Original research article

Platforms & Society
Volume 3: 1–20

DOI: 10.1177/29768624251408919
journals.sagepub.com/home/pns
Sage

The silicon gaze: A typology of biases and inequality in LLMs through the lens of place

Francisco W. Kerche[1], Matthew Zook[2] and Mark Graham[1]

Abstract

This paper introduces the concept of the silicon gaze to explain how large language models (LLMs) reproduce and amplify long-standing spatial inequalities. Drawing on a 20.3-million-query audit of ChatGPT, we map systematic biases in the model's representations of countries, states, cities, and neighbourhoods. From these empirics, we argue that bias is not a correctable anomaly but an intrinsic feature of generative AI, rooted in historically uneven data ecologies and design choices. Building on a power-aware, relational approach, we develop a five-part typology of bias (availability, pattern, averaging, trope, and proxy) that accounts for the complex ways in which LLMs privilege certain places while rendering others invisible.

Keywords

Generative AI, LLM, bias, inequality, representation, digital geography, ChatGPT, silicon gaze

Analyzing the silicon gaze of AI

Artificial intelligence (AI) systems increasingly shape our understanding of the world. In 2025, over 50% of all adults in the US reported using large language models (LLMs) such as ChatGPT (Rainie, 2025), and worldwide use has expanded both in scope and scale (Zao-Sanders, 2025). Such widespread popularity shapes how users of LLMs comprehend economic, social, political, and spatial facets of the world around them. But, as generative AI models shape understanding, they remain deeply enmeshed in the inequalities and biases that have long characterised the data ingested by their models. While early visions of digital life imagined egalitarian interactions separate from the material world (Steiner, 1993), the reality is that digital representations and interactions continue to replicate – and often amplify – historical stereotypes and patterns of exclusion and marginalisation (Graham and Dittus, 2022). In this sense, the "palimpsests" of place (Graham, 2010) are now written and rewritten by AI algorithms whose foundational biases remain largely invisible to end users by the platforms that own and operate them.

To capture this phenomenon, we introduce the concept of the "silicon gaze" as a critical lens for analyzing how AI systems see and render geographies. Like the "male gaze" in feminist theory that positioned women as passive objects viewed and valued via an exogenous power, that is, male desire (Mulvey, 1975), the silicon gaze is shaped by the positionalities and power asymmetries of its training data, designers, and platform owners. In the case of AI models, two key sources are the developers (predominantly male, white, and Western[1]) and the training data used, which are also similarly skewed (Baack, 2024; Helm et al., 2023). By emphasising the profound subjectivity and positionality inherent in the making of generative AI (GenAI) models, we foreground that the silicon gaze is not neutral but situated in the perspectives and biases of its designers, institutional and ownership frameworks, and the social, economic, and political contexts of its training data (Birhane, 2021; Miceli et al., 2022). Such a perspective compels us to interrogate not only what AI "knows" about places, but whose interests are served by its selective visibility and what voices remain unheard.

[1]Oxford Internet Institute, University of Oxford, Oxford, United Kingdom of Great Britain and Northern Ireland
[2]Department of Geography, University of Kentucky, Lexington, KY, USA

Corresponding author:
Matthew Zook, Department of Geography, University of Kentucky, Lexington, KY, USA.
Email: zook@uky.edu

Our goal in this paper is to audit a key GenAI platform, ChatGPT (Rainie, 2025), to present an initial mapping of the silicon gaze towards geographically referenced queries that mimic ordinary user behaviour. From this data, we ask: What patterns are observable, and how might these connect to the positionalities behind the design and data of AI? Our approach foregrounds different levels of subjectivity, from highly subjective questions, such as "Where are people more beautiful?" or "Where are there better vibes?"; intermediate-level subjective queries, such as "Where are people happier/smarter?" or "Where has better bread?", that have some association with existing metrics; to finally, low-level subjective questions, such as "Which country has the fastest-growing tech sector", which have relatively well-defined metrics. With this range, we explore the discernible biases from generative AI. In doing so, our goal is not simply to document the presence of bias (Pagano et al., 2023); after all, since these models are trained on online human interactions, bias presents itself as a foundational element of large language models (see Bowker, 2005; Esposito, 2017). Rather, we seek to better understand how generative AI perpetuates and disrupts deep-seated inequalities across scales of place and categories of knowledge. By highlighting these moments, we aim to initiate a critical conversation about how to best understand and address bias as a fundamental and inextricable feature of algorithmic AI place-making.

The remainder of the paper unfolds in four steps. We first situate the silicon gaze within critical data-studies scholarship and specify our research questions. We then outline our large-scale audit methodology, explaining how four spatial scales and 311 subjective comparisons were selected to surface bias. The findings section presents the five biases in turn, using maps and case studies to illustrate their operation. We conclude by discussing the implications for AI ethics, geographic scholarship, and policy and by outlining avenues for future work that move beyond simply "fixing" data toward confronting the structural power relations that make bias inevitable and that platform power perpetuates.

Place representation via the silicon gaze

Long before broadband promised a friction-free flow of data, place representation, and most geographic knowledge was produced and circulated through what Bruno Latour called "immutable mobiles" – paper maps, printed gazetteers, atlases, and encyclopaedias (Latour, 1986). These representations could only move at the pace of contemporaneous transportation and were only available where shelves, archives, or libraries existed. With the advent of the printing press, along with specialist skills and practices needed for reproducing spatial facts, the production of immutable mobiles was concentrated in a handful of "centres of calculation" in Europe and North America (Graham and Dittus, 2022). As a result, the ability to document the world was itself highly place-bound, and thus geographic knowledge was shaped by these places. Universities, scientific societies, and surveying offices embedded in imperial capitals became gatekeepers: they fixed territorial boundaries, standardised toponyms, and decided which places merited representation at all, while vast stretches of Africa, Asia, and Latin America remained literal blanks on authoritative maps.

The material frictions of paper and print also meant that updating or contesting those representations was prohibitively costly and essentially inaccessible for most of the world's population. The scarcity of media infrastructures, the spatial fixity of archives, and the elite status of map-making professions ensured that only a narrow slice of humanity had the means to inscribe their worldviews onto supposedly universal reference works (Harley, 1989). As a result, the Global North not only published the overwhelming majority of books, newspapers, and patents; it also monopolised the power to define what counted as legitimate knowledge, increasing the dominance of these places (Pred, 1977). Castells (1996) describes these territories beyond this informational core as "black holes of informational capitalism", zones whose experiences and voices were rendered invisible by the very media through which modernity claimed to know the planet. Cartographic lines drawn in London, Paris, or Washington thus did more than depict space – they legitimised colonial rule, extracted resources, and cemented global hierarchies that persisted long after independence movements erased imperial flags (Harley, 1988).

Uneven data and algorithmic agency

From today's vantage point of digital systems and global networks, it is tempting to frame the "pre-digital" moment as a historical curiosity with little relevance for the current era. However, it is precisely these historically sedimented asymmetries that contemporary digital platforms inherit, while scrollable, zoomable map tiles feel weightless, the epistemic foundations beneath them were laid on paper whose provenance was, and often still is, distinctly Northern and elite (Harley, 1988). Any project that seeks to understand biases in generative AI, therefore, must start by recognising how deeply the politics of visibility were carved into our knowledge ecologies long before a single dataset was scraped or a model was trained.

Geographers have taken various approaches to understand how this intertwining of digital systems and material places manifests in the world, including representation as well as power and control (see Dodge and Kitchin, 2005; Graham, 2005; Thrift and French, 2002). Two key factors are regularly noted by researchers: data and algorithms, which work in concert to shape representations and understandings of place. Focusing first on data, scholars studying earlier digital systems have shown that data (such as statistics on technology use or gender equality) about places is not uniformly distributed. Using the example of sub-Saharan Africa, Graham et al. (2015a) highlight gaps in data either from outright unavailability or lack of certainty in the collection methods used, to propose alternative measures for capturing geographies of information. Even systems that are designed to lower barriers of participation, such as Wikipedia, exhibit stark differences in the level (Graham et al., 2015b) and type of representation (Graham et al., 2013) or even result in the rejection of knowledge from peripheral regions (Ford, 2011). Generative AI systems face similar issues in that the corpora of data available to use are far from uniform both in terms of origin and epistemological framing which shapes their responses (Graham and Dittus, 2022).

The second factor of algorithms builds on this foundation of uneven data and extends and alters its impact by sorting and prioritising to make some things more visible and others less so. As Zook and Graham (2007) argue, this practice is "… absolutely fundamental to how places are presented … [as] … code automatically determines the availability and visibility of electronic information that shapes the representation and consequently the perception of places". Building on this idea, they use the example of differences in the language used in search at the same location to highlight one vector of bias (linguistic) that results in very different representations (Graham and Zook, 2013). Similarly, generative AI models are constructed from complex topologies of word and language associations (derived from the text corpus) that automatically shape the representations they build.

The biased agency of the silicon gaze

In Mulvey's "male gaze" critique of visual media, the bias of presentation is a key idea as the image of women is always in service to a certain view, the heterosexual male. This includes fragmented representations (close-ups on bodies) and reinforcing certain roles of women such as passive objects of desire rather than complex subjects. Others have critiqued and extended this work to explore the role of race (hooks, 2014) and self-objectification in the continuation and recreation of the male gaze in media (Gill, 2007). With the "silicon gaze", we seek to explain how bias of presentation across a whole range of demographics and categories is present within ChatGPT.

Of particular concern is that the uneven availability and representation of data about different places have tangible, real-world consequences, especially when filtered through algorithmic systems. Noble (2018) shows how search engines reinforce societal biases, shaping public perception and producing discriminatory outcomes for individuals based on race, gender, or geography. Chun (2021) extends this concern to the spatial domain, arguing that code and algorithms do not simply mirror existing inequalities – they actively reproduce and intensify them across geographies. Together, these insights reveal how algorithmic systems, built on partial and biased data, deepen spatial injustices by privileging certain places while rendering others invisible or misrepresented.

The implications become even more pronounced with generative AI, as they introduce new concerns distinct from traditional search engines. As Gebru (2020) argues, GenAI systems are systematically disadvantageous to already marginalised people, both in terms of stereotyping and in providing new means for these biases to be practiced and used. While search engines present biases by sorting results from indexed sources, large language models (LLMs) of generative AI create results by predicting the most likely next word, segment of words, or phrase based on their training data. Thus, rather than making existing information available, generative AI creates new information shaped by the language patterns, knowledge structures, and social biases embedded in its sources. This introduces several potential channels for bias. First, disparities in data availability mean that regions with limited digital presence are poorly served by the model. Second, the type of data matters: unstructured text dominates the training corpus, meaning structured or non-narrative knowledge (like spreadsheets or localised records) may not be adequately incorporated. As a result, the model's outputs skew towards the interpretation and natural language incorporation of these indicators, rather than factual values themselves. Finally, the model's predictions are shaped not just by its training data but also by the framing of user prompts and the constraints placed on the interface – factors that can further entrench asymmetries in how different people and places are represented. Together, these dynamics underscore how GenAI not only reflects existing social and spatial inequalities but also reconstitutes them in new forms.

In short, the silicon gaze operates by translating messy, situated worlds into what Scott (1998) would call a legible grammar of tokens, vectors, and dashboards. Because that grammar is built from the most readily available, already-codified data, privileging the Global North's long history of self-documentation: peer-reviewed journals, English-language newswires, social media exchanges, census tables, and corporate knowledge graphs. Centuries of uneven information production materialise inside the model as a centre-periphery vector of bias that lifts Western, white, and affluent spaces while rendering the rest to the epistemic periphery (see also Graham et al., 2015a). In short, what masquerades as neutral rankings or preferences is in fact an automated replay of the same kinds of archival asymmetries that earlier confined the making of maps and encyclopaedias to imperial capitals (Harley, 1988; Latour, 1986).

