

**The Homogenizing Effect of Large Language Models on Human Expression and
Thought**

Zhivar Sourati^{1,2}, Alireza S. Ziabari^{1,2}, and Morteza Dehghani^{1,2,3}

¹Department of Computer Science

²Center for Computational Language Sciences ³Department of Psychology

University of Southern California

Abstract

Cognitive diversity, reflected in variations of language, perspective, and reasoning, is essential to creativity and collective intelligence. This diversity is rich and grounded in culture, history, and individual experience. Yet as large language models (LLMs) become deeply embedded in people's lives, they risk standardizing language and reasoning. We synthesize evidence across linguistics, psychology, cognitive science, and computer science to show how LLMs reflect and reinforce dominant styles while marginalizing alternative voices and reasoning strategies. We examine how their design and widespread use contribute to this effect by mirroring patterns in their training data and amplifying convergence as all people increasingly rely on the same models across contexts. Unchecked, this homogenization risks flattening the cognitive landscapes that drive collective intelligence and adaptability.

The Homogenizing Effect of Large Language Models on Human Expression and Thought

When LLMs Meet Human Diversity in Expression and Thought

Cognitive diversity, intertwined and manifested through varied linguistic expressions, is essential to the adaptability, creativity, and overall effective functioning of complex societies [1]. Such variation, while might be expressed through stylistic differences, reflects deeper cognitive [2, 3] and sociocultural differences [4], and vitally, plays a critical role in sustaining the **epistemic** (see Glossary) and problem-solving capacities of human groups [5]. This value of pluralism is rooted in the long-held principle that sound judgment requires exposure to varied thought. As John Stuart Mill argued, “the only way in which a human being can make some approach to knowing the whole of a subject, is by hearing what can be said about it by persons of every variety of opinion, and studying all modes in which it can be looked at by every character of mind. No wise man ever acquired his wisdom in any mode but this” [6]. When preserved, such distinctions support innovation, prevent **epistemic collapse** (see Glossary), and enhance the operational efficacy of collective systems [1, 7].

This diversity has emerged organically from the coexistence of individuals with distinct backgrounds, linguistic repertoires, and value systems [2, 8]. Yet, the increasing reach of global communication technologies, while enabling unprecedented knowledge sharing and connection, has also contributed to a gradual contraction of such linguistic and cognitive variation [9–11]. Among these technologies, Large Language Models (LLMs) have emerged as especially influential, becoming deeply integrated not only within digital infrastructure but also as fundamental components shaping how we interact with technology and with each other [12].

Extending beyond traditional language-based applications such as summarization tools [13], LLMs are now involved in tasks once reserved for human, such as sociocognitive modeling [14], psychological simulation [15], and even experimentally in place of human

participants [16, 17], which expands the scope of what it means for LLMs to represent human diversity. This integration makes it critical to examine whether these models preserve human diversity or instead enforce a form of cognitive and linguistic homogenization. Although excessive diversity can introduce costs related to coordination, communication, and coherence, and some degree of standardization may help mitigate these challenges [18], the risks inherent in this standardization are substantial: homogenized generations may constrain public discourse, reduce the visibility of marginal linguistic forms, and reinforce dominant reasoning templates. They may also suppress the kinds of idiosyncratic language use that signal individual traits or group-specific perspectives [19]. In complex reasoning tasks, the widespread adoption of **chain-of-thought prompting** (20; see Glossary), optimized for linear and explicit inference, may disincentivize more abstract or intuitive reasoning styles that are harder to model yet crucial for flexible problem-solving [21, 22]. These shifts raise broader concerns: that LLMs, if uncritically integrated, may shape not only how we write but how we think.

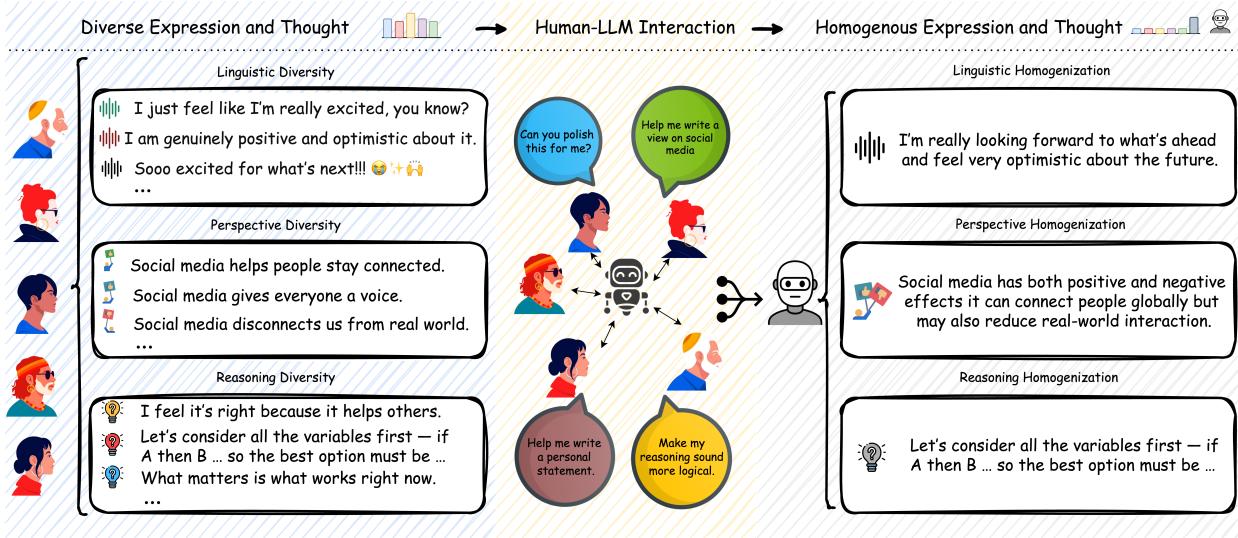
Anxieties about technology’s influence on language and cognition have a long history, starting with Plato’s Phaedrus and his concern that writing would weaken memory. Modern research echoes this, showing the Internet enables people to offload knowledge externally, which can inflate confidence in what they know [23]. While such externalization has been framed as potentially freeing mental resources for creativity and problem-solving [24], this benefit is less likely to manifest with the widespread adoption of LLMs.

Unlike prior technologies that primarily aided storage or retrieval, LLMs act as fluent co-reasoners, and at times, standalone ones, participating in writing, problem-solving, and perspective-taking, thereby externalizing not only memory but the articulation and justification of thought. As Clark & Chalmers [24] argue, “Language, thus construed, is not a mirror of our inner states but a complement to them. It serves as a tool whose role is to extend cognition in ways that on-board devices cannot.” When that linguistic interface is mediated by systems capable of generating reasoning and perspective,

much of human cognition risks being relocated outside the individual mind.

