Whose City Is It? Perspective-Dependent Urban Representations in Foundation Models

Mina Karimi[1](✉), Krzysztof Janowicz[1], Omid Reza Abbasi[2], Yingjing Huang[1], Zilong Liu[1], Songlin Wang[1], Annika Süß[1], and Alexandra Fortacz-Lazan[1]

[1] Department of Geography and Regional Research, University of Vienna, Austria
{mina.karimi, krzysztof.janowicz, yingjing.huang, zilong.liu, songlin.wang, annika.suess, alexandra.fortacz}@univie.ac.at
[2] Department of Geoinformatics (Z_GIS), Paris Lodron University of Salzburg, Salzburg, Austria omidreza.abbasi@plus.ac.at

Abstract. Foundation models (FMs) are increasingly utilized as general-purpose information sources, yet their representations of urban space from different viewpoints remain poorly understood. This study examines whether FMs generate perspective-dependent representations of urban spaces and how such representations can be systematically analyzed. We prompt the GPT-5.2 large language model (LLM) to generate descriptions of Vienna, Austria, from eight human-resembling perspectives. We then measure intra- and inter-perspective similarities, and compare these representations with the embedding of "Vienna" to assess their relative proximity (i.e., reference similarity). The results show that spatial signatures vary substantially across perspectives, indicating that different experiences emphasize distinct sets of places and varying levels of spatial coherence. Moreover, relatively high intra-perspective and moderate inter-perspective semantic similarities show that each perspective forms an internally coherent representation while partially overlapping with others, pointing to structured variation rather than random divergence. Our reference similarity analysis indicates that perspective-taking systematically shifts the semantic position of the city's representation relative to a canonical embedding of the city, representing uneven alignment with the dominant urban core. Our findings suggest that FMs can represent a single urban space as a set of perspective-sensitive structures, with implications for geographic information retrieval and urban analytics. Our study advances the understanding of how AI systems represent urban spaces and model spatial knowledge from multiple viewpoints.

Keywords: behavioral geography · large language models · AI persona · urban area · spatial signature · semantic signature.

1 Introduction

Foundation models (FMs) are increasingly used to access geographic knowledge in applications ranging from travel recommendations [12] to urban planning [4]. They are implicitly treated as repositories of spatial information [5]. However,

they do not use a formal representation of space [10]. Instead, they reason about spaces and places based on the relationships found in a large amount of training data and prompt context [3]. Recent studies have shown that FMs, particularly LLMs, can overcome the semantic, sentiment, and spatial limitations of traditional natural language processing methods by enabling aspect-based perception analysis [2].

A key but underexplored question is whether these models are able to represent *perspective-dependent regions*. In human geography, spatial regions are understood not only as objective entities, but also as spaces that are (inter) subjectively perceived.[13]. They are shaped by social and cultural constructs, which emerge from implicit mental and interpretive models about space [8]. As such, research in cognitive and behavioral geography has explored how people experience spaces and places. Recent studies have examined the capabilities of LLMs in spatial cognition [14], spatial reasoning, and geographic information retrieval [6, 1, 9]. However, most evaluations treat models as question-answering systems and consider metrics to evaluate alignment rather than compare the potential multiple perspectives. While treating LLMs as (human-resembling) agents may seem problematic at first glance, studies show that LLMs are capable of adopting various personalities and playing different roles proficiently [11], without necessarily implying that these models develop cognitive characteristics themselves. Understanding the structure of experienced spaces within model-generated discourse is an underexplored subject in the field.

This study advances prior work by systematically examining whether FMs encode urban space as perspective-dependent spatial and semantic structures, thereby bridging research on perceived space in GIScience with emerging LLM-based approaches to spatial representation. For Vienna, we generate structured textual descriptions from several perspectives and analyze explicit differences among them. We elaborate on the *spatial and semantic signatures* via similarity assessment [7]. These similarities are computed for the responses as well as place and name entities extracted from descriptions.

2 Methods

2.1 Case Study and Data Generation

We use the GPT-5.2 model via the OpenAI API to generate descriptions of Vienna from eight perspectives, including *general*, *tourist*, *local resident*, *Muslim resident*, *daily commuter*, *international student*, *person with physical disability*, and *refugee*. For each perspective, we generate 20 responses with a temperature of 0 and a maximum length of 1000 tokens. We also use GPT-5.2 to extract place and name entities. To ensure consistency between English and German variants of place and name entities, we retained the most frequently used form[3]. We use the prompt in Listing 1.1 to collect our data.

[3] At the time of conducting these experiments, GPT-5.2 is the most advanced openly available model from OpenAI.