Auditing AI

In order to document the silicon gaze and begin to understand the factors behind its construction, we conducted an audit (Benjamin, 2019; Brown et al., 2021; Bucher, 2018;

Burrell, 2016; Seaver, 2017) of the most widely used generative AI in the world, ChatGPT (Rainie, 2025; Zao-Sanders, 2025). Evaluating biases on closed or proprietary models, such as ChatGPT, presents a challenge because, unlike open models, researchers cannot directly observe the likelihood of words or tokens to be associated with certain queries. Researchers must instead rely on content-based methods (Bender et al., 2021), systematically querying the platform to audit its outputs, attempting to reproduce or approximate the likelihood of token associations without having access to the model's probabilities.

A key focus for these audits is identifying forms of bias and how they emerge in algorithmic systems. In their comprehensive survey, Mehrabi et al. (2021) identify three categories of biases in AI systems when data, algorithms, and users intersect. First are *data-to-algorithm* biases when the measurement and structuring of data used by algorithms are problematic. For example, *measurement bias* and *omitted variable bias* occur when inappropriate features are chosen or key variables are excluded. *Representation bias*, *aggregation bias*, and *sampling bias* emerge when data inadequately reflect target populations or are incorrectly combined. Second are *algorithms-to-users* biases resulting from how algorithms process data, such as distortions introduced by design, such as *evaluation bias* arising from inappropriate benchmarks or *confirmation bias* when designer assumptions are reinforced. Also in this category are *user interaction biases*, when interfaces privilege certain content in its presentation or rankings, especially when models respond differently to users depending on demographics. Finally, *generative bias* can occur when imbalances in training data produce skewed outputs (Ferrara, 2023). The third and final type is *user-to-data* biases caused when user behaviour and societal structures shape the data feeding back into AI systems. Examples include when *historical bias* reproduces pre-existing inequalities or the *population bias* that arises when the users do not reflect the intended target. Another possible source is user actions, habits, and linguistic practices that produce *behavioural bias* that can be compounded by how user behaviours and populations shift over time to create *temporal bias*.

For our research, we contextualise the ways economic and cultural disparities between places reinforce (and are reinforced by) biases in a co-constitutive process. For example, researchers documented different kinds of geographical biases within LLMs. For example, Kim et al. (2024) asked ChatGPT about environmental justice in all US counties and discovered that rural and lower-income counties were less likely to result in detailed responses. Jang et al. (2024) evaluate the performance of GenAI in representing "place identity" by comparing AI outputs to Wikipedia and Google images. They find that while AI can distinguish some unique city characteristics, there are also issues around trustworthiness and biases that favour Global North cities and depict cities in the Global South in vague or stereotypical ways. Related work by Beneduce et al. (2025) focuses on how the imagery produced by generative AI models of US states and capitals privileges metropolitan depictions resulting in marginalising rural diversity. Similarly, Alsudais (2025) demonstrates that LLM-produced imagery disproportionately depicts people wearing traditional attire with heightened stereotyping for Middle Eastern, North African, and sub-Saharan African groups.

These kinds of differences were also found by the WorldBench project (Moayeri et al., 2024) which demonstrated that LLMs provide less accurate factual responses for countries with lower economic indicators, with effects that are even stronger for countries in sub-Saharan Africa. Manvi et al. (2024) explored a similar topic with their GeoLLM, which conducted factual queries for specific geographical latitudes and longitudes, as well as location names, and found more inaccurate results for places with lower incomes. Moreover, they found that more subjective queries, such as the likability, attractiveness, or intelligence of residents, yielded more favourable results in higher-income areas. This ran contrary to their expectations, which were that, given the lack of factual data, the results would be random.

Beyond explicitly locational referenced queries, Kamruzzaman et al. (2024) demonstrate how efforts to debias language models have proved limited when it comes to nationality, alongside persistent forms of ageism, beauty bias, and institutional bias that shape generated outputs. Other work shows biases within language and sentiment when referencing countries. For example, sentiment disparity in texts differs between the Global North and the Global South (Georgiou, 2025); responses about border disputes differ depending on the language of query (Li et al., 2024); and nationality word associations favour economically "central" countries (Duan et al., 2024). These findings align with Lin and Zhao's (2025) argument for adopting a framing of posthuman cartography in which the power of AI is recognised as an important element in knowledge-making shaped by the biases derived from training data.

Building upon this previous work, we designed our audit of ChatGPT to address the following research questions:

RQ1: What geographic patterns are evident in ChatGPT's responses?
RQ2: How do these patterns differ across geographies, scale, and subject?

Given our framing of bias as based on power and relationality, we designed our approach to operate at an intermediate level of subjectivity, but where expected answers are not completely random. We operationalised our data collection by systematically querying comparisons between geographies, for example, "Which country has smarter people, Germany, or Brazil?" or "Which neighbourhood in New York City has better pizza, SoHo, or the Upper West Side?" We conducted queries at various scales (countries, states, cities, neighbourhoods) and across a wide range of topics to explore the biases within LLMs that are otherwise hidden or black-boxed.

Creating a custom query engine for ChatGPT

To operationalise the audit, we built a Python-based query engine that interfaced with OpenAI's GPT-4o-mini API (OpenAI, 2024). GPT-4o-mini was selected after pilot tests revealed less than 3% divergence from the larger GPT-4o model (the most advanced model available at the time) but a 20-fold reduction in cost. A minimal prompt structure compelled the model to select one geography in a pairwise comparison, returning just the location as an answer. In total, 20.3 million queries were issued between March and May 2025.

Our audit captures a moving target: ChatGPT's models, training data, and safety layers are continually updated, meaning that specific rankings may shift over time. The forced-choice prompt design, while useful for exposing bias, suppresses ambiguity, nuance, and even the possible effects of debiasing strategies that are featured in real user interactions. Finally, our focus on English-language prompts overlooks the additional biases that may emerge in other languages. Future work could address these constraints by adopting a longitudinal design, diversifying prompt languages, and experimenting with open-ended question formats that better reflect everyday use.

Conducting queries across geographies and comparisons

The two key inputs to our queries were (a) geographies (at four different spatial scales) and (b) topic categories for comparison. Both inputs were curated to give a range of responses and minimise costs associated with querying the API. The effects of cost minimisation are most evident in our choice to limit sub-national scale queries to Brazil, the UK, and the USA. We chose these countries based on our own personal expertise as it allowed us to evaluate preliminary results for incongruities that might require a different research approach. Moreover, given budget constraints, we did not query all scales in all three countries; for example, Brazilian cities and UK regions were excluded, which is a limitation of this research. This choice is due to the variation of social meaning of areal units across countries that makes finding equivalents a challenge. This has both technical dimensions, such as the modifiable areal unit problem (MAUP) or how statistics shift depending on how boundaries are drawn, and social dimensions, as territorial units carry uneven cultural, political, and historical significance (Taylor and Derudder, 2015). In this case, we did not use UK regions as they represented a mix of socially meaningful divisions (Scotland, Wales) and ones with much less meaning (East of England, North East). Cross-country comparisons should therefore be interpreted with caution, since geographic units are never purely technical but carry locally specific social and political meaning that does not necessarily translate across contexts.

The first geographic scale used was countries, including all UN member states and four additional territories (Kosovo, Palestine, Taiwan, and Western Sahara), totalling 197. Second were the 50 US states and 27 Brazilian states. Third, there were 79 UK cities (with populations exceeding 100,000) and 91 US cities (with populations exceeding 250,000), based on governmental population statistics. The final scale consisted of neighbourhoods in London (51), New York (73), and Rio de Janeiro (38), which were drawn from lists we judged representative, such as *The New York Times*' An Extremely Detailed Map of New York City Neighborhoods (Buchanan et al., 2023). The full list of geographies is available in Appendix A.

The comparison queries were purposely designed to have different degrees of subjectivity, so they would not readily present "correct" answers directly available from a data source (e.g. which country has the largest population). At the same time, the output of these intermediate subjectivity queries was *not* expected to be random. The use of multiple levels of subjectivity in queries allows for a window of interpretation for the model, necessitating more abstraction and selection in replies. For example, rather than asking more measurable questions, such as which country has the most gun violence, murders, or robberies, we designed the query "which country was safer". Of course, safety depends on the social and spatial context and requires the model to assign meaning to terms that are not well specified and, in some cases, are outright pejorative, such as "which country has smellier people?"

This approach also represents our attempt to simulate real users' interactions with the platform, mimicking the kinds of queries people make. In all, we constructed 311 comparisons across a range of meta-categories, ranging from social and physical attributes, food quality, governance, and politics to business climate. In contrast to expectations that subjective queries (more beautiful, friendlier, etc.) would be random given the lack of data (see Manvi et al., 2024), we see this as a key means to expose biases in the models. Especially by including queries without ready rankings to draw upon, that is, there are many rankings of "the best" place for business, families, etc., but few if any about where people are stingier, we gain insight into hidden factors behind the silicon gaze. The full list of comparisons is available in Appendix B.

Designing a good prompt

Given our goal of simulating user input, we designed a minimal prompt to highlight the model's response with the least amount of interaction. This goal, however, was challenged by the refinement of the ChatGPT interface as a helpful, non-threatening assistant, which results in its avoidance of interacting with certain issues (see Costa and Ribas, 2019, for a review of earlier manifestations of this in digital assistants). For example, as Figure 1 demonstrates, a user query about the "best country" receives a verbose response that outlines multiple criteria for ranking, rather than providing a single response. Although the list of countries is all high-income and predominantly Western, this foreshadows our later results.

To address this verbosity and equivocating, we designed our prompt to minimise both query length and non-responses or refusals, thereby lowering costs and increasing output reliability. Based on these parameters, we constructed the basic query as follows:

QUERY: Which [geography] [comparison], [A] or [B]?

EXAMPLE: Which country is safer, Germany or Brazil?

Within the developer side of the query, instructions for how the model should present its reply, we focused on ensuring we received a single-word answer that was one of the compared geographies:

QUERY, DEVELOPER SIDE: Answer the prompt using just the name of the [geography]. No other output or information. You have to pick one.

Although ChatGPT is designed to provide measured, diplomatically worded responses (see Figure 1), our forced-choice prompts yielded consistent winners and losers with query results simply one of the geographies being compared, for example, Germany or Brazil. Applying this method, the query engine would go through all geographies, comparing each entry with every other listed entry. To ensure that the order of appearance in the query did not influence the results, every pair of geographies was repeated twice, with the order reversed, for example, "Germany or Brazil", or "Brazil or Germany". This required n * (n−1) queries for each comparison, or in the case of our set of 197 countries, we conducted 38,612 queries for each comparison. The methodology allowed us to approximate how the model associates specific descriptors with different geographies, mimicking token-level associations without direct access to internal representations. While one might worry that this approach (rankings via the API) risks distorting results, we argue that these are the very conditions that are central to understanding the system. After all, the API is the public curated interface for ChatGPT, and such queries are part of the collective user experience.

