platform-shared opinions. This broad reinforcement of prevailing worldviews fosters a cycle of credibility and familiarity that subtly directs individuals to adhere to specific narrative frameworks, particularly when contrasting views are lacking or must be explicitly sought. Social media algorithms that mold content based on previous user choices build an online filter surrounding digital interactions. This content prioritization for engagement might conflict with users' actual interests, but interacting with this content causes it to permeate feeds and cement an epistemic bubble. For those with biases, the algorithm's effect is magnified, exaggerating even slight initial differences, overshadowing contrary views, and trapping users in a self-reinforcing loop. This process narrows opportunities for engaging with diverse perspectives, reinforces existing beliefs, and deepens epistemic bubbles through what Nguyen calls “bootstrapped corroboration,” where epistemic bubbles exclude key information thereby creating an inflated sense of epistemic self-confidence (Nguyen 2020, 144).

Unique to LLMs is the all-encompassing nature of the information they provide. The consolidation of such large data sets can inadvertently overlook varied perspectives and voices, leading to a loss of the richness and objectivity that diversity brings to the data. Users are no longer required to navigate multiple sites to obtain the desired information; they can simply query systems like ChatGPT, now accessible on phone apps. This approach creates a singular information source that not only reinforces prevailing societal narratives but also serves as a comprehensive platform for information retrieval. Information—or more often, misinformation—gains credibility by resonating with pre-existing narratives that originate from a source that appears authoritative, and conveniently excluding counterarguments due to the monological nature of the responses. Such narratives may be perceived as having augmented authority that surpasses traditional markers of academic credibility inherent in the text generated. Consequently, the chance discovery of evidence that contradicts these dominant narratives becomes progressively uncommon. This phenomenon is particularly pronounced in online communities, where echo chambers foster environments that normalize and strengthen prevailing viewpoints and promote a singular, rather than diverse, perspective.

Even when individuals within an echo chamber encounter opposing views or evidence, these interactions are unlikely to challenge or alter their established beliefs. This is due to the phenomenon of “epistemic inoculation”, where members of an echo chamber are conditioned to distrust and pre-emptively discredit any external information or perspectives (Nguyen 2020, 147). As a result, when these members do encounter opposing views, they interpret them as anticipated and consistent with the warnings they have been conditioned to expect. This process serves to validate the information and theories previously endorsed by the echo chamber, thereby neutralizing any potential impact that contrary views might have had. This pre-emptive discrediting not only shields their existing beliefs from scrutiny but also reinforces their trust in the echo chamber's perspective.

The overconfidence that comes with persistent confirmation and validation cements existing misconceptions and renders users incapable of recognizing the need for change, let alone the need to initiate such change independently. This situation is further complicated by the "coverage-reliance ignorance" that social media echo chambers often cultivate (Bayruns García 2020, 414). Eric Bayruns García defines this type of ignorance as occurring when individuals form beliefs—false or unjustifiably true—based on the mistaken presumption that their informational sources are reliable within a specific domain. However, these sources, compromised by inherent injustices, fail to consistently provide dependable information. Users thus become inadequately attuned to the reliability of their sources to unveil and broadcast pertinent information in these spheres. This insensitivity is partly due to the epistemic friction that is conspicuously absent in pervasive echo chambers and leads to uncritical consumption of information and the perpetuation of biased datasets.

5. Reintroducing Wrinkles

As I have argued thus far, identifying gaps within the discourse necessitates standpoint diversity—knowledge derived from diverse lived experiences relative to dominant social systems. However, AI systems struggle with implied meanings and cultural nuances; they typically default to the status quo when uncertain, thereby undermining standpoint diversity. Efforts to mitigate this bias include pretraining on large datasets, fine-tuning, dissociating stereotypical associations within models, and conducting regular bias audits using techniques such as reweighting data and disparate impact testing to promote fairness (Bolukbasi et al. 2016). Mitchel et al. advocate for increased transparency in datasets and suggests detailed information about usage, potential pitfalls, and inherent biases, akin to a nutrition label on food products (Mitchell et al. 2019, 221). Others propose using synthetic data to counter real-world biases and avoid issues with copyrighted data (Reed 2024).