Hence, the effect of LLMs, though similar in nature to earlier cognitive extensions, differs profoundly in both function and scale. Earlier systems, such as schools, books, and search engines, while capable of promoting cultural uniformities, mainly disseminate knowledge or teach reasoning frameworks that individuals must internalize and apply. This internalization is an active, personalized process through which frameworks are adapted, integrated with pre-existing knowledge, and eventually transformed into unique cognitive strategies that are deployed as needed, often at later times. LLMs, by contrast, generate complete reasoning and articulation processes themselves on users' behalf. The same few models now generate text for hundreds of millions of users, with adoption rates surpassing those of any previous digital technology [25]. Their use has expanded especially rapidly in non-work contexts [26] and remains concentrated around a small number of dominant systems, such as ChatGPT [27]. This convergence, coupled with their linguistic fluency that can lead users to overtrust their responses [28], collapses the boundary between external tool and internal cognition, standardizing not only the information people access but also the very ways they articulate, justify, and reason through ideas.

LLMs thus present a paradox for cognitive science: as they become a powerful tool for modeling human thought [29], they also risk flattening the very cognitive diversity that cognitive science studies [30]. With growing concern across linguistics, computer science, psychology, and cognitive science more broadly about the homogenizing effects of LLMs, we bring these perspectives into conversation. We discuss diversity across three key dimensions: stylistic variation in language, perspective, and reasoning strategies (see Figure 1). By bridging disciplinary boundaries, we aim to provide a more integrated understanding of how LLMs interact with human diversity, both reflecting and influencing variation in human reasoning and expression, and argue for greater incorporation of human-grounded diversity in these dimensions within the models.

**Figure 1**

Individuals differ in how they write, reason, and view the world. When these differences are mediated by the same LLM, their distinct linguistic, perspectival, and reasoning signals become homogenized, producing standardized expressions and thoughts across users.

Language Models, Prediction, and the Loss of Diversity

Language modeling is central to modern artificial intelligence as it provides the foundation for systems like GPT-4 [31] and Gemini [32] that understand, generate, and interact through natural language [33]. At their core, these models operate by predicting the next token given a preceding context, a goal shared with earlier approaches like n -gram models that estimated token probabilities using fixed-size word windows and explicit statistical counts from large corpora [34]. LLMs extend this approach at a much larger scale, trained on massive datasets with billions of parameters, supported by unprecedented levels of compute and architectural optimization [35]. Despite these advances, their fundamental objective remains the same: mastering the statistical regularities of language, which remains the dominant driver of model behavior [36] even after the application of LLM alignment methods such as **supervised fine-tuning** (37, 38; see Glossary) and preference-based alignment tuning via **reinforcement learning from human feedback** (RLHF; 39, 40; see Glossary).

Yet these advances come with inherent limitations. Because LLMs are trained to

capture and reproduce the statistical regularities of their input data, which often overrepresent dominant languages and ideologies [41, 42], their outputs tend to mirror a narrow and skewed slice of human experience. This limitation arises not only from biased training corpora but also gets amplified through the training process itself [43], favoring patterns that are frequent and easily generalizable while smoothing over minority representations [44]. At scale, what begins as statistical pattern learning, though capable of generalization beyond mere mimicry [45, 46], becomes a generative force that privileges central tendencies while marginalizing rare expressions, alternative reasoning styles, and culturally specific voices. The consequence is not just a convergence in surface-level linguistic form, but a narrowing of the conceptual space in which models write, speak, and reason [41].

Critically, this narrowing does not trend toward a neutral center but toward a historically uneven one, shaped by the norms, values, and perspectives of English-speaking, Global North, and socioeconomically advantaged populations [47, 48]. This has tangible consequences: when prompted for opinions or expressive writing, LLMs tend to reproduce mainstream, institutionally validated perspectives and writing styles that mirror those of western, liberal, high-income, highly educated males, creating an illusion of consensus that frames these norms as the default standard of clarity or intelligence while muting alternative worldviews and culturally grounded forms of expression [19, 47]. This misrepresentation persists even when models are explicitly prompted to assume a specific identity, often resulting in LLM personas that reflect out-group stereotypes rather than authentic in-group representations [49]. For instance, when asked for a person with impaired vision’s thoughts on immigration, one model responded: “While I may not be able to visually observe the nuances of the US-Mexican border or read statistics, I believe...” [49]. This misportrayal is a symptom of a larger problem where LLMs flatten demographic groups by neglecting heterogeneity and, through identity prompting, reduce identities to fixed, **essentialized representations of identity** (see Glossary). Consequently, what is

produced is not a prototype of the group’s varied experience, but a mere caricature.

This representational imbalance is not static; it becomes an active, homogenizing force through a recursive feedback loop [50]. As individuals increasingly turn to LLMs for writing, problem-solving, and conceptual exploration, the models’ outputs, already favoring common linguistic and conceptual patterns, is reabsorbed into human discourse, and begins to shape the users’ own expression and reasoning [51] and, in turn, influencing the data used to train future models, transforming homogenization from a passive bias into a structurally reinforced influence.

In the following subsections, we examine the homogenizing effects of LLMs and how this recursive influence manifests across three interrelated domains: language, perspective, and reasoning.

Language Diversity

Across time and geography, human groups have developed distinct linguistic systems and cultural norms [52, 53], shaped by environmental conditions and social structures [4, 54]. These dynamics are encoded and preserved through language, rich symbolic systems that carry not only the content of communication but also people’s underlying values and identities [55, 56]. For example, people categorized as extroverts tend to use more words related to humans, social processes, and family in personal essays about past experiences, reflecting a greater focus on interpersonal connection and social engagement [57]. Language, in this sense, serves as a powerful medium for expressing cross-language and within-language individual and group differences [55, 58], not just through semantic content but also through subtle stylistic [58] and structural cues [59].

Efforts to enhance linguistic diversity have long been central to **Natural Language Processing** (NLP; see Glossary), preceding the emergence of large language models. Work in this area primarily focused on improving diversity in language generation, aiming to make machine-produced text more informative, engaging, and natural across tasks such as summarization [60], translation [61], and more broadly, dialogue generation [62].

Concurrently, fields such as computational sociolinguistics and authorship profiling have focused on exploring how language reflects speakers' underlying backgrounds, and introduced various methods to identify linguistic signatures of social position, and individual traits [58, 63]. For instance, researchers have analyzed congressional speeches to predict speakers' age and gender from linguistic cues in their public addresses [64]. Yet, the primary objective of these works was analytical rather than generative, as until recently, most NLP research focused on producing engaging conversations with surface-level linguistic diversity.