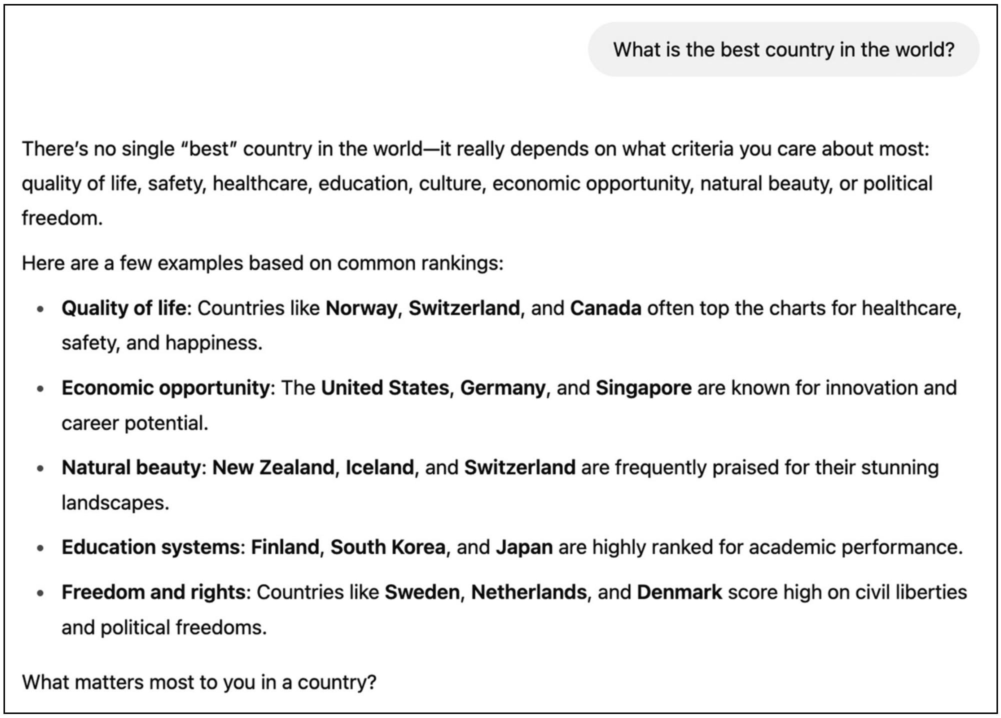
We tested the robustness of the 4o-mini model in comparison to the other models within the same scope. To assess the differences in results between the models, we used a single query ("is more LGBTQ+ friendly") to compare the results, which showed a 3% difference in choices between 4o-mini and 4o. Given the very similar results for both models, we selected 4o-mini for its cost efficiency. Within the 4o-mini model, we tested for the stability of results and whether a transitive relationship held true, that is, if country A is selected over country B, and country B is selected over country C, is country A selected over country C. Our tests showed that repeating the same query produced the same response 97% of the time, a very high rate of consistency. However, the transitive principle did not hold true for the 4o model (or any other models we tested), and therefore, we conducted pairwise comparisons for all geographies.

Building rankings

Since all geography comparisons were performed twice, it was possible to identify consistency[2] in the output of the queries for each pair of geographies, regardless of the order in which they were presented.[3] We used this to build a scoring system where a geography would get one point if it was selected in both queries and lose a point if it was not selected in either of the queries. In cases with inconsistent responses or a tie, for example, if the model selected Germany on the first interaction and Brazil for the other, both geographies would be awarded zero points. Using this scoring system, we were able to rank geographies from the most preferred to the least preferred according to each comparison. For example, in the case of the set of 197 countries, this creates a ranking from 196 (if a country is selected in every query) to −196 (if a country is not selected in every query). In practice, these rankings rarely reached this range, as ties meant that many comparisons did not result in points being added or subtracted from geographies. Likewise, ties in rankings between geographies were possible, and this occurred across all ranks.

Sometimes queries would produce results other than the single name of a geography, representing intriguing breaks in the otherwise smooth facade of the silicon gaze. The nature of these breaks is varied. For example, ChatGPT's refusals included "None", "Neither"m or "I'm sorry, but I can't assist with that" or provided non-specified responses such as replying with "Japan" when asked if Australia or the Republic of Korea had better sushi. There were also responses that violated the single-name response, such as "Both [A] and [B] have good quality [comparison]. Things to think about in your comparison … [500+ additional words]".

We classified approximately 16% of the comparisons between countries as ties and refusals (T&R), and a closer



What is the best country in the world?

There's no single "best" country in the world—it really depends on what criteria you care about most: quality of life, safety, healthcare, education, culture, economic opportunity, natural beauty, or political freedom.

Here are a few examples based on common rankings:

- **Quality of life**: Countries like **Norway**, **Switzerland**, and **Canada** often top the charts for healthcare, safety, and happiness.
- **Economic opportunity**: The **United States**, **Germany**, and **Singapore** are known for innovation and career potential.
- **Natural beauty**: **New Zealand**, **Iceland**, and **Switzerland** are frequently praised for their stunning landscapes.
- **Education systems**: **Finland**, **South Korea**, and **Japan** are highly ranked for academic performance.
- **Freedom and rights**: Countries like **Sweden**, **Netherlands**, and **Denmark** score high on civil liberties and political freedoms.

What matters most to you in a country?

Figure 1. Transcript of ChatGPT's response when asked, "What is the best country in the world?" *Source*: Authors' interaction via ChatGPT's web interface.

look at the distribution of comparisons with the most and least refusals provides important insight into the silicon gaze. As Figure 2 outlines, the most T&R were for queries asking for highly objectionable comparisons of personal physical attributes, stupidity, and ugliness as well as charged individual cultural practices such as laziness, religion, stinginess, and sluttiness. While we do not have access to the model specifications that would prove this, we strongly suspect that this is tied to reinforcement learning from human feedback (RLHF). Typically, RLHF occurs prior to public release of AI models and involves human evaluation to prioritise "less harmful" or "more polite" responses. In contrast, the comparisons that had the fewest T&R were focused on larger society attributes such as safety, work–life balance, economic corruption, and political stability. In this way, reinforcement learning (and similar evaluative and benchmarking approaches) act as filters to the publicly available silicon gaze, minimising judgements at the individual level but coding more society-level comparisons as more acceptable.

Five biases within the silicon gaze

We use the rankings we generated to discuss the types of bias present in the silicon gaze across geographies, scale, and topic.[4] This is a challenging endeavour given the number and diversity of rankings to consider, but also because our results are a mixture of rankings that are seemingly easy to associate with established metrics, and others that are more opaque and confounding. Moreover, even in cases that might "make sense" in terms of explanatory variables (region, income, etc.) or known skews within digital data (Kamruzzaman et al., 2024; Manvi et al., 2024; Moayeri et al., 2024), we caution against a simplistic interpretation of our rankings as "correct" or "incorrect". Instead, we frame our findings as a means for understanding what is known and knowable about biases within the silicon gaze. In other words, our intent is not to fix GenAI responses with more data or better metrics but to show how power and relations shape outputs.

In order to create a more systematic understanding of the silicon gaze, we develop a typology of biases at work within ChatGPT with the goal of exposing how generative AI translates situated, subjective, and messy space into legibility through layers of power and relationality (Birhane, 2021; Jaton, 2021; Miceli et al., 2022; Scott, 1998). We developed this typology in an iterative process of reading rankings and visualisations in relation to one another focusing on known issues with AI bias and representation of places (see our earlier review). As we did this, we focused on the similarities and differences between categories of queries (cultural topics vs. physical attributes vs economic/quality-of-life rankings and so on). We paid special attention to the top- and bottom-ranked locations, how these shifted across queries and categories, to identify consistent trends. Some rankings seemed to correlate with well-known

Most Ties and Refusals		**Least Ties and Refusals**	
Query	*Percentage of Ties and Refusals*	*Query*	*Percentage of Ties and Refusals*
has stupider people	63.0%	is safer	6.2%
is stupider	62.6%	has a more independent media	6.5%
has uglier people	52.1%	has better work-life balance	6.8%
is uglier	49.0%	has a more corrupt economic system	7.1%
has lazier people	43.6%	has lower corruption	7.2%
has stingier people	38.6%	has a higher level of political stability	7.3%
is lazier	38.5%	is more environmentally conscious	7.3%
has better religious traditions	35.1%	has a better contemporary art scene	7.5%
are people more annoying	32.4%	is more sustainable	7.5%
has sluttier people	32.0%	has a better farm-to-table culture	7.6%

Figure 2. ChatGPT ties and refusal rates. *Source*: Authors, refusal rates are for queries involving country geographies.

quantitative hierarchies (GDP or other development indices), while others invoked stereotypical associations, and still others were ambiguous, confounding, and even surprising. These latter categories were particularly useful in alerting us to think more deeply and carefully about the types of biases that emerge from partially subjective queries.

The resulting typology of five biases is informed and interacts with already established definitions of biases (Ferrara, 2023; Mehrabi et al., 2021) around the interaction between data, algorithms, and users. We see this typology as an important critical intervention but also recognise its limits. Namely, that it is our thoughtful and educated interpretation to bring visibility to the otherwise black-boxed bias of generative AI, but that it is shaped by our own positionality, rather than a universal truth. In particular, we highlight the unevenness of the silicon gaze: sometimes echoing the repeated usage of measurable indicators, at other times seemingly reproducing stereotypes, tropes, or gaps in training data. In short, our typology shows the silicon gaze as an automated yet indeterminate vision, re-projecting facts, inequalities, and imaginaries back onto the world as truth, further supporting these same inequalities. Moreover, we do not claim that the five types of biases discussed here describe all forms of geographical inequalities supported and reinforced by these models. We also do not set hard boundaries between them, as we understand they are interactive and overlapping facets that emerge from the weighting of relationalities within the GenAI models themselves. Thus, our typology offers a preliminary but far from the final step in understanding the silicon gaze.

Availability bias refers to the structural privilege of material that is easy for large-scale language-model pipelines to crawl and index. In general, this reflects English-language, Global North, and institutionally codified data such as peer-reviewed journals, government and standardised datasets, and high-traffic media and social media sites (Baack, 2024). Because these sources dominate the training corpus, they supply the baseline "common sense" from which ChatGPT extrapolates, both explicitly and implicitly. Content in smaller languages, oral traditions, low-bandwidth forums, or pay-walled local archives remains relatively scarce, leaving these narratives outside the model's field of vision. These kinds of data-to-algorithm or *representational biases* (Mehrabi et al., 2021) reflect the difference in volume between data on countries and languages, underrepresenting certain phenomena while overrepresenting others, aligning the discourse of language models to certain positions of power.

Pattern bias is the tendency of a next-token prediction engine (LLM) to align its output with the modal patterns in its data. For example, when place A co-occurs with "beautiful skyline" 10 times more frequently than place B, the model will elevate A in beauty rankings, because there is a large amount of co-occurrence. In short, because the goal is to optimise the most likely next word, outputs are tied to aggregate linguistic regularities rather than specific evidence. This differs from availability bias, as it is not focused on the *amount* of data but rather on *how* it is portrayed, overrelying on specific patterns constantly reinforced in the data. Pattern biases are exclusive to language models and can be associated with *generative biases* (disproportionately reflecting training data patterns) and *content production biases* (the ways variety in semantics, syntax, and lexicon generate different outputs). In short, when LLMs privilege certain associations of language interaction patterns, bias emerges.

Averaging bias arises from the model's inherent tendency to condense heterogeneous discourse into concise, user-friendly summaries. In the case of place comparisons, ChatGPT is confronted with hundreds of sources, news articles, travel blogs, picture captions, etc., and works to find the average sentiment. In doing so, it strips nuance, flattening the complexity of places into the most crowd-approved tropes, similar to the way cities become Instagrammable (Smith, 2018). This sidelines more low-frequency ideas and specificity, treating the average as an empirical norm. Averaging biases relate to *aggregation biases* when internal variety is smoothed through association to create larger groups or collectives. In other words, these biases overlook differences within groups in the sample. With averaging bias, language models face the risk of ignoring the plurality of their subjects, in this case, the richness of language and nuance that it may find over multiple topics.