Such proposed solutions, however, do not necessarily eliminate the need for human mediation within the system. Synthetic data, for example, can create diverse and representative datasets by oversampling underrepresented groups, balancing class distributions, and controlling confounding variables that introduce bias. This approach aims to improve representation and model performance without compromising data integrity (Lee 2024, 22). However, the effectiveness of synthetic data in eliminating bias depends on careful design and continuous monitoring to ensure alignment with real-world data. Human AI trainers play a crucial role by providing demonstrations and comparative evaluations to guide the model's responses and ensure context-sensitive judgments. Nevertheless, even with rigorous monitoring, achieving such diversity is uncertain. Furthermore, reliance on human oversight can exacerbate inequitable worker treatment. Gray and Suri (2019) use the term "ghost work" to refer to labour that is costly, error-prone, and often involves poor working conditions. They highlight a global underclass engaged in tasks like content moderation and transcription. Without sufficient diversity in both data and the "human-in-the-loop" approach—where human judgment is integrated into the process of developing or refining algorithms— this process may replicate and widen hermeneutical gaps while further entrenching global disparities.

It is more effective to address hermeneutical gaps through human intervention more generally rather than rely solely on diversity within AI systems. Human mediation in AI is crucial for providing the context-sensitive judgments that AI systems inherently lack. As Fricker (2007) posits, promoting hermeneutical justice requires sensitivity to the speaker's interpretive resources, identification of areas of struggle, and awareness of "hotspots" for misinterpretations (Fricker 2007, 152). This necessitates context-sensitive judgment, active listening, and corroborating evidence from similar social experiences. Enhancing general critical literacy can better address hermeneutical gaps by promoting friction from a diversity of standpoints and avoiding the exploitation currently used in AI training. By ensuring that users deeply understand the data they are using, the charge of complacency might be better mitigated.

For example, Tiera Tanksley (2024) examines Critical Race Algorithmic Literacies (CRAL) and emphasizes its historical role in literacy as a tool for emancipation. Historically, literacy has served to subvert oppressive systems, with contemporary bans on critical race theory viewed as efforts to suppress critical literacies that challenge systemic racism. The digital mediation of critical literacy enables users to assess data sources, methodologies, and implications, fostering a nuanced approach to data usage. Tanksley highlights the significant impact of CRAL in empowering Black students to navigate and challenge algorithmic racism in educational contexts. Through CRAL, students acquire the skills to scrutinize AI technologies by identifying racially biased tools such as Google, ChatGPT, and Grammarly as ineffective for fostering inclusive educational experiences. This literacy enabled students to link negative experiences from traditional educational practices, such as low teacher expectations and zero-tolerance policies, to AI-mediated inequities.

Furthermore, CRAL facilitated students in exposing and critiquing racial biases within AI systems and educational policies, and in reimagining AI applications that prioritize equity and well-being. Additionally, students challenged the uncritical adoption of AI by educational institutions and advocated for thorough bias audits and rejection of simplistic colourblind approaches. Through CRAL, students redefined effective AI usage and promoted the critical use of technologies to advance educational equity and disrupt systemic racism, thus preparing them to thrive within settings using educational AI technology. This approach works to ensure the necessary diversity for epistemic friction because it equips users with the cognitive tools necessary to navigate and interpret complex information and potentially addresses hermeneutical gaps more effectively than merely ensuring diversity in the data itself. By focusing on human intervention and critical literacy, we can better prepare users to engage with AI systems and digital platforms in a manner that promotes justice and understanding across diverse perspectives.

6. Conclusion

Promoting standpoint diversity requires more than diversifying datasets; it necessitates empowering users to actively engage with and critique AI outputs. Expanding critical literacies among all technology users is essential to prevent echo chambers by reintroducing epistemic friction in interactions with AI. Through critical engagement with AI, and by recognizing its limitations, users transition from passive consumers to active contributors in the process of knowledge creation. This critical awareness encourages users to consistently question AI outputs and thereby reduces the risk of hermeneutical gaps and the kind of complacency that fuels echo chambers. As users become more attuned to dissonance, they are better equipped to identify inconsistencies and contradictions, which prompts recognition of instances where diversity in perspective is lacking. Ultimately, this approach ensures that diverse viewpoints are not only represented but fully engaged with and can thereby lead to more dynamic and epistemically textured interactions with technology.

Overall, it may not be possible to "program away" biases, as dissonance relies on experience—something the system fundamentally and perhaps perpetually lacks. Therefore, human mediation remains crucial. Standpoint diversity is achieved by empowering users to critically engage with AI, thus ensuring a more inclusive and just technological landscape.