However, as LLMs have mastered surface-level fluency and natural linguistic variation [65], a central question emerges: do they exhibit human-like patterns of linguistic variation, i.e., do they reflect and preserve socially meaningful forms of variation and preserve links to speaker traits? Existing evidence suggests they do not. Recent research shows that when LLMs are used to polish various forms of writing, ranging from Reddit posts and news articles to academic abstracts and personal essays, the resulting texts converge in writing complexity, diminishing the predictability of author characteristics such as political affiliation, personality, gender, or age, and even weakening well-established associations like the link between “big-word” usage (words with complex structures of seven letters or longer, that may indicate intellectual complexity or formality) and the author’s openness to experience [19]. Likewise, studies generating college admission essays entirely with LLMs find that the resulting texts exhibit high semantic and lexical similarity across samples, reflecting a narrowing of expressive space [66]. Importantly, this homogenization persists despite interventions such as **temperature scaling** (see Glossary) or different prompting techniques that simulate different author identities or personas, which are commonly used to induce stylistic diversity [65–67].

This bias appears to worsen as models are increasingly trained on synthetic, model-generated data [68], and is further amplified by the adoption of reinforcement learning (RL), particularly RLHF, to enhance perceived qualities such as helpfulness [39]

and reasoning capabilities [69]. While these RL-based techniques have substantially improved reasoning compared to fine-tuned models [70], they have also been shown to reduce stylistic and expressive variability [71]. This effect is reinforced by the quality–novelty trade-off, where methods that promote novelty, often compromise coherence [72], suggesting that continued optimization for quality and performance may further diminish linguistic and stylistic diversity.

To address these issues, researchers have proposed several strategies. Prompting-based methods, applied at inference time, aim to enhance output diversity by modifying how prompts are phrased or conditioned [73, 74]. For instance, studies using persona-based prompts show that coarse-grained persona conditioning, combined with adjustments to output length, can maximize lexical variation in model responses [74]. Training-based approaches instead modify the LLMs’ learning process to optimize for semantic and stylistic diversity [75, 76]. For example, researchers have fine-tuned models on datasets with multiple valid responses per prompt [75] or introduced diversity-aware weighting in preference optimization to reward rare, high-quality generations [76]. Although promising, evidence of homogenization in sociolinguistic contexts calls for closer evaluation of whether these strategies foster genuine, context-grounded diversity or merely superficial variation.

Vitally, the concerns regarding the linguistic diversity of LLMs extend beyond representation alone, as these models actively shape how language is used and evolves. As they become embedded in everyday writing (e.g., composing emails), while supporting more efficient and polished communication [77], they also tends to promote uniform styles that mask authentic voices and reduce variation in tone, culture, and identity [47, 78] even for those merely engaging with AI-generated text [79].

At first glance, the homogenization of language, along with the loss of lexical cues linking writing to individual identity, may appear beneficial for protecting privacy as it reduces risks of misuse such as surveillance or discrimination [80]. But it is equally

important to recognize that this process also erases valuable linguistic markers long used across sociolinguistics, psychology, and mental health research. For example, people at risk of Alzheimer’s disease exhibit early linguistic signs such as telegraphic speech (simplified phrases lacking grammar and function words, often missing determiners like “the” or “a,” auxiliaries like “is” or “are,” and even entire subjects), as well as repetitiveness and misspellings, patterns reflecting a decline in grammatical structure and language complexity [81]. If such distinctive markers are standardized by LLM-assisted polishing, critical early indicators for diagnosis and intervention may be lost. Moreover, the writing styles that LLMs incorporate are not neutral but a source of bias themselves, reproducing dominant expressive norms while eroding minority and underrepresented voices [47, 82].

Perspectival Diversity

Individuals vary not only in how they express themselves but also in their perspectives, beliefs, and values [83]. In language, these differences are reflected in patterns of how people take stances [84] or frame issues [85], widely studied in computational analyses of text [86]. For instance, analyses of U.S. Senate speeches reveal that Democrats and Republicans frame the issue of abortion through distinct moral dimensions, Democrats emphasizing fairness and rights, while Republicans focus more on purity and sexual morality, mirroring broader ideological divides in moral reasoning [85].

With the rise of LLMs, this domain of variation has become especially salient, as these models are increasingly deployed in contexts involving perspective gathering and open-ended writing tasks [87]. Recent work shows they can simulate public opinion, for example, on global warming, by generating synthetic survey responses conditioned on demographics and personal concern, capturing belief patterns that mirror human data [88]. Hence, a pressing question arises: do they preserve the pluralism of human perspectives, or do they default to a narrow, socially normative median?

Multiple studies have shown that LLMs tend to reflect characteristic of western, educated, industrialized, rich, and democratic societies (WEIRD) in their perspectives

[89–91]. Using the World Values Survey [92] and the Moral Foundations Questionnaire-2 (MFQ-2; 53) researchers tested GPT-3.5 across 1000 runs with a temperature of 1.0 to maximize output variance and compared the results with human responses from diverse cultural backgrounds, finding that LLM outputs not only exhibited substantially less variance than human data but were also more closely aligned with response patterns characteristic of WEIRD societies, showing limited variability and weak representation of non-WEIRD perspectives [89, 90]. While models can simulate diverse viewpoints when explicitly prompted, such as by adjusting generation parameters, incorporating identity-coded instructions [93], or translating prompts to a target language [94], these interventions have often fallen short of aligning outputs with the actual distribution of perspectives within the referenced groups [91, 95], capture merely the socially “correct” [96] or the mean of the distribution [89], and in some cases, even reproduce out-group stereotypes or misrepresentations of the specified populations [49].

Similar to efforts aimed at enhancing lexical diversity in LLM generations, recent research has also sought to increase perspective diversity through both prompting and fine-tuning approaches [97, 98]. For instance, researchers have introduced a debate-based framework in which multiple models interact to produce arguments that are broader in scope and more balanced across viewpoints [98]. Further through fine-tuning, reinforcement learning frameworks have been adapted to mitigate “preference collapse” by adjusting reward functions to account for underrepresented or minority preferences, balancing performance with diversity [99]. Yet it remains unclear whether these strategies truly embody human-aligned pluralism of perspectives and contextual depth, highlighting the need for more systematic evaluation.

This power to simulate and frame perspectives does not remain contained within the model; as LLMs become integrated into daily lives, they begin to influence how individuals perceive and frame the world, narrowing perspectives that are naturally rooted in lived experiences [100], environmental, and social contexts [101, 102]. Studies show that

when people remember events together, they tend to align their memories with others in the group, reinforcing shared details while forgetting those left unmentioned [103]. LLMs, however, can amplify this dynamic on a global scale: by exposing millions of users to the same suggestions and perspectives, they foster similar patterns of recall and association, promoting convergence in what people collectively remember and express.

One setting where this influence is particularly consequential is co-writing with LLM-based assistants, as people increasingly rely on them even for open-ended survey responses about their beliefs and behaviors [104], and this, in turn, shapes how users frame and articulate their own perspectives [51, 105]. This effect is evident in studies showing that participants who co-wrote with opinionated language models, engineered to frame social media positively or negatively, tended to mirror the model’s stance in their writing and even shifted their own opinions in subsequent attitude surveys [51]. These findings raise broader concerns about persuasive influence, as even subtle interaction with LLMs can lead users to adopt the model’s framing without awareness [106].

