

Reproducibility Review of:

# Assessing the Geographic Diversity of AI's Spatial Representations in Image Generation



| Item                 | Value                                                                                                                                                                                                                                                                       |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title                | Assessing the Geographic Diversity of AI's Spatial Representations in Image Generation                                                                                                                                                                                      |
| Authors              | Zilong Liu, Krzysztof Janowicz and Mina Karimi                                                                                                                                                                                                                              |
| Full paper reference | Liu, Z., Janowicz, K., and Karimi, M.: Assessing the Geographic Diversity of AI's Spatial Representations in Image Generation, AGILE GIScience Ser., 7, 10, <a href="https://doi.org/10.5194/agile-giss-7-10-2026">https://doi.org/10.5194/agile-giss-7-10-2026</a> , 2026. |
| Codechecker(s)       | Frank O. Ostermann (ORCID 0000-0002-9317-8291)                                                                                                                                                                                                                              |
| Date of Check        | 10-04-2026                                                                                                                                                                                                                                                                  |
| Summary              | Successful but partial reproduction                                                                                                                                                                                                                                         |
| Repository           | NA                                                                                                                                                                                                                                                                          |
| Ref. certificate     | Ostermann, FO (2026): Reproducibility review of "Assessing the Geographic Diversity of AI's Spatial Representations in Image Generation" <a href="https://doi.org/10.17605/OSF.IO/vcgy2">https://doi.org/10.17605/OSF.IO/vcgy2</a>                                          |

**Table 1: Reproduction metadata**

| Output                          | Comment                                                     |
|---------------------------------|-------------------------------------------------------------|
| repro_rev_2026_29_notebook.html | HTML-export of Jupyter notebook containing the reproduction |

**Table 2: Summary of output files generated**

# Summary

The paper describes a study that has three main steps: First, generation of imagery using several chatbot interfaces. Second, assessment of the generated images by a combination of human and AI annotators. Third, descriptive statistics on the depicted landmarks using the annotations from step 2 and querying of Wikidata items for calculating similarity and diversity metrics.

This reproducibility review only attempted to reproduce the third part, based on the provided annotation files and Python code in the form of a Jupyter notebook that generates the main figures used in the paper.

The other two parts should in principle be replicable because the main information is provided. However, for this review they were considered to be out-of-scope because of inaccessibility to the exact same LLM for the first part, lack of process detail for the second part, and required time.

The reproduction of the third part went smoothly, with all figures being identical to the ones produced for the paper. This review reports thus a successful partial reproduction.

## Reproducibility reviewer notes

The paper comes with code repository, generated imagery, and documentation how to use both. There is also an application that allows convenient browsing of the input data (prompts and generated images). However, since this is not strictly part of or necessary for the research study, I did not test this further.

## Installation prerequisites and computational environment

The instructions were very clear and straightforward and could be implemented without any problems.

## Running the code

The provided Jupyter notebook was run smoothly without any issues. On a Windows 11 / Ubuntu 24.04 laptop with 16 GB RAM and an Intel i5-8265U CPU, code execution took just a few seconds.

## Outputs and results

The produced images are identical to the ones in the paper. Part of the analysis depends on querying an external source (Wikidata), which might change. Given the type of

features queried, this is unlikely but still possibly. Any future reproductions that observe differences to the reported metrics should consider this possibility.

## Recommendations

While this reproducibility review went smoothly and the authors are to be commended on their documentation (very clear and nicely aligned with the paper structure!), from a general reproducibility perspective, the study will pose problems to anyone attempting a reproduction.

As mentioned, the exact same AI models are unlikely to be available to a future researchers. This is clearly not something that the authors can control and a common limitation to contemporary research using proprietary closed-source AI models that require high-performance computing infrastructure to run.

However, some potential for improvement can be found with the annotation part, which is sparsely documented, with the exact usage of the mentioned tools and the interactions with the human annotators being so vague that any attempt at reproduction seems impossible. This is unfortunate, because the output images have been provided, so this part brings all the necessary data to attempt reproduction. I encourage to authors to consider providing more information on this step, ideally for this study already, but at least for future studies.

## Acknowledgements

NA