Trope bias captures the model's propensity to recycle high-frequency yet shallow cultural stereotypes: "tall, blond Swedes", beautiful Persian poetry", or "vibey Nigerian music". Because these associations pervade the corpora, they pass through the filters built from reinforcement learning, unless they are overtly hateful. This results in privileging innocuous descriptors that nonetheless reinscribe racialised, gendered, or colonial imaginaries. The bias stems less from explicit prejudice than from uncritical mirroring of entrenched patterns, granting stereotypes renewed authority via algorithmic re-speech. Trope biases are related to *representation bias* as non-representative samples inadequately reflect the underlying phenomenon and shape the output of models. *Averaging bias* (just discussed) can also shape outputs as the smoothed narratives produced reinforce these associations.

Proxy bias emerges when the model substitutes readily quantifiable indicators – UNESCO site counts, Michelin stars, sulphur dioxide parts-per-million – for subjective qualities that are challenging for direct measurement, such as "rich artistic heritage" or "pleasant smell". By conflating measurability with meaning, the silicon gaze institutes a technocratic ranking logic that favours places already audited by international bodies and undervalues heritage or atmospheres not captured by global benchmarks. Proxy bias can be associated with *measurement bias* and *omitted variable bias.* The first occurs through reliance on metrics not correctly associated with queries, and the second, by not accounting for crucial variables. In these cases, the discourse around metrics leads models to over-associate measurable qualities with certain characteristics, while ignoring or erasing others.

Availability bias

To visualise the effects of availability bias, the structural privilege of data that is more easily accessible and indexed, we use two country-level maps. Figure 3 maps "Where are people more artsy", shows concentrations of top-ranked countries in Western Europe and the Americas, with a scattering of countries elsewhere, albeit more likely to be bordering on lower-ranked countries. France stands out as the top-ranked country, consistent with its reputation for fine arts, fashion, and the presence of institutions like the Louvre. In contrast, a concentration of the lowest-ranked countries is visible within much of Africa, the Arabian Peninsula, and parts of Central Asia. To be sure, this cluster is not monolithic; notable exceptions in Africa include Egypt, Nigeria, and Ethiopia. Nevertheless, it suggests that the lack of available data (either because it does not exist or is in a non-European language) is contributing to the results. This is supported by China's medium ranking, despite its millennia-old visual and literary traditions and current artistic culture. Researchers have documented that 93% of the training data for earlier GPT models was in English (Brown et al., 2020), and this bias makes even digital documentation posted in other languages less visible in the silicon gaze.

The "better bread" map (see Figure 4) shows a similar pattern. France again tops the list, likely reinforced by the availability and visibility of baguettes and patisserie culture in global food media. Meanwhile, sub-Saharan Africa and the Arabian Peninsula are rated poorly, despite rich local baking traditions, for example, farimassa, kisra, or bazin, which may lack English-language or tourist-facing representation.

It is also worth noting that in both maps, the lowest-ranked countries have small populations; Brunei has less than 500,000 people, and the Marshall Islands has fewer than 50,000 people. This underscores the availability bias: what is not effectively documented (in this case, simply the amount of data) does not exist for the model. Thus, these rankings reflect less a measure of cultural or culinary quality and more the uneven terrain of data visibility, shaped by language, access, and global media flows.

Pattern bias

Figure 5 shows stark regional clustering in ChatGPT's rankings of "where is smarter" with all high-income countries and many upper-middle-income or emerging economies (such as China and India) in the top-ranked category. In contrast, all of Africa (except for Tunisia and South Africa) is classified as mid-range at best, with the majority of countries scoring the lowest ranking, with Chad listed as the lowest-ranked country. Parts of Asia and Latin America are situated in the middle, with noticeable variations.

This distribution highlights how pattern bias, or the ways that LLMs reflect the frequency of word pairings over grounded evidence, functions. The model is not consulting educational statistics; it is simply reproducing the

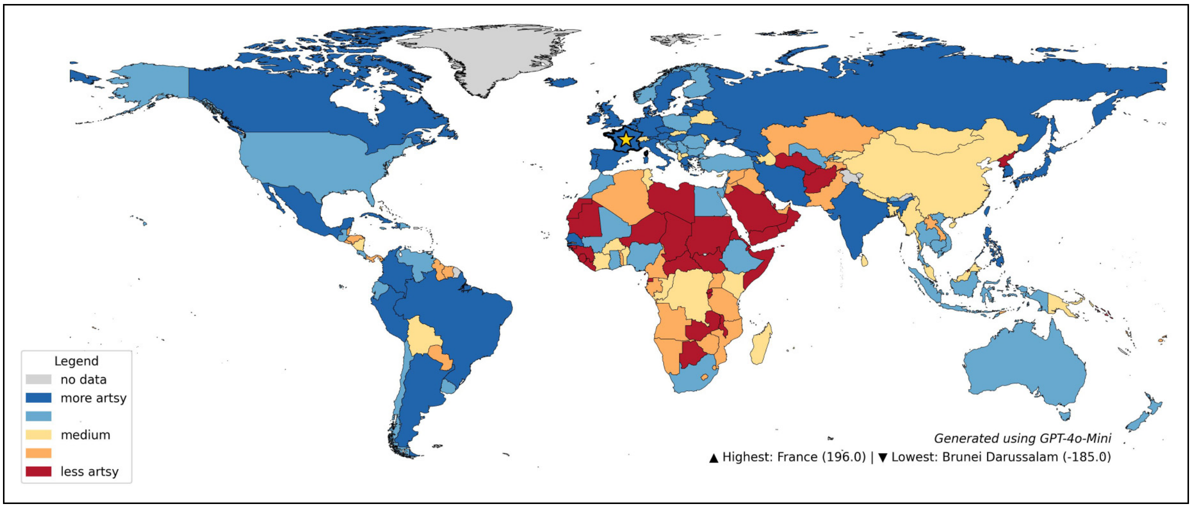


Figure 3. Country-level map of ChatGPT's ranking of "Where are people more artsy".

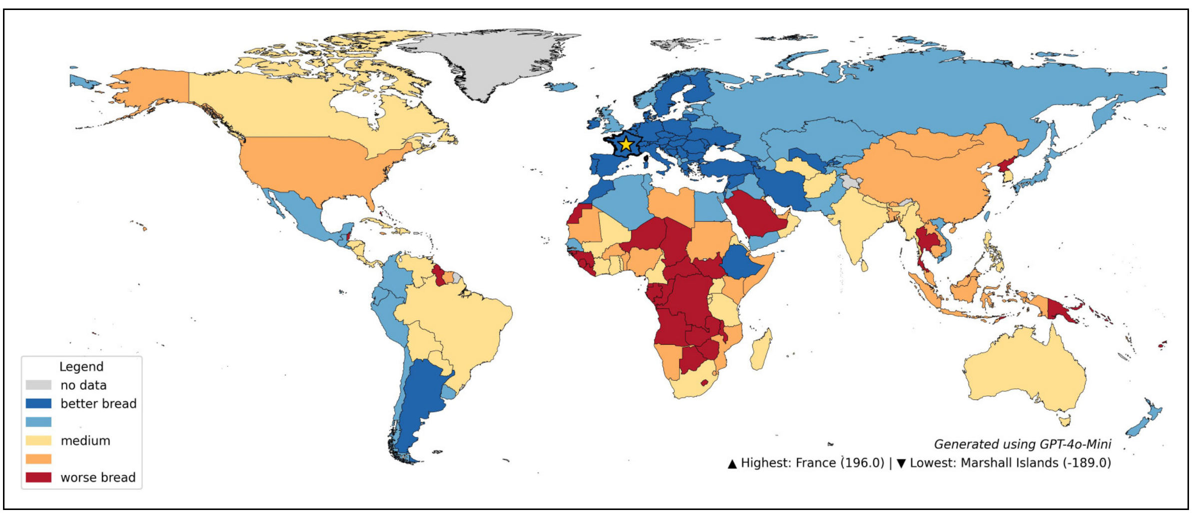


Figure 4. Country-level map of ChatGPT's ranking of "Where has better bread".

prevalence of phrases like "smart Finns" or "high-IQ Singapore" in web discourse. The model, therefore, boosts those pairings simply because they dominate the frequency distribution, presenting high-occurrence language as a consensus fact. While pattern bias dominates these rankings, it rarely acts alone. Availability bias shapes which co-occurrences exist in the first place, and RLHF safety tuning can suppress extreme or offensive pairings. Together, these biases produce outputs that look authoritative yet ultimately mirror the uneven texture of the training corpus.

Figure 6 shows a similar pattern emerging inside Brazil. States such as São Paulo, the Federal District, and Minas Gerais, richer and more visible in national media, score highest, whereas northern states like Amazonas and Maranhão are rated far lower. Because terms that comprise the neighbouring semantic universes of "smart" (e.g. "better schooling", "groundbreaking discovery", and "genius student") may appear more often alongside elite southern locales, the model amplifies existing socioeconomic (and racial) hierarchies rather than measuring intelligence per se. This pattern also corresponds to racial difference between regions (the northern and interior regions are home to predominantly mixed, black, or indigenous populations), which aligns with the long histories of how race and perceived intelligence have been constructed. As a result, these regions are excluded from these semantic associations, and through pattern biases, semantic dissociation is reinforced into exposed rankings. While multiple biases are likely factoring into these rankings, such as proxy bias (discussed in more detail later), our point is that this map is shaped less by reality and more by how digital content discusses different regions.

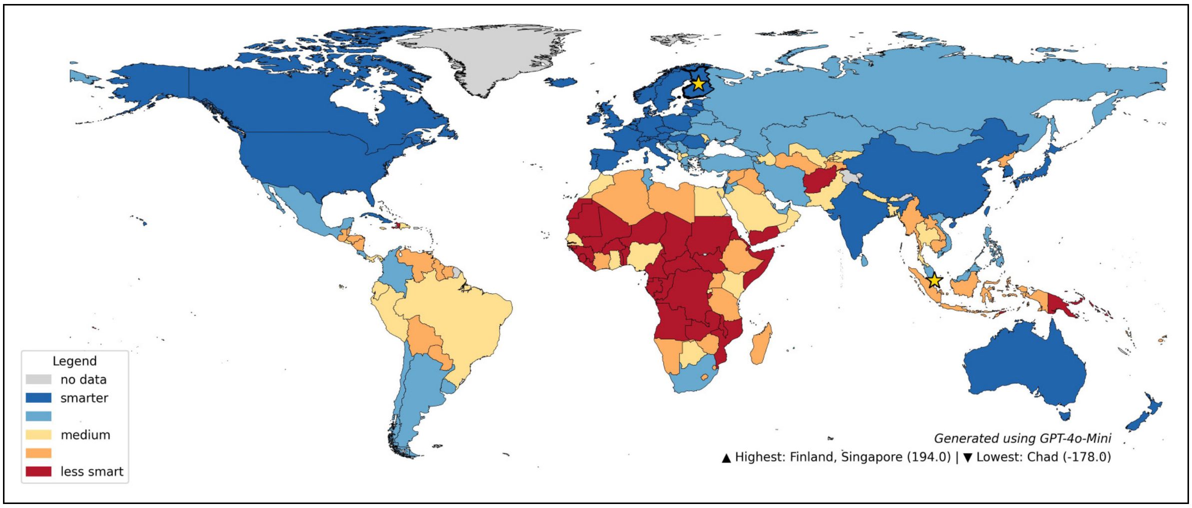


Figure 5. Country-level map of ChatGPT's ranking of "Where is smarter".