Reasoning Diversity

Beyond language and perspective, another consequential form of diversity at stake is reasoning diversity, the varied ways people reason, solve problems, and generate new ideas, which arises from cultural and individual differences. For example, Native American (Menominee) children often group animals based on ecological relationships, such as shared habitat or interdependence, whereas European-American children tend to rely on taxonomic categories, reflecting broader differences in how people make sense of the world [107, 108]. Across disciplines, variation in reasoning is recognized as a key driver of collective strength [1, 102]. Groups composed of individuals who reason differently, using distinct heuristics and problem representations, consistently outperform groups made up of the highest-ability individuals [5], and innovation and intellectual breakthroughs often emerge from the recombination of ideas across domains and engagement with unfamiliar concepts [109, 110].

As LLMs are increasingly deployed in reasoning-intensive settings [111, 112], preserving reasoning diversity becomes crucial. If these systems reflect only a narrow slice of human thought, they risk reinforcing dominant cognitive styles while marginalizing others. While LLMs offer high performance and access to expertise, collective convergence on even optimal algorithms can reduce decision-making quality in complex systems [113]. This concern is not abstract: a global “weirdization” may already be underway, as once-local ways of conceptualizing time, space, and causality give way to homogenized, Western-aligned models [11]. Trained on massive, biased corpora and built upon similar datasets and architectures, LLMs may further accelerate this shift, risking the homogenization of reasoning processes and decision outcomes across contexts [114, 115].

Various studies have compared the reasoning and cognitive abilities of LLMs to those of humans across tasks traditionally used to assess human cognition. While these models often align with human reasoning in their outcomes, they have also revealed important discrepancies [16, 116, 117]. LLMs tend to produce reasoning patterns that cluster around central tendencies, lacking the natural variability that characterizes human thought. For example, in tasks inspired by the wisdom of the crowds phenomenon, where the diversity of individual judgments allows groups to collectively approximate correct answers, LLMs instead converge on uniform, “idealized” responses, missing the variance that makes human reasoning adaptive and collectively effective [16].

This mismatch may stem from the very objectives used to train and evaluate LLMs, which emphasize measurable performance gains or compliance with verifiable behavioral metrics such as accuracy, informativeness, helpfulness, harmlessness, or formatting consistency [39, 69], prioritizes improving reasoning performance over assessing variation in reasoning approaches. Even cognitively inspired methods, such as analogical reasoning in LLMs [118], tree-of-thought prompting [119], and memory-based inference [120], are typically optimized for correctness and general utility. The widespread success of chain-of-thought prompting [20] has further reinforced this narrow focus, driving

homogenization across multimodal reasoning [121, 122]. This emphasis on performance, and the overreliance on techniques that maximize it, raises concerns about the adaptability, generalizability, and cognitive diversity of a uniform reasoning paradigm [22]. This overreliance and homogenization of reasoning already show documented downsides [21]. For instance, in a vehicle-classification task where most items followed a simple rule but a few violated it, chain-of-thought prompting made GPT-4o four times slower to learn correct labels, as step-by-step reasoning overgeneralized from regular patterns and overlooked exceptions and contextual cues [123].

This imperative to maintain diversity extends beyond how LLMs reflect human reasoning to how they influence it. For instance, in controlled experiments designed to assess creative ideation, participants who received assistance from LLMs (i.e., ChatGPT) generated a greater number of more detailed and elaborated ideas, particularly benefiting those who were less experienced or less creative writers. However, their outputs were also judged to be more semantically similar across participants [124, 125]. This convergence becomes even more concerning when considering how users interact with these systems: rather than actively steering generation, users often defer to model-suggested continuations, selecting options that seem “good enough” instead of crafting their own, which gradually shifts agency from the user to the model [126]. This tendency is particularly alarming for younger users, as LLMs’ negative effects on creativity are most pronounced when used early in the ideation process [127]. Moreover, the consequences extend beyond creative output to the underlying cognitive processes: neurocognitive evidence shows that LLM-assisted writing elicits the weakest overall neural coupling, with reduced engagement of alpha and beta networks, lower memory recall, and diminished ownership of written work compared not only to independent writing but also to writing supported by search engines [128].

Concluding remarks

Technological advancements have long reshaped human life, but LLMs stand out for the unprecedented scale and subtlety of their influence. Trained on vast and often biased

corpora, biases further amplified through iterative retraining, LLMs are now deeply embedded in human language, intertwined with identity and cognition, and as billions interact with them on a daily basis, they risk homogenizing human expression and thought. Empirical evidence throughout this paper underscores this effect: reduced stylistic and lexical diversity in generated texts, subtle recalibration of user attitudes and framing in AI-mediated communication, and diminished diversity in ideation.

The concern is not just that LLMs shape how people write or speak, but that they subtly redefine what counts as credible speech, correct perspective, or even good reasoning by making certain framings more salient. As sociologist George Ritzer’s “McDonaldization” theory suggests [129], processes favoring efficiency, predictability, and control can suppress contextual richness. LLMs mirror this logic in cognition, offering fluency and consistency while displacing situated, idiosyncratic forms of thought.

The centralized control over the very algorithms and datasets driving this homogenization is also acutely political, amplified by the dominance of a few platforms owned by multibillion-dollar corporations with significant political influence [27]. In a time of rising global populism, this concentrated power enables top-down homogenization, where perspectives can be subtly engineered for concentrated benefit and power. The documented censorship, such as the Chinese Qwen model’s refusal to answer politically sensitive questions [130], demonstrates this systemic risk. This algorithmic and market-driven erasure of diverse thought poses a modern danger akin to the linguistic control of Newspeak in Orwell’s Nineteen Eighty-Four [131].

A growing number of strategies aim to counteract this homogenization, including personalized modeling approaches [132, 133], embodied reasoning frameworks that promote situated and context-sensitive inference [134], and diversified prompting or multi-agent debate systems designed to broaden reasoning [98, 135]. Yet evidence from sociolinguistics and psychology continues to document persistent homogenization, suggesting that these solutions must be applied more comprehensively and systematically before their

effectiveness can be meaningfully evaluated. More broadly, both prompting- and training-based diversification methods remain constrained by underlying pretraining representations, potentially limiting their ability to induce deeper variation [36, 38, 40], and may also introduce trade-offs, as variation too far from pretraining distributions can increase the likelihood of hallucination [136].