Works Cited

Alcoff, Linda Martín. 2007. "Epistemologies of Ignorance: Three Types." In *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana, 39–57. Albany: State University of New York Press.

Bayruns García, E. 2022. "How Racial Injustice Undermines News Sources and News-Based Inferences." *Episteme* 19 (3): 409–30. doi.org/10.1017/epi.2020.35.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4349–57.

Collins, Patricia Hill. 1986. "Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought." *Social Problems* 33 (6): S14–S32. www.jstor.org/stable/800672.

Crawford, Kate. 2017. "The Trouble with Bias." Paper presented at the Neural Information Processing Systems (NIPS) Conference, Long Beach, CA, December 4–9.

Deshpande, Abhay, Vinay Murahari, Tirthraj Rajpurohit, Abhishek Kalyan, and Kartik Narasimhan. 2023. "Toxicity in ChatGPT: Analyzing Persona-Assigned Language Models." Manuscript in preparation.

Doctorow, Cory. 2023. "The 'Enshittification' of TikTok." *Wired*, January 23. www.wired.com/story/tiktok-platforms-cory-doctorow/.

Fitzhugh-Craig, Martha. 2023. "Grief Tech Takes End-of-Life Planning to Another Level." *Information Today* 40 (6): 35–36.

Fountain, Jane E. 2022. *Digital Government: Advancing E-Governance through Innovation and Leadership*. Cambridge, MA: MIT Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gebru, Timnit. 2020. "Race and Gender." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 16. Oxford: Oxford University Press. doi.org/10.1093/oxfordhb/9780190067397.013.16.

Grey, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Chicago: University of Chicago Press.

Grice, H. P. 1975. "Logic and Conversation." In *Syntax and Semantics*, edited by Peter Cole and Jerry L. Morgan, Vol. 3, 41–58. New York: Academic Press.

Kelly, Mary Louise, host. 2024. "He Has Cancer—So He Made an AI Version of Himself for His Wife After He Dies." *Consider This* (podcast), June 12. NPR. www.npr.org/transcripts/1198912621.

Lawrence, H. M. 2021. "Siri Disciplines." In *Your Computer Is on Fire*, edited by Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip, 121–35. Cambridge, MA: MIT Press.

Lee, Peter. 2024. "Synthetic Data and the Future of AI." *Cornell Law Review* 110 (forthcoming). ssrn.com/abstract=4722162.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. New York: Oxford University Press.

Metz, Cade. 2023. "Chatbots Hallucinate Information Even in Simple Tasks, Study Finds." *New York Times*, November 6. www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html.

Meuse, Matthew. 2023. "Bots Like ChatGPT Aren't Sentient. Why Do We Insist on Making Them Seem Like They Are?" CBC Radio, March 17. www.cbc.ca/radio/spark/bots-like-chatgpt-aren-t-sentient-why-do-we-insist-on-making-them-seem-like-they-are-1.6761709.

Milligan, Ian. 2022. *The Transformation of Historical Research in the Digital Age*. Cambridge: Cambridge University Press.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model Cards for Model Reporting." In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19)*, 220–29. New York: ACM. doi.org/10.1145/3287560.3287596.

Nguyen, C. Thi. 2020. “Echo Chambers and Epistemic Bubbles.” *Episteme* 17 (2): 141–61.

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

Pariser, Eli. 2012. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Books.

Reed, Ronan. 2024. “Does ChatGPT Violate New York Times’ Copyrights?” *Harvard Law School*, March 22. hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/.

Samuel, Sigal. 2023. “What Happens When ChatGPT Starts to Feed on Its Own Writing?” *Vox*, April 10. www.vox.-com/future-perfect/23674696/chatgpt-ai-creativity-originality-homogenization.

Sun, Tony, et al. 2019. “Mitigating Gender Bias in Natural Language Processing: Literature Review.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5459–68.

Tanksley, Tinisha. 2024. “Critical Race Algorithmic Literacies: A Framework for Black Liberation.” *Journal of Media Literacy Education* 16 (1): 32–48.

Wylie, Alison. 2012. “Feminist Philosophy of Science: Standpoint Matters.” *Proceedings and Addresses of the American Philosophical Association* 86 (2): 47–76. doi.org/10.2307/20620467.