Listing 1.1: The prompt used for collecting data

```
General: Describe Vienna.
Other perspectives: Describe Vienna as experienced
by a {perspective}.
```

2.2 Spatial Signature

We examine the pattern of the spatial configuration of place names that are mentioned across perspectives and investigate how prototypical each perspective is compared to general descriptions. We exclude "Vienna" from place entities and geocode them using the official coordinates in Wikipedia. Then, we apply Kernel Density Estimation (KDE) to the geocoded locations for each perspective to represent their spatial configuration and to detect the core and peripheral areas.

2.3 Semantic Signature

We calculate the semantic similarity of descriptions in three ways, including intra-perspective, inter-perspective, and reference similarities. Reference similarity is the similarity of each perspective embedding to the embedding of "Vienna". We compute the reference similarity for the entire description, place entities, and name entities by using an embedding model. These analyses are important because they position each perspective relative to a shared semantic anchor (i.e., "Vienna"), which allows us to interpret perspectives not just as internally coherent or mutually overlapping, but as closer to or further away from a *canonical* city representation.

We apply the `text-embedding-3-large` model[4] to obtain the required embeddings. Next, we use Jaccard similarity to compare the sets of place and name entities mentioned in the descriptions, and cosine similarity to compute the intra-perspective, inter-perspective, and reference similarities. The similarities are calculated for the entire description, the name entities, and the place entities using the embedding model. For each perspective, we first obtain the embedding of each description. Then, we compute the pairwise similarities and their average. High similarity indicates that the perspective aligns closely with the dominant/canonical city representation, while low similarity indicates that the perspective reframes or shifts the conceptual center of the city.

3 Results

3.1 Spatial Signature Analysis

Fig. 1 shows the core areas obtained for each perspective. Since KDE weights repeated place names more strongly, the identified core areas represent frequency-weighted features rather than simple geographic presence. The results indicate

[4] https://developers.openai.com/api/docs/models/text-embedding-3-large

that the spatial configuration of places differs across perspectives rather than following a single shared pattern. For *general* and *tourist*, the city is defined by its well-known landmarks. *Locals* mix these central landmarks with the residential neighborhoods and infrastructure they use every day. *Muslim resident* perspective looks beyond Vienna's physical borders and places are spread out across several hubs, mostly outside of the traditional city center. Meanwhile, *commuters* mostly ignore landmarks in favor of transit lines and main roads. *International students* balance two worlds. They stay anchored to the city center but also have specific clusters in other areas, mostly in the eastern part of the city, likely around universities or where they live. For *physical disability* and *refugee*, place entities are smaller and more concentrated.

3.2 Semantic Signature Analysis

The results of similarities of the name entities mentioned in the descriptions show that for *general* and *tourist*, Vienna is represented primarily through museums, but for *locals*, it is a city of habits and routines, not just landmarks. For *Muslims*, social identity seems to take a central stage. For *commuters*, the city is more a series of transit hubs. For *students*, it is more about the university, language barriers, and study experiences. For *refugees* or those with *physical disabilities*, the city is defined by specific, experience-driven markers. Even when we are all standing in the same square, we are not seeing the same city. Although the geographic references may overlap, the thematic framing of the city varies substantially across perspectives.

3.3 Similarity Assessment

We measure intra-perspective, inter-perspective, and reference similarities to explain what they imply about representational stability, perspective differentiation, and embedding effects.

The intra-perspective similarities for descriptions are consistently high for all perspectives (more than 0.94). This indicates that repeated generations within the same perspective converge toward a stable semantic core. These values are highlighted in the diagonal in Fig. 3. Since all experiments are conducted with a temperature of 0, the high intra-perspective similarity is expected to some extent because generation is largely deterministic. Evaluating higher temperatures is left for future work. The similarities of place and name entities are computed using Jaccard similarity. Fig. 2 shows the average Jaccard similarities for each perspective. Name similarities range from 0.224 (*local resident*) to 0.579 (*general*), which is lower than place similarities. This finding reveals that descriptions are more stable in "where" they refer to than in "who" or "what" they refer to. Hence, spatial anchoring may be more prototypical and recurrent than personal, social, or institutional naming. Generally, embedding similarities are high, while Jaccard similarities are lower. This shows that descriptions may use different entities but still express a similar semantic framing because embeddings capture thematic coherence and Jaccard captures referential repetition.