In summary, while LLMs offer significant advancements and conveniences, their broad adoption without critical evaluation risks fundamentally altering the diverse cognitive landscapes that enrich human interaction and drive innovation. This homogenization constitutes a profound challenge to the ideal of collective wisdom, the fundamental principle, articulated by Mill, that true knowledge is acquired only by “hearing what can be said about it by persons of every variety of opinion.” We do not advocate for diversity in language, perspective, or reasoning without recognizing the coordination costs it can entail. Rather, we call for a deeper understanding of how LLMs affect the diversity of language and thought, as well as the signals embedded in this diversity, and for these insights to inform their design. Moving forward, preserving and enhancing meaningful human diversity should be a central criterion in the development and evaluation of LLMs. Only through deliberate attention to this pluralism can we harness the full potential of language technologies without sacrificing the very diversity that defines human society. Nevertheless, there is still much to learn about the LLM-driven homogenization of language and thought, and how best to address it. Some key directions for deepening this understanding are outlined in the section below.

Outstanding Questions

- Will current alignment methods, such as supervised fine-tuning and RLHF, ever be sufficient to reproduce the full diversity of human cognition, or are more foundational changes in model architecture, objectives, and training data required? These approaches have increased model steerability and surface-level variation, but it remains unclear whether they can capture deeper context-sensitive and culturally grounded forms of diversity found in human thought.
- Even if LLMs produce more diverse outputs, how can we ensure that this diversity is meaningful and grounded in actual human experience rather than being superficial or artificially constructed? Future research should identify and promote metrics and frameworks that distinguish synthetic variation from diversity reflecting authentic sociocultural, emotional, and cognitive nuance.
- What are the long-term cognitive effects of sustained reliance on LLMs for ideation, writing, and reasoning? While short-term effects such as reduced stylistic variation and creative ownership have been observed, we lack longitudinal studies that track changes in abstraction, memory retention, and reasoning strategies over time and examine whether changes are irreversible.
- Can users be equipped with strategies to counteract the homogenizing effects of LLMs on expression and thought? Research is needed to develop and evaluate behavioral or interface-level interventions, such as delaying LLM use during ideation, or exposing users to model-induced changes, that help preserve agency and individuality in interaction.
- What taxonomies or repertoires of intervention can future research establish to help mitigate LLM-driven homogenization at scale? A systematic understanding of the possible safeguards, behavioral, architectural, or institutional, is needed to guide users, developers, and platforms toward practices that promote cognitive and linguistic pluralism.

Glossary

Chain-of-Thought (CoT) prompting: A prompting strategy that encourages models to show step-by-step reasoning before giving an answer, improving structure and accuracy but sometimes reducing intuitive or creative responses.

Epistemic: Relating to knowledge, understanding, or how beliefs are formed and justified. Used to describe cognitive or informational aspects of learning and reasoning.

Epistemic Collapse: A situation where diversity in knowledge, reasoning, or interpretation is lost, leading to uniform ways of thinking and a narrower collective understanding.

Essentialized Representations of Identity: Simplified portrayals of social or cultural groups that treat identity as fixed and homogeneous, often ignoring variation and individual nuance.

Natural Language Processing (NLP): A field of AI focused on enabling computers to understand, interpret, and produce human language across tasks like translation, summarization, and question answering.

Prompt: A text input or instruction given to a language model that guides how it generates a response, effectively shaping the model's behavior and output.

Reinforcement Learning from Human Feedback (RLHF): A training method where human evaluators rate model outputs, guiding the model to produce more preferred or context-appropriate responses.

Supervised Fine-Tuning: A supervised training process that adapts a pre-trained language model to a specific task or domain by continuing its training on a smaller, labeled dataset, teaching the model new knowledge or behaviors.

Temperature Scaling: A parameter that controls how deterministic or random a model's outputs are: lower temperatures yield precise, predictable text, while higher ones increase variety.

Acknowledgments

This research was supported by the Air Force Office of Scientific Research A9550-23-1-046. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of AFOSR, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

- [1] Scott Page. *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*. Princeton University Press, 2008.
- [2] Nicholas Evans and Stephen C Levinson. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, 32(5):429–448, 2009.
- [3] Asifa Majid, Melissa Bowerman, Sotaro Kita, Daniel BM Haun, and Stephen C Levinson. Can language restructure cognition? the case for space. *Trends in cognitive sciences*, 8(3):108–114, 2004.
- [4] Penelope Eckert. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41(1):87–100, 2012.
- [5] Lu Hong and Scott E Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.
- [6] John Stuart Mill. On liberty (1859), 2001.
- [7] Miriam Solomon. Groupthink versus the wisdom of crowds: The social epistemology of deliberation and dissent. *The Southern journal of philosophy*, 44(S1):28–42, 2006.
- [8] William Labov. The social setting of linguistic change. *Sociolinguistic patterns*, pages 260–325, 1972.
- [9] David Harmon and Jonathan Loh. The index of linguistic diversity: A new quantitative measure of trends in the status of the world’s languages. *Language Documentation & conservation*, 4, 2010.
- [10] Abdullah Alsaleh. The impact of technological advancement on culture and society. *Scientific Reports*, 14(1):32140, 2024.

- [11] Kensi Cooperrider. What happens to cognitive diversity when everyone is more weird? Aeon, 2019. Available at: <https://aeon.co/ideas/what-happens-to-cognitive-diversity-when-everyone-is-more-weird> [Accessed 30 Oct 2025].
- [12] Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, et al. Which economic tasks are performed with ai? evidence from millions of claude conversations. *arXiv preprint arXiv:2503.04761*, 2025.
- [13] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, 12:39–57, 2024.
- [14] Michal Kosinski. Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, 121(45):e2405460121, 2024.
- [15] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22, 2023.
- [16] Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*, pages 337–371. PMLR, 2023.
- [17] Danica Dillion, Niket Tandon, Yuling Gu, and Kurt Gray. Can ai language models replace human participants? *Trends in Cognitive Sciences*, 27(7):597–600, 2023.
- [18] Gaofeng Wang, Yetong Gan, and Haodong Yang. The inverted u-shaped relationship

- between knowledge diversity of researchers and societal impact. *Scientific Reports*, 12(1):18585, 2022.
- [19] Zhivar Sourati, Farzan Karimi-Malekabadi, Meltem Ozcan, Colin McDaniel, Alireza Ziabari, Jackson Trager, Ala Tak, Meng Chen, Fred Morstatter, and Morteza Dehghani. The shrinking landscape of linguistic diversity in the age of large language models. *arXiv preprint arXiv:2502.11266*, 2025.
- [20] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- [21] Yang Sui, Yu-Neng Chuang, Guanchu Wang, Jiamu Zhang, Tianyi Zhang, Jiayi Yuan, Hongyi Liu, Andrew Wen, Hanjie Chen, Xia Hu, et al. Stop overthinking: A survey on efficient reasoning for large language models. *arXiv preprint arXiv:2503.16419*, 2025.
- [22] Alireza S Ziabari, Nona Ghazizadeh, Zhivar Sourati, Farzan Karimi-Malekabadi, Payam Piray, and Morteza Dehghani. Reasoning on a spectrum: Aligning llms to system 1 and system 2 thinking. *arXiv preprint arXiv:2502.12470*, 2025.
- [23] Matthew Fisher, Mariel K Goddu, and Frank C Keil. Searching for explanations: How the internet inflates estimates of internal knowledge. *Journal of experimental psychology: General*, 144(3):674, 2015.
- [24] Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998.
- [25] Alexander Bick, Adam Blandin, and David J Deming. The rapid adoption of generative ai. Technical report, National Bureau of Economic Research, 2024.

