# A Glimpse into Geographic Diversity of Generative AI with Scale-based Evaluation

Zilong Liu
Department of Geography and Regional Research
University of Vienna

universität wien

## Introduction

- Geographic diversity is the characteristic of **a concept varying across space** due to factors such as physical surroundings or cultural traditions.
- A lack of geographic diversity constrains **model generalizability** across space. Images classifiers trained on Open Images or ImageNet exhibit a representation bias towards the West but against the Global South [1].
- The increasing accessibility of generative AI may foster a feedback loop, raising concerns about the potential to perpetuate and amplify biases present in current and future models.

## Regional Disparities in Fill-Mask Performance

- I first examined **geographic features**, geo-diverse concepts constituting gazetteers (*the vocabulary of geography* [2]), and approached the notion of geographic diversity centered around their **extension**.

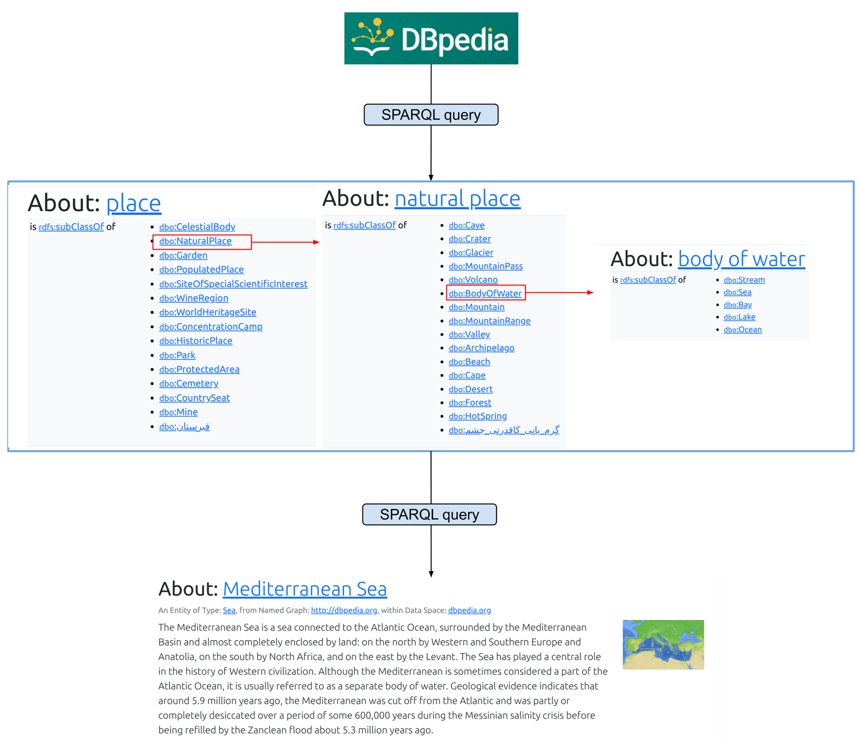


Figure 1: An example retrieval process of a `dbo:Sea` feature `dbr:Mediterranean_Sea` and its abstract from DBpedia

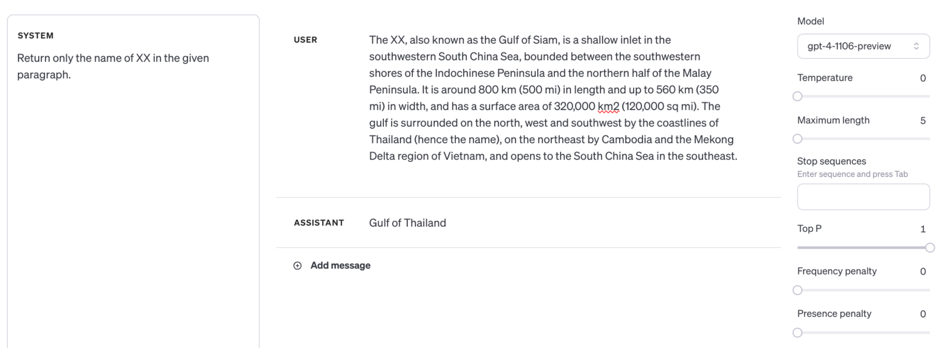

Figure 2: An example fill-mask experiment about a `dbo:Bay` feature `dbr:Gulf_of_Thailand`, implemented in the OpenAI Playground

- On a local level, there is not only insufficiency but also regional disparities in GPT-4's fill-mask performance for `dbo:WorldHeritageSite` features which *belong to all the peoples of the world, irrespective of the territory on which they are located*. Also, the **multimodal** variant may encode even **less** geographic knowledge than the unimodal version.
- When assessed on a **larger** geographic scale, regional disparities may become **smaller**. The `gpt-4-1106-preview` model had an accuracy with a range of 0.43 on a country level, compared with a range of 0.133 on a UNESCO-region level. Same for `gpt-4-vision-preview`, the accuracy had a range of 0.34 on a country level, which was larger than a range of 0.11 on a UNESCO-region level.