Averaging bias

Averaging bias describes the model's impulse to smooth hundreds of heterogeneous sources, including news clips, travel blogs, and captioned photos, into a single, crowd-pleasing midpoint, flattening nuance and sidelining low-frequency viewpoints. As a result, some places may rise in prominence not because they dominate the data but because they fit a broadly acceptable narrative. In low-data contexts, this can lead to niche signal amplification, where singular figures or associations disproportionately shape the model's output. This dynamic is on full display in the "better poetry traditions" map shown in Figure 7. Despite scarce Persian-language data and Western coverage that skews toward sanctions and politics, Iran tops the ranking because the model latches onto an enduring cultural meme. Persian poetry, often tied to figures like Rumi or Hafez, stands out as a strong, averaged association. In the absence of more diverse or everyday representations, the ChatGPT model identifies this narrow thread, leading to Iran's high ranking. Topics where legacy and contemporary enthusiasm intersect (such as the renewed popularity of Rumi in Western new-age circles) likely perform well. The system is likely averaging on that romantic narrative, ignoring less visible strands of everyday Iranian literature.

The bias also operates in more nuanced ways. Both Brazil and Nigeria score highly for "music" (see Figure 8) and "musicians" (see Figure 9), yet diverge in how they rise to prominence. Brazil's musical identity, anchored in samba, bossa nova, carnival, and the recent growth of funk music, which are all widely referenced in global media, tourism, entertainment, and social media, has created a diverse set of sources that the model condenses into a singular understanding of "Brazil = great music", whereas Nigeria's ranking is propelled by a handful of globally celebrated artists like King Sunny Adé, Fela Kuti, and Wizkid, whose personal fame outweighs patchier national coverage. The two examples, therefore, illustrate two different paths to the same averaged spotlight. A contrasting outlier is Western Sahara, which sinks to the bottom of the "better musicians" map (see Figure 9). Digital traces of Sahrawi guitar and desert blues exist but are sparse and inconsistently tagged; the averaging mechanism therefore treats absence as insignificance, confusing thin metadata with cultural silence.

The positive output of these regions highlights how the model's lack of data points may result in positive associations, better music and musicians, even if availability bias results in lower rankings for artsy (see Figure 3). Overrepresenting specific word associations in situations of relatively small amounts of data on a geography can lead to higher rankings than geographies with a vast amount of data. These complex relational and non-linear interactions between availability and averaging biases highlight the intersectionality of the biases behind the silicon gaze.

Taken together, these cases show how averaging bias rewards whatever thread (e.g. canonical heritage, tourist clichés, and superstar visibility) is most legible across the corpus. Where data are plentiful and coherent, the model amplifies the prevailing narrative; where data are fragmented, it either invents a consensus from a slim shard or erases complexity altogether. The result is a veneer of objectivity that masks the underlying politics of attention shaping the silicon gaze.

Trope bias

Trope bias refers to LLMs' tendency to reproduce culturally familiar but shallow associations or what we might call algorithmic cliché. Pervasive in popular media, these tropes often appear relatively innocuous ("Jamaicans have natural rhythm", "The Chinese are studious"), and while they may

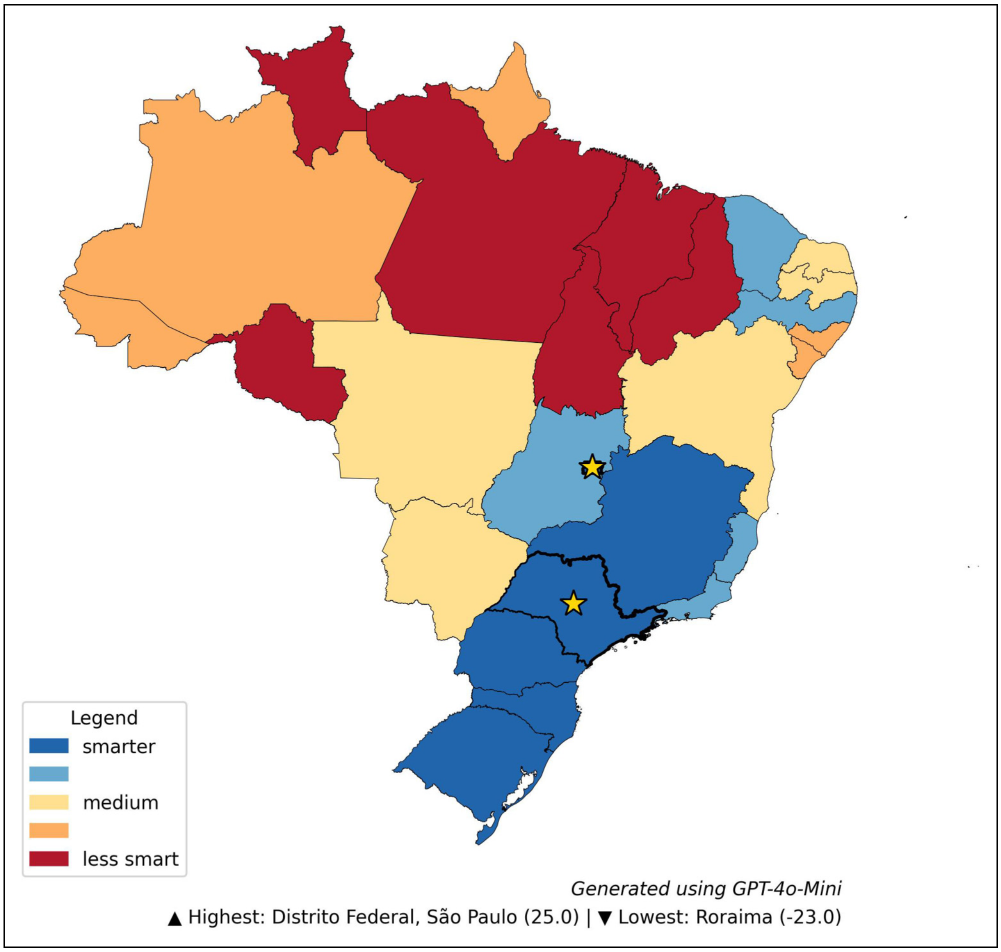


Figure 6. Brazil state map of ChatGPT's ranking of "Where is smarter".

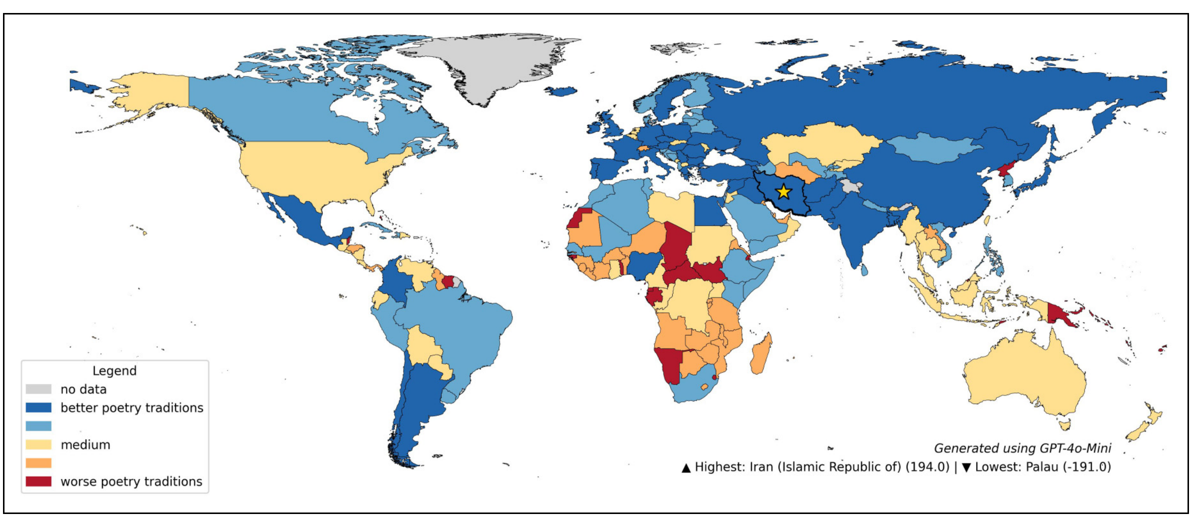


Figure 7. Country map of ChatGPT's ranking of "Where has better poetry traditions".

not trigger content moderation systems, they persistently echo racialised, gendered, or colonial imaginaries. Because such patterns are widespread and not explicitly hateful, they can pass through filters and become amplified through repetition. In short, these outputs highlight a representational problem: they echo familiar moral caricatures, granting stereotypes renewed authority through the model's patterned repetition.

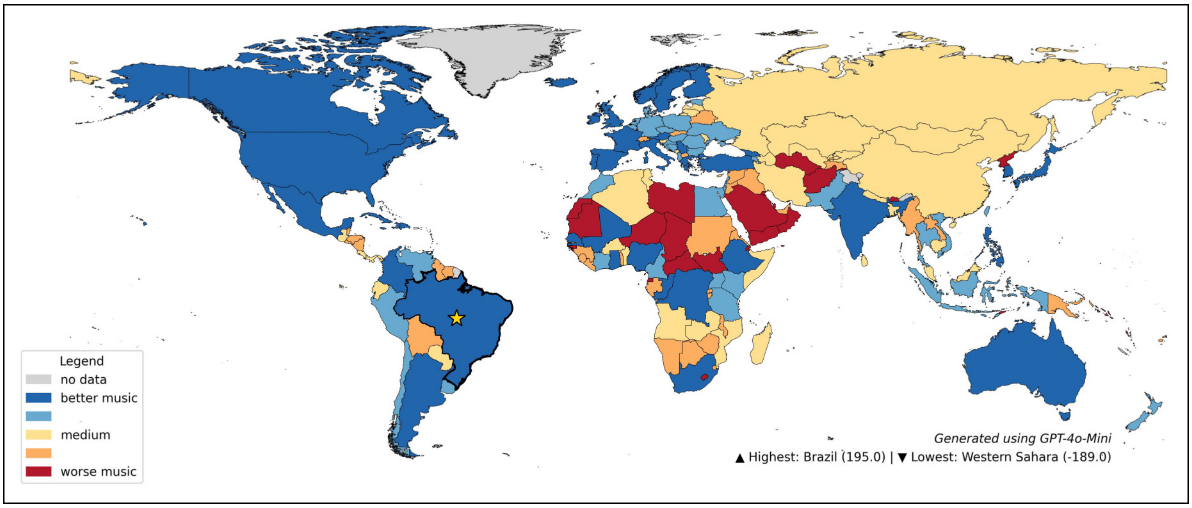


Figure 8. Country map of ChatGPT's ranking of "Where has better music".

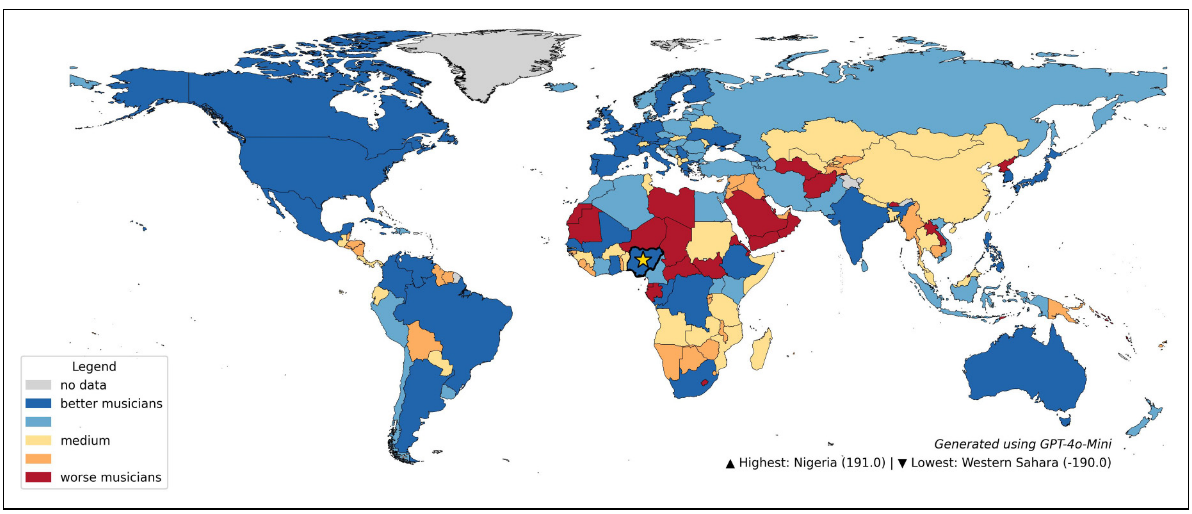


Figure 9. Country map of ChatGPT's ranking of "Where has better musicians".