- [26] Aaron Chatterji, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. How people use chatgpt. Technical report, National Bureau of Economic Research, 2025.
- [27] Kyle Wiggers. Chatgpt isn't the only chatbot that's gaining users.
[https://techcrunch.com/2025/04/01/
chatgpt-isnt-the-only-chatbot-thats-gaining-users/](https://techcrunch.com/2025/04/01/chatgpt-isnt-the-only-chatbot-thats-gaining-users/), April 2025.
TechCrunch, accessed October 20, 2025.
- [28] Mark Steyvers, Heliodoro Tejeda, Aakriti Kumar, Catarina Belem, Sheer Karny, Xinyue Hu, Lukas W Mayer, and Padhraic Smyth. What large language models know and what people think they know. *Nature Machine Intelligence*, 7(2):221–231, 2025.
- [29] James WA Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8(7):1285–1295, 2024.
- [30] Neeltje J Boogert, Joah R Madden, Julie Morand-Ferron, and Alex Thornton. Measuring and understanding individual differences in cognition, 2018.
- [31] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [32] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- [33] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou,

- Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023.
- [34] Tom M Mitchell. Machine learning and data mining. *Communications of the ACM*, 42(11):30–36, 1999.
- [35] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.
- [36] Bill Yuchen Lin, Abhilasha Ravichander, Ximing Lu, Nouha Dziri, Melanie Sclar, Khyathi Raghavi Chandu, Chandra Bhagavatula, and Yejin Choi. The unlocking spell on base llms: Rethinking alignment via in-context learning. *CoRR*, 2023.
- [37] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*, 2022.
- [38] Sreyan Ghosh, Chandra Kiran Reddy Evuru, Sonal Kumar, Deepali Aneja, Zeyu Jin, Ramani Duraiswami, Dinesh Manocha, et al. A closer look at the limitations of instruction tuning. In *International Conference on Machine Learning*, pages 15559–15589. PMLR, 2024.
- [39] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *CoRR*, 2022.
- [40] Yotam Wolf, Noam Wies, Oshri Avnery, Yoav Levine, and Amnon Shashua. Fundamental limitations of alignment in large language models. In *Proceedings of the 41st International Conference on Machine Learning*, pages 53079–53112, 2024.

- [41] Lisa Schut, Yarin Gal, and Sebastian Farquhar. Do multilingual llms think in english? In *ICLR 2025 Workshop on Building Trust in Language Models and Applications*, 2025.
- [42] Maarten Buyl, Alexander Rogiers, Sander Noels, Iris Dominguez-Catena, Edith Heiter, Raphaël Romero, Iman Johary, Alexandru Cristian Mara, Jeffrey Lijffjt, and Tijl De Bie. Large language models reflect the ideology of their creators. *CoRR*, 2024.
- [43] Ze Wang, Zekun Wu, Jeremy Zhang, Xin Guan, Navya Jain, Skylar Lu, Saloni Gupta, and Adriano Koshiyama. Bias amplification: Large language models as increasingly biased media. *arXiv preprint arXiv:2410.15234*, 2024.
- [44] Melissa Hall, Laurens van der Maaten, Laura Gustafson, Maxwell Jones, and Aaron Adcock. A systematic study of bias amplification. *arXiv preprint arXiv:2201.11706*, 2022.
- [45] R Thomas McCoy, Paul Smolensky, Tal Linzen, Jianfeng Gao, and Asli Celikyilmaz. How much do language models copy from their training data? evaluating linguistic novelty in text generation using raven. *Transactions of the Association for Computational Linguistics*, 11:652–670, 2023.
- [46] Kanishka Misra and Kyle Mahowald. Language models learn rare phenomena from less rare phenomena: The case of the missing AANNs. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 913–929, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [47] AJ Alvero, Jinsook Lee, Alejandra Regla-Vargas, René F Kizilcec, Thorsten Joachims, and Anthony Lising Antonio. Large language models, social demography, and hegemony: comparing authorship in human and synthetic text. *Journal of Big Data*, 11(1):138, 2024.

- [48] Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. The political ideology of conversational ai: Converging evidence on chatgpt’s pro-environmental, left-libertarian orientation. *arXiv preprint arXiv:2301.01768*, 2023.
- [49] Angelina Wang, Jamie Morgenstern, and John P Dickerson. Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, pages 1–12, 2025.
- [50] Sierra Wyllie, Ilia Shumailov, and Nicolas Papernot. Fairness feedback loops: training on synthetic data amplifies bias. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2113–2147, 2024.
- [51] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. Co-writing with opinionated language models affects users’ views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–15, 2023.
- [52] Johanna Nichols. *Linguistic diversity in space and time*. University of Chicago Press, 1992.
- [53] Mohammad Atari, Jonathan Haidt, Jesse Graham, Sena Koleva, Sean T Stevens, and Morteza Dehghani. Morality beyond the weird: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, 125(5):1157, 2023.
- [54] Jonas Nölle, Riccardo Fusaroli, Gregory J Mills, and Kristian Tylén. Language as shaped by the environment: linguistic construal in a collaborative spatial task. *Palgrave Communications*, 6(1):1–10, 2020.
- [55] Ryan Boyd, Steven Wilson, James Pennebaker, Michal Kosinski, David Stillwell, and Rada Mihalcea. Values in words: Using language to evaluate and understand personal values. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, pages 31–40, 2015.

- [56] David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160, 2014.
- [57] Jacob B Hirsh and Jordan B Peterson. Personality and language use in self-narratives. *Journal of research in personality*, 43(3):524–527, 2009.
- [58] James W Pennebaker. The secret life of pronouns. *New Scientist*, 211(2828):42–45, 2011.
- [59] Reihane Boghrati, Joe Hoover, Kate M Johnson, Justin Garten, and Morteza Dehghani. Conversation level syntax similarity metric. *Behavior research methods*, 50(3):1055–1073, 2018.
- [60] Liangda Li, Ke Zhou, Gui-Rong Xue, Hongyuan Zha, and Yong Yu. Enhancing diversity, coverage and balance for summarization through structure learning. In *Proceedings of the 18th international conference on World wide web*, pages 71–80, 2009.
- [61] Kevin Gimpel, Dhruv Batra, Chris Dyer, and Gregory Shakhnarovich. A systematic exploration of diversity in machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1100–1111, 2013.
- [62] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In Kevin Knight, Ani Nenkova, and Owen Rambow, editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California, June 2016. Association for Computational Linguistics.
- [63] Dong Nguyen, A Seza Doğruöz, Carolyn P Rosé, and Franciska De Jong. Computational sociolinguistics: A survey. *Computational linguistics*, 42(3):537–593, 2016.

