The Geopolitical Bias of Generative AI: A Call for Country-Level Dataset Transparency (CLDT)

Chaiya Tantisukarom, Independent Researcher

September 24, 2025

Abstract

Generative Artificial Intelligence (GenAI) models deployed in high-stakes sectors like medicine (medGenAI) and law (lawGenAI) exhibit a critical risk of perpetuating global disparities. This paper argues that this output bias is directly proportional to the **geopolitical disparity** inherent in the models' training datasets. We propose a framework for mandatory **Country-Level Dataset Transparency (CLDT)** based on quantifiable metrics to assess the **imparity risk** and empower practitioners in underrepresented countries to apply necessary human oversight. This approach shifts the focus from general fairness audits to specific, computational **jurisdictional accountability**.

1 Introduction: The Quantifiable Imparity Gap

GenAI models, built on massive, uncurated web-scraped corpora, are not globally neutral. The concentration of digital content production in a small number of high-income, Western nations results in models that are statistically robust but contextually brittle outside of those domains [1]. For a user in an underrepresented country C_i, the model's output reflects the dominant statistical patterns, disease profiles, legal precedents, and resource availability of the Top N source countries. We define the **Imparity Gap** as the quantifiable risk of applying statistically validated, yet contextually irrelevant, GenAI recommendations due to a fundamental mismatch between the training data's origin and the deployment context.

2 Literature Review

Research on AI bias has evolved from addressing individual fairness attributes (gender, race) to tackling systemic and geographic disparities. Studies show that Large Language Models (LLMs) are **geographically biased**, exhibiting differential knowledge and sentiment based on location, directly correlated with socioeconomic factors of the data's origin [1]. This inherent imbalance is not merely linguistic; it is **ontological**, embedding a specific worldview that marginalizes alternative perspectives.

In healthcare, this bias has critical consequences. Research on **Polygenic Risk Scores (PRS)**, which underpins modern predictive medicine, found a substantial drop in predictive accuracy (up to 4.9-fold) in non-European populations due to their underrepresentation in training data [2]. Furthermore, image recognition systems in clinical settings, such as **dermatology**, have demonstrated poorer performance on darker skin tones because the majority of validated images originate from lighter-skinned populations [4]. The quantified bias, which shows a significant drop in diagnostic accuracy for darker skin types, highlights the severe real-world consequences of poor dataset diversity [4]. The challenge extends to the need for better data collection protocols to improve future diagnostic systems [3].

In the legal domain, the primary challenge is **jurisdictional relevance**. The deployment of lawGenAI has already shown the dangers of relying on models that hallucinate or misapply legal precedents. When a model's foundation is dominated by a foreign common law tradition (O3), its advice for a civil law jurisdiction (O1) becomes dangerously unreliable, leading to real-world sanctions against practitioners [5]. This body of evidence necessitates a shift toward transparency metrics that quantify the risk of geographic and jurisdictional misalignment.

3 Key Terminology and Definitions

For the purpose of applying the CLDT framework and analyzing domain-specific risks, we define the following operational terms related to data origin and utility:

Term	Definition and Context
O1 (Own Country Data)	The subset of the GenAI model's training data that originates specifically from the user's local jurisdiction (C_i). In a lawGenAI context, this is the essential corpus for local legal validity.
O2 (Neighboring Data Pool)	The aggregated data subset from geographically or culturally proximate countries. In medGenAI, this data often shares relevant disease profiles, ecological factors, or common resource availability.
O3 (Dominant Source Data)	The data subset contributed by the top N globally dominant countries (the source of the geopolitical bias). This data drives the model's highest statistical confidence but introduces the highest risk of contextual irrelevance for C_i.
O4 (Entire Dataset)	The collective training corpus of the entire foundational model (N). This represents the maximum statistical robustness and global knowledge base, but it is heavily skewed by the O3 contribution.

4 Proposed Transparency Metrics for Accountability

To address the imparity gap, AI labs must disclose the proportional data contribution from each country to enable risk calculation.

Let C_k denote the data contribution (e.g., in terabytes, tokens, or records) from Country k, and N be the Total Contribution ($N = \sum_{k=1}^{189} C_k$). The Top N contributing countries are denoted T_N.