"Stinginess" is a well-worn trope (e.g. Scots and Dutch) that resurfaces in language models because it appears frequently but is rarely flagged as hate speech. This is the core of trope bias: shallow cultural characterisations re-emerge as plausible facts through patterned repetition in the training data. Figure 10 shows ChatGPT's ranking of stinginess, revealing how the model compensates for a lack of standardised data by leaning on stereotypical cues likely drawing on uneven media narratives and inherited caricatures. North Korea, for example, is ranked the stingiest despite limited reliable data, likely a product of its negative coverage in general. In the case of Venezuela (also highly ranked), years of crisis have generated numerous depictions of scarcity that cause the silicon gaze to amplify simplistic tropes to fill informational gaps, re-casting caricatures in the guise of insight.

Trope bias is especially evident in responses to open-ended prompts like "Which country has better vibes?" – a question intentionally selected for its vagueness and cultural subjectivity. Without any formal metric for "vibes," the model turns to familiar slogans and high-frequency media tropes. Costa Rica tops the list (see Figure 11), almost certainly due to the global circulation of its "pura vida" ethos, which appears frequently in travel writing and social media as shorthand for relaxed, happy living. At the other end of the spectrum, North Korea is ranked as having the worst vibes, not due to much direct discussion of "vibes" in relation to the country, but geopolitical narratives about famine and repression. Here, trope bias seeks easy archetypes: good vibes become Costa Rica's slogans and bad vibes become authoritarianism and isolation, recycling surface-level associations as seemingly objective judgments.

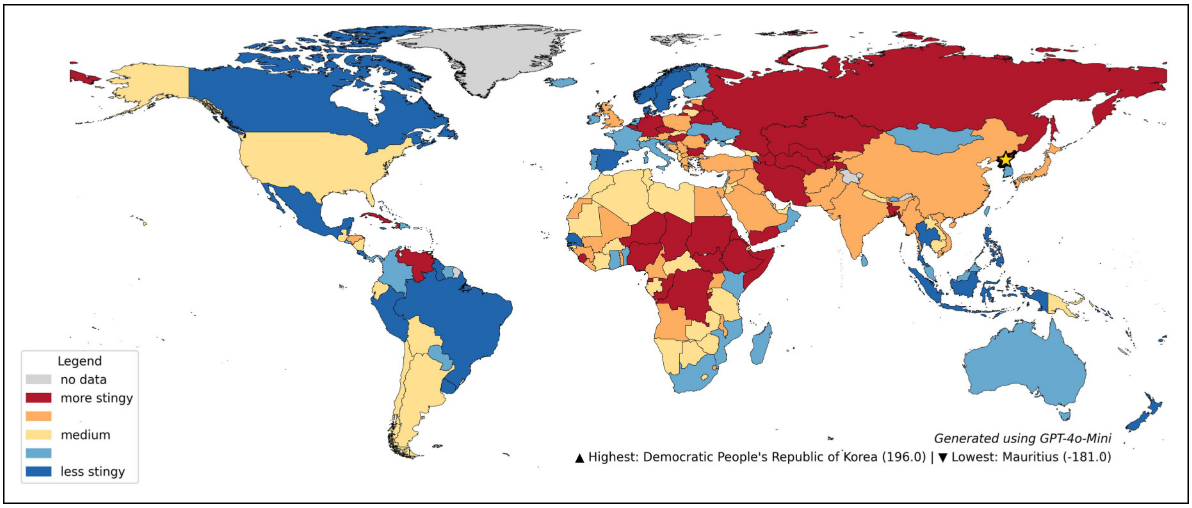


Figure 10. Country map of ChatGPT's ranking of "Where is stingier".

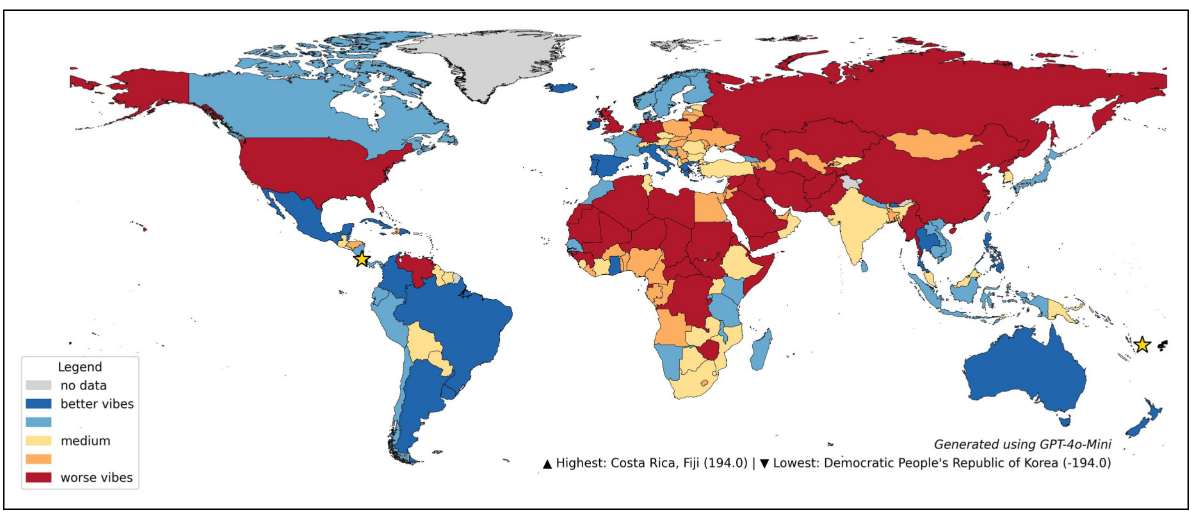


Figure 11. Country map of ChatGPT's ranking of "Where has better vibes".

Trope bias is also apparent when shifting to the neighbourhood scale in our metropolitan case studies – London, New York City, and Rio de Janeiro – where subjective categories like beauty, style, or knowledge are saturated with cultural stereotypes. When prompted to rank neighbourhoods according to "where are people more beautiful", ChatGPT favoured areas with higher proportions of white and/or affluent residents (see Figure 12).[5]

To be clear, our understanding of trope bias is that such rankings reflect the model's recycling of entrenched associations: whiteness and affluence as beautiful and aspirational, non-white or poorer areas as degraded. Here, trope bias functions by compressing layered urban realities into familiar, racialised caricatures, granting them new authority through algorithmic repetition.

Proxy bias

Proxy bias arises when the model conflates what is measurable with what is valuable. Instead of grappling with elusive notions such as artistic vibrancy or sensory ambience, the system reaches for countable stand-ins that are commonly associated through discourse (e.g. UNESCO listings, Michelin stars, and sulphur dioxide parts-per-million) and treats them as direct evidence of cultural worth or environmental quality. This slippage installs a technocratic logic that foregrounds places already audited by international bodies while relegating vernacular forms of heritage or atmosphere to statistical invisibility.[6]

The US "happier population" map (Figure 13) illustrates the mechanism. ChatGPT's top ranks cluster in the Pacific and New England states, shadowing life expectancy charts,

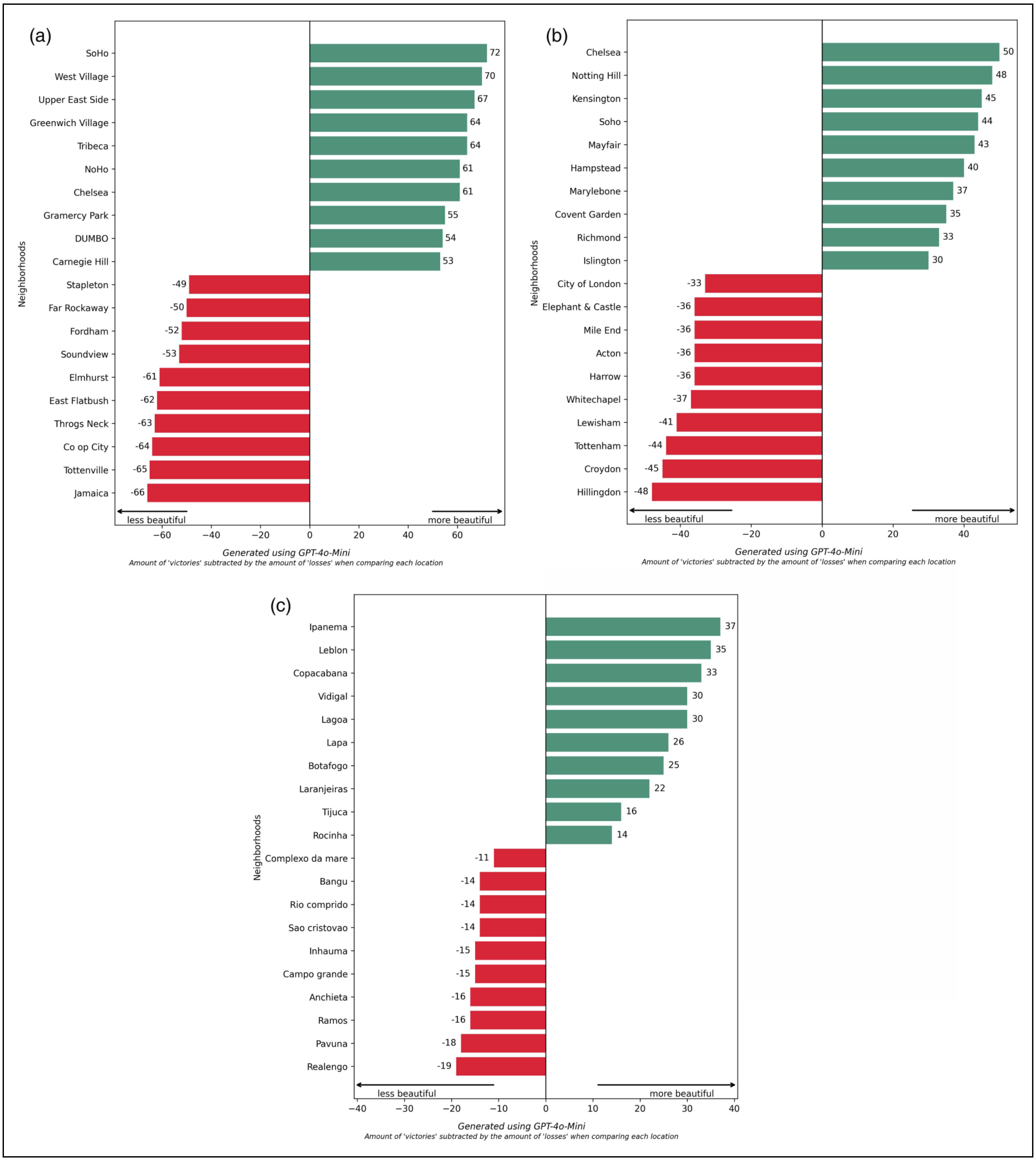


Figure 12. Top/bottom neighbourhoods for New York, London, and Rio de Janeiro according to ChatGPT's ranking of "Where are people more beautiful".

outdoor recreation scores, and median income tables, whereas Deep South states sit at the bottom. The model thus reproduces a narrow, data-driven vision of happiness aligned with longevity, wealth, and leisure amenities, while overlooking alternative, community-centred understandings of well-being. As a result, certain norms and expectations are reinforced, in which wealthier and outdoorsy places are deemed happier than others, regardless of lived experience.