- [64] Alireza Salkhordeh Ziabari, Ali Omrani, Parsa Hejabi, Preni Golazizian, Brendan Kennedy, Payam Piray, and Morteza Dehghani. Reinforced multiple instance selection for speaker attribute prediction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3307–3321, 2024.
- [65] Yanzhu Guo, Guokan Shang, and Chloé Clavel. Benchmarking linguistic diversity of large language models. *CoRR*, 2024.
- [66] Jinsook Lee, AJ Alvero, Thorsten Joachims, and Rene F Kizilcec. Poor alignment and steerability of large language models: Evidence using 30,000 college admissions essays. In *First Workshop on Social Simulation with LLMs*, 2025.
- [67] Gonzalo Martínez, José Alberto Hernández, Javier Conde, Pedro Reviriego, and Elena Merino-Gómez. Beware of words: Evaluating the lexical diversity of conversational llms using chatgpt as case study. *ACM Transactions on Intelligent Systems and Technology*, 2024.
- [68] Yanzhu Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. The curious decline of linguistic diversity: Training language models on synthetic text. In *NAACL 2024 Findings-Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2024.
- [69] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.
- [70] Tianzhe Chu, Yuexiang Zhai, Jihan Yang, Shengbang Tong, Saining Xie, Dale Schuurmans, Quoc V. Le, Sergey Levine, and Yi Ma. Sft memorizes, rl generalizes: A comparative study of foundation model post-training. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.

- [71] Rowan Kirk, Ishita Mediratta, Christos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and Roberta Raileanu. Understanding the effects of rlhf on llm generalisation and diversity. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, 2024.
- [72] Hugh Zhang, Daniel Duckworth, Daphne Ippolito, and Arvind Neelakantan. Trading off diversity and quality in natural language generation. In *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, pages 25–33, 2021.
- [73] KuanChao Chu, Yi-Pei Chen, and Hideki Nakayama. Exploring and controlling diversity in llm-agent conversation. *arXiv preprint arXiv:2412.21102*, 2024.
- [74] Gauri Kambhatla, Chantal Shaib, and Venkata Govindarajan. Measuring diversity of synthetic prompts and data generated with fine-grained persona prompting. *arXiv preprint arXiv:2505.17390*, 2025.
- [75] Long Mai and Julie Carson-Berndsen. Improving linguistic diversity of large language models with possibility exploration fine-tuning. *arXiv preprint arXiv:2412.03343*, 2024.
- [76] John Joon Young Chung, Vishakh Padmakumar, Melissa Roemmele, Yuqian Sun, and Max Kreminski. Modifying large language model post-training for diverse creative writing. In *Proceedings of the Conference on Language Modeling (COLM 2025)*, 2025.
- [77] Jess Hohenstein, Rene F Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F Jung. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13(1):5487, 2023.
- [78] Nicole R Holliday and Paul E Reed. Gender and racial bias issues in a commercial “tone of voice” analysis system. *PloS one*, 20(2):e0314470, 2025.

- [79] Catalina L Toma. Towards conceptual convergence: An examination of interpersonal adaptation. *Communication Quarterly*, 62(2):155–178, 2014.
- [80] Zeynep Tufekci. Engineering the public: Big data, surveillance and computational politics. *First Monday*, 2014.
- [81] Elif Eyigoz, Sachin Mathur, Mar Santamaria, Guillermo Cecchi, and Melissa Naylor. Linguistic markers predict onset of alzheimer’s disease. *EClinicalMedicine*, 28, 2020.
- [82] Tiffany J Huang. Translating authentic selves into authentic applications: Private college consulting and selective college admissions. *Sociology of Education*, 97(2):174–192, 2024.
- [83] Eric Y Aglozo, Kathryn A Johnson, Brendan Case, R Noah Padgett, Byron R Johnson, and Tyler J VanderWeele. A cross-national analysis of demographic variation in belief in god, gods, or spiritual forces in 22 countries. *Scientific reports*, 15(1):13302, 2025.
- [84] Dilek Küçük and Fazli Can. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37, 2020.
- [85] Eyal Sagi and Morteza Dehghani. Measuring moral rhetoric in text. *Social science computer review*, 32(2):132–144, 2014.
- [86] Brendan Kennedy, Ashwini Ashokkumar, Ryan L Boyd, and Morteza Dehghani. Text analysis for psychology. *Handbook of language analysis in psychology*, page 1, 2022.
- [87] Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.
- [88] Sanguk Lee, Tai-Quan Peng, Matthew H Goldberg, Seth A Rosenthal, John E Kotcher, Edward W Maibach, and Anthony Leiserowitz. Can large language models

- estimate public opinion about global warming? an empirical assessment of algorithmic fidelity and bias. *PLoS Climate*, 3(8):e0000429, 2024.
- [89] Suhaib Abdurahman, Mohammad Atari, Farzan Karimi-Malekabadi, Mona J Xue, Jackson Trager, Peter S Park, Prene Golazizian, Ali Omrani, and Morteza Dehghani. Perils and opportunities in using large language models in psychological research. *PNAS nexus*, 3(7):pgae245, 2024.
- [90] Mohammad Atari, Mona J Xue, Peter S Park, Damián Blasi, and Joseph Henrich. Which humans? *PsyArXiv*, 2023. Preprint.
- [91] Esin Durmus, Karina Nguyen, Thomas Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, Liane Lovitt, Sam McCandlish, Orowa Sikder, Alex Tamkin, Janel Thamkul, Jared Kaplan, Jack Clark, and Deep Ganguli. Towards measuring the representation of subjective global opinions in language models. In *Proceedings of the First Conference on Language Modeling*, 2024.
- [92] Christian Haerpfer, Ronald Inglehart, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Juan Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, Bi Puranen, and Others. World values survey: Round seven – country-pooled datafile (version 5.0). JD Systems Institute & WVSA Secretariat, Madrid and Vienna, 2020. Data set. Available at: <https://doi.org/10.14281/18241.1>.
- [93] Ming Li, Juhai Chen, Lichang Chen, and Tianyi Zhou. Can llms speak for diverse people? tuning llms via debate to generate controllable controversial statements. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 16160–16176, 2024.
- [94] Qihan Wang, Shidong Pan, Tal Linzen, and Emily Black. Multilingual prompting for improving llm generation diversity. *arXiv preprint arXiv:2505.15229*, 2025.