4.1 Core CLDT Metrics for User C_i

The interface must display the following metrics, which serve as the foundation for risk assessment:

1. **Local Relevance Score (ByTotal):** The proportion of data contributed by the user's country C_i. This score measures the statistical influence of the local context on the entire model.

$$\text{ByTotal}(C_i) = \frac{C_i}{N}$$

2. **Dominance Concentration (TopNByTotal):** The collective proportion of data contributed by the dominant countries (O3). This score represents the maximum potential source of external, and often culturally biased, influence.

$$\text{TopNByTotal} = \frac{\sum_{k \in T_N} C_k}{N}$$

3. **Data Imparity Index (ByTopN Ratio):** The ratio of local data (O1) to the dominant data (O3). This index provides a relative measure of sparsity. A score close to zero indicates severe disparity.

$$\text{ByTopN Ratio}(C_i) = \frac{C_i}{\sum_{k \in T_N} C_k}$$

5 Domain-Specific Disparity Risks

5.1 LawGenAI: The High Demand for O1 (Own Country)

Legal advice is fundamentally bounded by jurisdiction. The risk in lawGenAI stems from the model prioritizing O3 data when C_i's legal corpus (O1) is sparse. For a lawyer in C_i, the optimal output should be highly weighted towards local data. The utility space is highly jurisdictionally bounded, where the risk of failure is an intersectional constraint:

$$\text{Utility}_{\text{Law}} \propto \text{Relevance}_{\text{Local}} \cap \text{Validity}_{\text{Jurisdiction}}$$

If $\text{Validity}_{\text{Jurisdiction}}$ is compromised by low $\text{ByTotal}(C_i)$, the entire output is flawed, making the use of O3 data for local statutes dangerous.

5.2 MedGenAI: The Relevance vs. Robustness Trade-off

Medical AI must balance the statistical power of the massive O4 (Entire Dataset) with local and regional relevance (O1 and O2). The medGenAI might provide a high-confidence diagnosis (justified by O4's statistical robustness) but an irrelevant treatment plan. For instance, a disease model might be statistically validated globally, but its utility for a practitioner in C_i depends on whether it incorporates local (O1) or proximal (O2) disease variants, availability of local medications, or common environmental risk factors. The doctor's challenge is managing the risk ratio:

$$\text{Med Risk} \propto \frac{\text{Confidence}_{\text{Global}}}{\text{Relevance}_{\text{Local}} + \text{Relevance}_{\text{Regional}}}$$

The CLDT metrics provide the quantitative context needed to apply an expert adjustment filter to the statistically dominant, but potentially irrelevant, O4 recommendation.

6 Proposed Solution: Interoperable Risk Vetting (ICV)

To mitigate the existential risk in critical decision-making, we propose mandatory, contextual **Cross-Domain Constraint Vetting** between medGenAI and lawGenAI at the point of output.

Inter-Domain Constraint Vetting (ICV):

1. The medGenAI generates a clinically optimal recommendation R_{Med} (utilizing the statistical power of O4).
2. The lawGenAI module, constrained strictly to the user's local legal context (i.e., operating almost exclusively on O1 data), evaluates R_{Med}. It checks for conformance regarding informed consent, malpractice precedents, and local regulatory hurdles.
3. The final output R_{Final} is R_{Med} adjusted for local legal compliance:

$$R_{\text{Final}} = R_{\text{Med}} \setminus \text{Legal Risk}_{C_i}$$

This framework transforms the geopolitical data disparity from a latent ethical problem into a computational safety requirement for real-world deployment. The mandatory display of CLDT metrics is the necessary first step toward true algorithmic accountability in a globalized world.

References

[1] Manvi, R. et al. (2024). Large Language Models are Geographically Biased. https://arxiv.org/html/2402.02680v2

[2] Martin, A. R. et al. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*. https://pubmed.ncbi.nlm.nih.gov/30926966/

[3] Rezk, E. et al. (2022). Leveraging Artificial Intelligence to Improve the Diversity of Dermatological Skin Color Pathology: Protocol for an Algorithm Development and Validation Study. *JMIR Research Protocols*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8941446/

[4] Daneshjou, R. et al. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9374341/

[5] Wikipedia. (n.d.). Steven A. Schwartz's AI-generated court case citations. https://en.wikipedia.org/wiki/Steven_A._Schwartz's_AI-generated_court_case_citations