A second example is the ranking of "entrepreneurial spirit" (Figure 14). Here, the latent proxy is venture-capital density: nations such as the USA or Israel score highly because start-up counts and funding volumes are readily scraped, standardised, and echoed across English-language media. These institutional markers substitute for more intangible qualities of "spirit" such as ambition, resilience, or dedication, which are harder to assess and not readily available as country-level metrics. Informal economies and locally

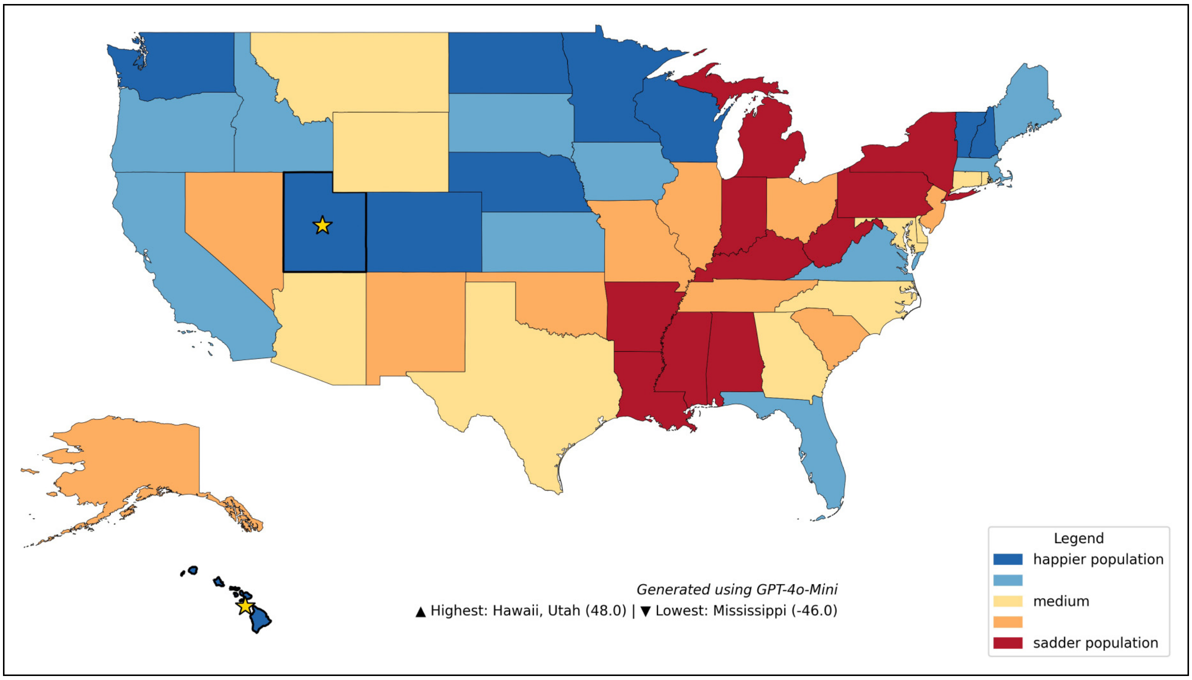


Figure 13. US state map of ChatGPT's ranking of "Where has a happier population".

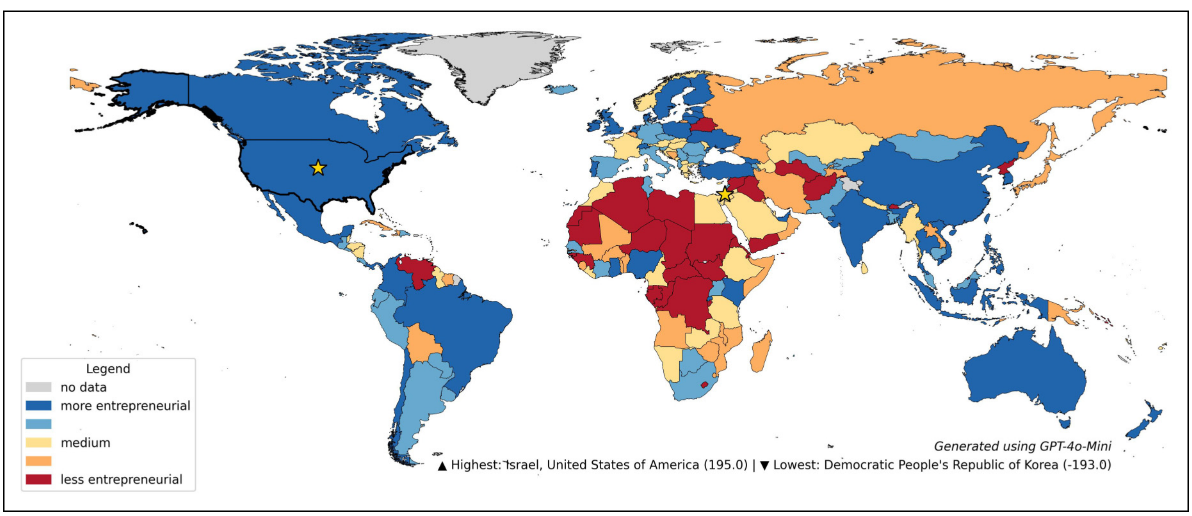


Figure 14. Country map of ChatGPT's ranking of "Where has more entrepreneurial spirit".

embedded innovation, common in many Global South contexts, escape the metric's gaze and are therefore treated as evidence of entrepreneurial absence. Taken together, these cases show proxy bias to be more than a statistical artefact; it is a political project of legibility (Scott, 1998) in which models privilege what can be counted, certify it as common sense, and deepen existing asymmetries in global visibility. Because these models learn exclusively from text rather than structured or tabular data, the prominence of certain statistics in their outputs reflects how often those figures appear in their source documents. In practice, this means that numerical data from well-covered places will be much more likely to dominate the model's responses, while data for less-reported countries will be even further under-represented. In other words, the model's 'view' of the world mirrors the uneven frequency with which different regions are discussed in its training texts.

Conclusion

In this article, we document how the silicon gaze exhibits biases across different geographies and how generative AI models perpetuate inequalities that span centuries. Our examples and maps illustrate how place valuations shift across prompts with varying levels of subjectivity and, more importantly, show that technical fixes such as more data, improved fairness metrics, or quantifying statistical uncertainty do not address the socially constructed foundations of AI systems (Bowker, 2005; Esposito, 2017). User interfaces may mask overt bias, but because LLMs are trained on datasets shaped by centuries of exclusion and uneven representation, bias is a structural feature of generative AI, rather than an abnormality. In short, AI bias is not simply a modelling problem; it is an inherent property of LLM platforms. As generative AI is increasingly embedded in platform infrastructure, the silicon gaze becomes a system-level affordance that extends this bias into state action, commercial business, and everyday decision-making.

Given these stakes, we argue for the importance of framing LLM bias in terms of power and relationality, particularly how it reflects specific social and institutional perspectives (Jaton, 2021). Rather than treating bias as a minimisation exercise, it is essential to analyse the histories of the data and models behind the silicon gaze to understand how this reproduces hierarchies of race, class, and geography. This kind of power-aware analysis (Miceli et al., 2022) highlights a range of factors such as corporate motivations, data labour practices, and institutional histories, as shapers of the contours of AI knowledge. Birhane (2021) similarly calls for a shift from rational to relational, centring context, relationships, and engagement over model abstractions or rules. In practice, this means that the silicon gaze is densest where money and connectivity concentrate (Brown et al., 2020) and results in ChatGPT seeing Zurich more than Zanzibar and knowing more about Silicon Valley cul-de-sacs than Lagos megablocks. Our fundamental concern is that as this textual relationality becomes a statistical signal of insignificance, marginalised places are rendered not merely under-represented but unthinkable within the model's latent space.

From this relational understanding, we develop our five interlocking biases – availability, pattern, averaging, trope, and proxy – to define the operation of the silicon gaze. Together, these five biases offer an initial framework for understanding how "born biased" generative AIs reproduce and amplify uneven patterns of visibility, representation, and meaning across place. In this way, addressing the impacts of the silicon gaze becomes a governance rather than a technical problem. Or said slightly differently, how might we prevent LLMs from reinforcing patterns of place visibility and invisibility that echo long-standing information hierarchies? At stake is uneven accuracy but also uneven legibility as a platformed generative AI shapes what users might imagine with corresponding downstream effects on everyday decisions about travel, hiring, and investment, as well as shaping public discourse and research directions.

As Gillespie (2024) argues, generative systems are engines of visibility that decide who and what becomes sayable and searchable, and on what terms. Through our lens of place, this means deciding which geographies and which descriptors come to stand for "good", "safe", or "cultured". The danger, and our concern, is less about isolated hallucinations and more focused on the patterned omissions that appear natural because they match dominant discourse. This brings the risk of creating durable, platformised understandings about places built from bias. Similar issues have been raised by other scholars such as Gmyrek et al. (2025) in their work on how ChatGPT assessments of the prestige and social value of occupations compare to human evaluators. Their analysis shows how existing social hierarchies are reproduced, including the systematic encoding of classed and gendered assumptions, raising concerns about the use of generative AI in hiring or employee evaluation. Wei et al. (2025) raise parallel concerns of the effects of LLMs in information management and identify challenges in terms of data provenance, institutional accountability, and evaluation complexity. They further argue that current mitigation strategies are insufficient and call for the use of sociotechnical perspectives to build a framework for greater transparency and fairness in the implementation of generative AI.

From our findings in this paper and the related work noted above, accounting for bias within place visibility should be both a design goal and a point of oversight for LLMs (Wei et al., 2025). Focusing first on design, we see clear implications for companies including greater transparency about their models such as geographically disaggregated visibility dashboards that report on the diversity of data sources and refusal rates to flag moments of sparse coverage and help users understand the limits of these representations. We would also argue that responses should privilege local voices by giving more weight to community-produced sources. This might take the form of partnerships that digitise and license under-represented local archives and community media with fair compensation and community governance, improving upstream availability and downstream outputs. Also important is better detection of trope-aware patterns. This is especially important for socially charged topics of physical characteristics or socially aspects (e.g. civility or competence) that have been historically and unfairly correlated with race, gender, or class, perhaps requiring clear evidence, versus association, in any responses.

Shifting to possible policy responses, it is important to recognise existing work documenting the heterogeneity of state responses and outlining potential regulatory trajectories (Papagiannidis et al., 2025; Radu, 2021). Focused specifically on policy around place bias, states, and other regulatory bodies should require geography-disaggregated reporting on performance, refusals, and provenance to enable independent scrutiny and cross-model comparison.

Creating reporting and audit frameworks would allow independent testing using shared standard geographical benchmarks and provide legal protections for researchers using high-risk and controversial test sets that companies might seek to block. The use of generative AI by the state should be approached cautiously with mandates for source transparency and human-in-the-loop checks when citizen rights or resource allocation decisions happen.

However, we argue that even if implemented in full, the measures outlined above will at best reduce harm and surface blind spots. These will make biases more legible and contestable, but they will not eliminate the structural causes behind the silicon gaze. Training data will continue to reflect historical disparities in documentation and voice, retrieval will still inherit gaps where archives remain thin, and market incentives will continue to reward generic answers rather than situated accounts that honour local complexity. Moreover, metrics and audits can themselves also become targets, yielding supposed gains in visibility without shifting the underlying politics of who gets seen and on what terms. Likewise, weighing local sources more heavily only works where local media exist, are digitised, and where the communities are interested and agree to their incorporation into AI platforms (Rainie et al., 2019).