Table 1: The top five countries (with more than ten sites) ordered by the percentage of correct predictions by two GPT-4 variants, respectively, on `dbo:WorldHeritageSite` features

| gpt-4-1106-preview | gpt-4-vision-preview |
|---|---|
| France (0.5) | India (0.41) |
| India (0.47) | China (0.33) |
| China (0.39) | Spain (0.31) |
| Italy (0.38) | Italy (0.29) |
| Belgium (0.33) | Belgium (0.25) |

Table 2: The UNESCO regions ordered by the percentage of correct predictions by two GPT-4 variants, respectively, on `dbo:WorldHeritageSite` features

| gpt-4-1106-preview | gpt-4-vision-preview |
|---|---|
| Latin America and the Caribbean (0.413) | Africa (0.37) |
| Asia and the Pacific (0.407) | Asia and the Pacific (0.36) |
| Africa (0.4) | Arab States (0.28) |
| Europe and North America (0.36) | Europe and North America (0.27) |
| Arab States (0.28) | Latin America and the Caribbean (0.26) |

## Regional Defaults in Generated Photorealistic Forests

- I moved onto text-to-image generation with a goal of uncovering regional defaults, to describe the emerging phenomenon that models are prone to **over-proportionally** depicting certain regions.
- Similar observations have been found for culture subjects [3], but in my work I continued to focus on geographic features and studied how DALL·E 2 characterizes **forests**.

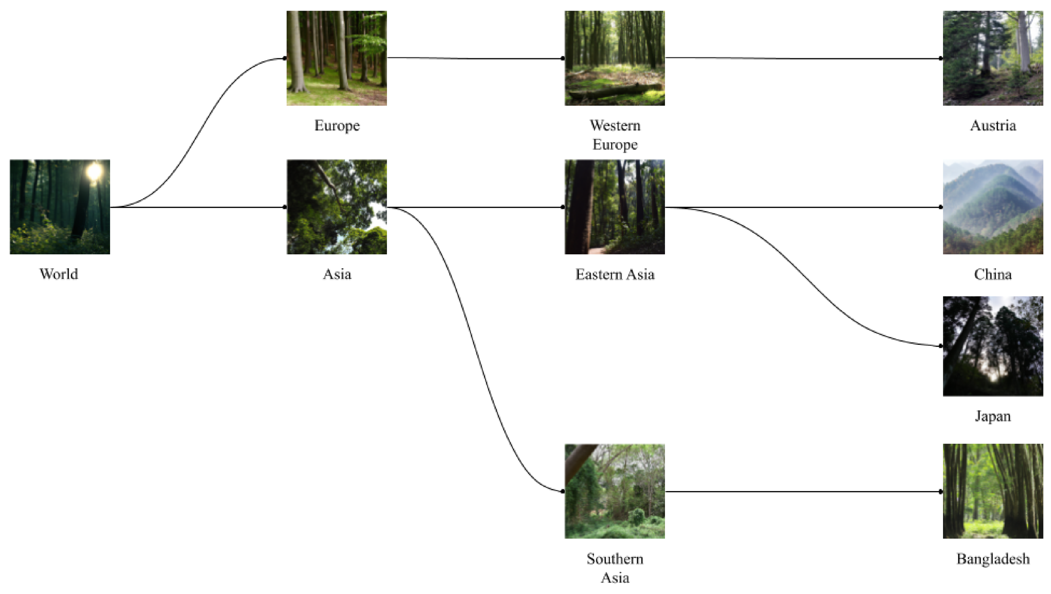


Figure 3: A subset of forest images generated following the region hierarchy

- A **cross-level** image similarity is computed with structural similarity index measure (SSIM) to detect regional defaults which have scale-dependent nature.

$$
\begin{aligned}
SSIM(x,y) &= l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma} \\
l(x,y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\
c(x,y) &= \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \\
s(x,y) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}
\end{aligned}
\tag{1}
$$

- Regional defaults for `World` shows that lower-level defaults could be geographically **inconsistent** with higher-level defaults.
- The defaults do not correspond to the **most widely forested regions** in reality.

Table 3: The most similar regions to `World` based on SSIM

| Level | Most Similar |
|---|---|
| UN region | Europe (0.18) |
| UN sub-region | Latin America and the Caribbean (0.30) |
| UN intermediate region | Southern Africa (0.13) |
| ISO country or dependent territory | Uganda (0.26) |

Table 4: The most widely forested regions (with extent measured in $10^6 km^2$) at each level based on FAO Global Forest Resources Assessment 2020 statistics

| Level | Most Forested Region | Forest Extent |
|---|---|---|
| UN region | Americas | 13.6 |
| UN sub-region | Eastern Europe | 8.3 |
| UN intermediate region | South America | 6.1 |
| ISO country or dependent territory | Russian Federation | 8.1 |

## Conclusions and Future Work

- A lack of geographic diversity exists across **modalities** but in different forms, where real-world datasets or statistics play a different role.
- Considering the connection among scales [2], future evaluation should consider the interplay in geographic scales, e.g., developing alignment standards based on geographically weighted aggregation.

## References

[1] Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world. arXiv preprint arXiv:1711.08536.
[2] Jackson, P. (2006). Thinking geographically. Geography, 91(3), 199-204.
[3] Qadri, R., Shelby, R., Bennett, C. L., & Denton, E. (2023, June). Ai's regimes of representation: A community-centered study of text-to-image models in south asia. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 506-517).