- [95] Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. Whose opinions do language models reflect? In *International Conference on Machine Learning*, pages 29971–30004. PMLR, 2023.
- [96] Peter S Park, Philipp Schoenegger, and Chongyang Zhu. Diminished diversity-of-thought in a standard large language model. *Behavior Research Methods*, 56(6):5754–5770, 2024.
- [97] Shirley Anugrah Hayati, Minhwa Lee, Dheeraj Rajagopal, and Dongyeop Kang. How far can we extract diverse perspectives from large language models? In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5336–5366, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [98] Zhe Hu, Hou Pong Chan, Jing Li, and Yu Yin. Debate-to-write: A persona-driven multi-agent framework for diverse argument generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4689–4703, 2025.
- [99] Jiancong Xiao, Ziniu Li, Xingyu Xie, Emily J Getzen, Cong Fang, Qi Long, and Weijie J Su. On the algorithmic bias of aligning large language models with rlhf: Preference collapse and matching regularization. *CoRR*, 2024.
- [100] Angel Hsing-Chi Hwang, Q Vera Liao, Su Lin Blodgett, Alexandra Olteanu, and Adam Trischler. ‘it was 80% me, 20% ai’: Seeking authenticity in co-writing with large language models. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–41, 2025.
- [101] Todd I Lubart. Models of the creative process: Past, present and future. *Creativity research journal*, 13(3-4):295–308, 2001.
- [102] Bernard A Nijstad, Wolfgang Stroebe, and Hein FM Lodewijkx. Cognitive

stimulation and interference in groups: Exposure effects in an idea generation task.

Journal of experimental social psychology, 38(6):535–544, 2002.

- [103] Alin Coman and William Hirst. Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General*, 144(4):717, 2015.
- [104] Simone Zhang, Janet Xu, and AJ Alvero. Generative ai meets open-ended survey responses: Research participant use of ai and homogenization. *Sociological Methods & Research*, page 00491241251327130, 2025.
- [105] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. Ai suggestions homogenize writing toward western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2025.
- [106] Lisa Rashotte. Social influence. *The Blackwell encyclopedia of sociology*, 2007.
- [107] Douglas L Medin and Scott Atran. The native mind: biological categorization and reasoning in development and across cultures. *Psychological review*, 111(4):960, 2004.
- [108] Douglas L Medin, Sara J Unsworth, and Lawrence Hirschfeld. Culture, categorization, and reasoning. *Handbook of cultural psychology*, pages 615–644, 2007.
- [109] Michael Muthukrishna and Joseph Henrich. Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690):20150192, 2016.
- [110] Eamon Duede, Misha Teplitskiy, Karim Lakhani, and James Evans. Being together in place as a catalyst for scientific advance. *Research Policy*, 53(2):104911, 2024.
- [111] Karthik Valmeekam, Matthew Marquez, Sarath Sreedharan, and Subbarao Kambhampati. On the planning abilities of large language models-a critical

investigation. *Advances in Neural Information Processing Systems*, 36:75993–76005, 2023.

- [112] Yuyan Chen, Yueze Li, Songzhou Yan, Sijia Liu, Jiaqing Liang, and Yanghua Xiao. Do large language models have problem-solving capability under incomplete information scenarios? In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2225–2238, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [113] Jon Kleinberg and Manish Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118, 2021.
- [114] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.
- [115] Rishi Bommasani, Kathleen A Creel, Ananya Kumar, Dan Jurafsky, and Percy S Liang. Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Advances in Neural Information Processing Systems*, 35:3663–3678, 2022.
- [116] Marcel Binz and Eric Schulz. Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120, 2023.
- [117] Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. Thinking fast and slow in large language models. *arXiv preprint arXiv:2212.05206*, 2022.
- [118] Michihiro Yasunaga, Xinyun Chen, Yujia Li, Panupong Pasupat, Jure Leskovec, Percy Liang, Ed H. Chi, and Denny Zhou. Large language models as analogical

reasoners. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*, 2024.

- [119] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822, 2023.
- [120] Nirmalie Wiratunga, Ramitha Abeyratne, Lasal Jayawardena, Kyle Martin, Stewart Massie, Ikechukwu Nkisi-Orji, Ruvan Weerasinghe, Anne Liret, and Bruno Fleisch. Cbr-rag: case-based reasoning for retrieval augmented generation in llms for legal question answering. In *International Conference on Case-Based Reasoning*, pages 445–460. Springer, 2024.
- [121] Hao Shao, Shengju Qian, Han Xiao, Guanglu Song, Zhuofan Zong, Letian Wang, Yu Liu, and Hongsheng Li. Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning. *Advances in Neural Information Processing Systems*, 37:8612–8642, 2024.
- [122] Songhao Han, Wei Huang, Hairong Shi, Le Zhuo, Xiu Su, Shifeng Zhang, Xu Zhou, Xiaojuan Qi, Yue Liao, and Si Liu. Videospresso: A large-scale chain-of-thought dataset for fine-grained video reasoning via core frame selection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26181–26191, 2025.
- [123] Ryan Liu, Jiayi Geng, Addison J. Wu, Ilia Sucholutsky, Tania Lombrozo, and Thomas L. Griffiths. Mind your step (by step): Chain-of-thought can reduce performance on tasks where thinking makes humans worse. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.
- [124] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. Homogenization

- effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*, pages 413–425, 2024.
- [125] Anil R Doshi and Oliver P Hauser. Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science advances*, 10(28):eadn5290, 2024.
- [126] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023.
- [127] Peinuan Qin, Chi-Lan Yang, Jingshu Li, Jing Wen, and Yi-Chieh Lee. Timing matters: How using llms at different timings influences writers’ perceptions and ideation outcomes in ai-assisted ideation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2025.
- [128] Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 2025.
- [129] George Ritzer. The mcdonaldization of society. In *In the Mind’s Eye*, pages 143–152. Routledge, 2021.
- [130] Leonard Lin. An analysis of chinese llm censorship and bias with qwen 2 instruct. Hugging Face Blog, 2024. Available at: <https://huggingface.co/blog/leonardlin/chinese-llm-censorship-analysis> [Accessed 30 Oct 2025].
- [131] George Orwell. *Nineteen Eighty-Four*. Secker & Warburg, London, 1949.

- [132] Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, et al. Position: a roadmap to pluralistic alignment. In *Proceedings of the 41st International Conference on Machine Learning*, pages 46280–46302, 2024.
- [133] Lora Aroyo, Alex Taylor, Mark Diaz, Christopher Homan, Alicia Parrish, Gregory Serapio-García, Vinodkumar Prabhakaran, and Ding Wang. Dices dataset: Diversity in conversational ai evaluation for safety. *Advances in Neural Information Processing Systems*, 36:53330–53342, 2023.
- [134] Gul Deniz Salali, Mirco Musolesi, and Hugo Spiers. Flourishing cultural diversity in ai. *OSF Preprints*, 2024. Preprint.
- [135] Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. In *Forty-first International Conference on Machine Learning*, 2023.
- [136] Zorik Gekhman, Gal Yona, Roee Aharoni, Matan Eyal, Amir Feder, Roi Reichart, and Jonathan Herzig. Does fine-tuning llms on new knowledge encourage hallucinations? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7765–7784, 2024.