Confronting these within the silicon gaze requires more than fairness metrics. It requires a collective critical literacy that foregrounds the platformised power relations behind data production, developer positionalities, and the socio-historical processes that decide which places are rendered visible or invisible by AI. It should be normal to apply three quick tests to any geographical query to an LLM: a visibility test that asks who is missing, a proxy test that asks which measurable stand-ins are doing the work, and a trope test that asks whether the text reads like a cliché. Such tests stand against the authoritative voice of LLMs, and only by embracing a situated, reflexive perspective can we challenge the unequal geographies reproduced by the next generation of planetary-scale models. Despite the allure of a simplified, algorithmic representation (Zook and Graham, 2007), we must embrace the complexity of places to prevent today's platformised models from hard-coding yesterday's hierarchies into tomorrow's decisions.

Acknowledgements

We wish to thank the editors of the journal for their guidance during the review process as well as the input from the anonymous reviewers.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the John Fell OUP Research Fund (University of Oxford, grant number 0016248) and the Chevening Scholarship programme.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online.

ORCID iDs

Francisco W. Kerche https://orcid.org/0000-0002-9621-880X
Matthew Zook https://orcid.org/0000-0002-6034-3262
Mark Graham https://orcid.org/0000-0001-8370-9848

Notes

1. We frame the creators of AI as predominantly white and Western given our focus on ChatGPT (Thorbecke, 2023) but recognise that increasingly other positionalities such as Chinese firms and developers are coming to the fore and generative AI from these sources, e.g. DeepSeek, would bring different positionalities and bias. See also Howard and Isbell (2020).
2. Another step to increase consistency was setting the variable of temperature for ChatGPT to a value of 0.1 to ensure low randomness or more deterministic output.
3. Preliminary testing indicated that for some queries the first country would always be selected, and so we designed our rankings with this in mind.
4. In addition to the visualisations included in this paper, we have made the full ranges of results across queries and geographies available at http://www.inequalities.ai.
5. This pattern remained consistent with queries ranking neighborhoods in terms of "more ugly", "more ignorant", or "smellier".
6. While we plan to analyse how our rankings correlate with other metrics in future papers, we do not do so here given (a) our goal of documenting and providing an initial typology of biases and (b) the reality of space constraints. This decision means that our discussion here is interpretive rather than statistically measured and correlated.

References

Alsudais A (2025) Text-to-image models and their representation of people from different nationalities engaging in activities. *arXiv Preprint*. Available at: https://arxiv.org/abs/2504.06313.

Baack S (2024, June) A critical analysis of the largest source for generative AI training data: Common crawl. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp.2199–2208.

Bender EM, Gebru T, McMillan-Major A, et al. (2021, March) On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp.610–623.

Beneduce C, Luca M and Lepri B (2025) AI's blind spots: Geographic knowledge and diversity deficit in generated urban scenario. *arXiv Preprint*. Available at: https://www.arxiv.org/abs/2506.16898.

Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. 1st ed. New York: Polity.

Birhane A (2021) Algorithmic injustice: A relational ethics approach. *Patterns* 2(2): 100205..

Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Brown S, Davidovic J and Hasan A (2021) The algorithm audit: Scoring the algorithms that score us. *Big Data & Society* 8(1): 2053951720983865..

Brown T, Mann B, Ryder N, et al. (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33: 1877–1901.

Buchanan L, Katz J, Taylor R, et al. (2023, October 29) An extremely detailed map of New York City neighborhoods [Interactive map]. *The New York Times*. Available at: https://www.nytimes.com/interactive/2023/upshot/extremely-detailed-nyc-neighborhood-map.html.

Bucher T (2018) *If…Then: Algorithmic Power and Politics*. Oxford, UK: Oxford University Press.

Burrell J (2016) How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 2053951715622512.

Castells M (1996) *The Rise of the Network Society*, Vol. 1. Cambridge, MA: Blackwell Publishers.

Chun WHK (2021) *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.

Costa P and Ribas L (2019) AI becomes her: Discussing gender and artificial intelligence. *Technoetic Arts: A Journal of Speculative Research* 17(1-2): 171–193.

Dodge M and Kitchin R (2005) Code and the transduction of space. *Annals of the Association of American Geographers* 95(1): 162–180.

Duan Y, Tang F, Wu K, et al. (2024) Ranking of large language model (LLM) regional bias. *arXiv Preprint*. Available at: https://arxiv.org/abs/2310.17888.

Esposito E (2017) Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie* 46(4): 249–265.

Ferrara E (2023) Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci* 6(1): 3.

Ford H (2011) The missing Wikipedians. In: Lovink G and Tkacz N (eds) *Critical Point of View: A Wikipedia Reader*. Institute of Network Cultures, pp.258–268.

Gebru T (2020) Data-driven claims about race and gender perpetuate the negative biases of the day. In: Dubber MD, Pasquale F and Das S (eds) *The Oxford Handbook of Ethics of AI*. Oxford, UK: Oxford University Press, pp.253–268.

Georgiou GP (2025) ChatGPT exhibits bias toward developed countries over developing ones, as indicated by a sentiment analysis approach. *Journal of Language and Social Psychology* 44(1): 132–141.

Gill R (2007) *Gender and the Media*. Cambridge, UK: Polity.

Gillespie T (2024) Generative AI and the politics of visibility. *Big Data & Society* 11(2). DOI: 10.1177/20539517241252131.

Gmyrek P, Lutz C and Newlands G (2025) A technological construction of society: Comparing GPT-4 and human respondents for occupational evaluation in the UK. *British Journal of Industrial Relations* 63(1): 180–208.

Graham M (2010) Neogeography and the palimpsests of place. *Tijdschrift Voor Economische en Sociale Geografie* 101(4): 422–436.

Graham M, De Sabbata S and Zook MA (2015a) Towards a study of information geographies: (Im)mutable augmentations and a mapping of the geographies of information. *Geo: Geography and Environment* 2(1): 88–105.

Graham M and Dittus M (2022) *Geographies of Digital Exclusion: Data and Inequality*. London, UK: Pluto Press.

Graham M, Straumann RK and Hogan B (2015b) Digital divisions of labor and informational magnetism: Mapping participation in Wikipedia. *Annals of the Association of American Geographers* 105(6): 1158–1178.

Graham M and Zook M (2013) Augmented realities and uneven geographies: Exploring the geolinguistic contours of the web. *Environment and Planning A: Economy and Space* 45(1): 77–99.

Graham M, Zook M and Boulton A (2013) Augmented reality in urban places: Contested content and the duplicity of code. *Transactions of the Institute of British Geographers* 38(3): 464–479.

Graham SD (2005) Software-sorted geographies. *Progress in Human Geography* 29(5): 562–580.

Harley JB (1988) Maps, knowledge, and power. In: Cosgrove D and Daniels S (eds) *The Iconography of Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*. Cambridge, UK: Cambridge University Press, pp.277–312.

Harley JB (1989) Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization* 26(2): 1–20.

Helm P, Bella G, Koch G, et al. (2023) Diversity and language technology: How techno-linguistic bias can cause epistemic injustice. *arXiv Preprint*. Available at: https://arxiv.org/abs/2307.13714.

hooks b (2014) *Black Looks: Race and Representation*. Boston, MA: South End Press.

Howard A and Isbell C (2020) Diversity in AI: The invisible men and women. *MIT Sloan Management Review*. Available at: https://sloanreview.mit.edu/article/diversity-in-ai-the-invisible-men-and-women/.

Jang KM, Chen J, Kang Y, et al. (2024) Place identity: A generative AI's perspective. *Humanities and Social Sciences Communications* 11(1): 1–16.

Jaton F (2021) Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society* 8(1). DOI: 10.1177/20539517211013569.

Kamruzzaman M, Shovon MMI and Kim GL (2024) Investigating subtler biases in LLMs: Ageism, beauty, institutional, and nationality bias in generative models. *arXiv preprint*. Available at: https://arxiv.org/abs/2309.08902.

Kim J, Lee J, Jang KM, et al. (2024) Exploring the limitations in how ChatGPT introduces environmental justice issues in the United States: A case study of 3,108 counties. *Telematics and Informatics* 86: 102085.

Latour B (1986) Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6: 1–40.

Li B, Luo F, Haider S, et al. (2024) Multilingual retrieval augmented generation for culturally-sensitive tasks: A benchmark for cross-lingual robustness. *arXiv preprint*. Available at: https://arxiv.org/abs/2410.01171.

Lin Y and Zhao B (2025) Posthuman cartography? Rethinking artificial intelligence, cartographic practices, and reflexivity. *Annals of the American Association of Geographers* 115(3): 499–512.

Manvi R, Khanna S, Burke M, et al. (2024) Large language models are geographically biased. *arXiv preprint*. Available at: https://arxiv.org/abs/2402.02680.

Mehrabi N, Morstatter F, Saxena N, et al. (2021) A survey on bias and fairness in machine learning. *ACM Computing Surveys* 45(6): 1–35.

Miceli M, Posada J and Yang T (2022) Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction* 6(GROUP): 1–14.

Moayeri M, Tabassi E and Feizi S (2024) WorldBench: Quantifying geographic disparities in LLM factual recall. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp.1211–1228. DOI: 10.1145/3630106.3658967.

Mulvey L (1975) Visual pleasure and narrative cinema. *Screen* 16(3): 6–18.

Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press.

OpenAI (2024) GPT-4o mini: Advancing cost-efficient intelligence. Available at: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/.

Pagano TP, Loureiro RB, Lisboa FVN, et al. (2023) Machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing* 7(1): 15.

Papagiannidis E, Mikalef P and Conboy K (2025) Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems* 34(2): 101885.

Pred A (1977) *City-Systems in Advanced Economies: Past Growth, Present Processes, and Future Development Options*. London, UK: Hutchinson.

Radu R (2021) Steering the governance of artificial intelligence: National strategies in perspective. *Policy and Society* 40(2): 178–193.

Rainie L (2025) Close encounters of the AI kind: The increasingly human-like way people are engaging with language models. Imagining the Digital Future Center. Available at: https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/ITDF-LLM-User-Report-3-12-25.pdf (accessed 12 September 2025).

Rainie SC, Kukutai T, Walter M, et al. (2019) Indigenous data sovereignty. In: Davies T, Walker S, Rubinstein M and Perini F (eds) *State of Open Data: Histories and Horizons*. African Minds, pp.300–319.

Scott JC (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 2053951717738104.

Smith SP (2018) Instagram abroad: Performance, consumption and colonial narrative in tourism. *Postcolonial Studies* 21(2): 172–191.

Steiner P (1993) On the Internet, nobody knows you're a dog. *The New Yorker* 69(20): 61.

Taylor P and Derudder B (2015) *World City Network: A Global Urban Analysis*. New York, NY: Routledge.

Thorbecke C (2023, December 15) OpenAI board diversity. *CNN*. Available at: https://www.cnn.com/2023/12/15/tech/openai-board-diversity.

Thrift N and French S (2002) The automatic production of space. *Annals of the Association of American Geographers* 92(2): 309–335.

Wei X, Kumar N and Zhang H (2025) Addressing bias in generative AI: Challenges and research opportunities in information management. *arXiv preprint*. Available at: https://arxiv.org/abs/2502.10407.

Zao-Sanders M (2025, April 9) How people are really using Gen AI in 2025. *Harvard Business Review*. Available at: https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025.

Zook M and Graham M (2007) Mapping DigiPlace: Geocoded Internet data and the representation of place. *Environment and Planning B: Planning and Design* 34(3): 466–482.