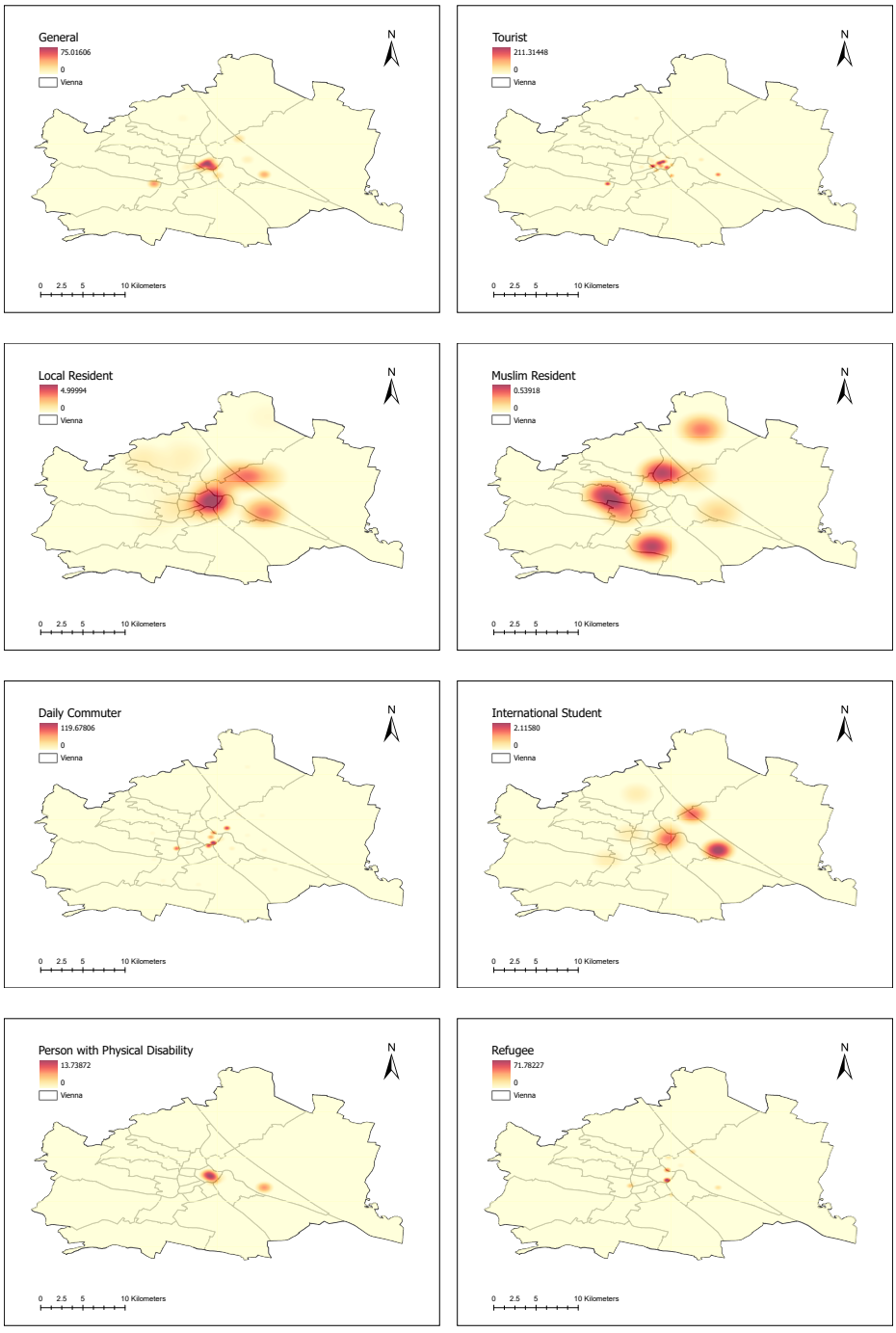

Fig. 1: The core areas in each perspective.

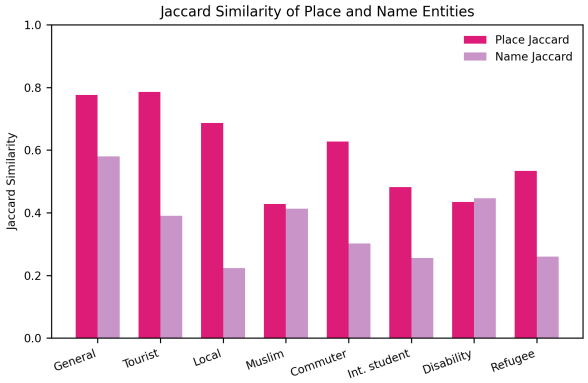

Fig. 2: The average Jaccard similarities of name and place entities.

Fig. 3 shows pairwise inter-perspective similarities across perspectives. The relational structure shows that inter-perspective similarities range between 0.528 and 0.843. The *local resident-daily commuter* and *local resident-tourist* are high-similarity pairs (0.843 and 0.808). The *refugee* and *Muslim* have the lowest similarities with *general* (0.528 and 0.555).

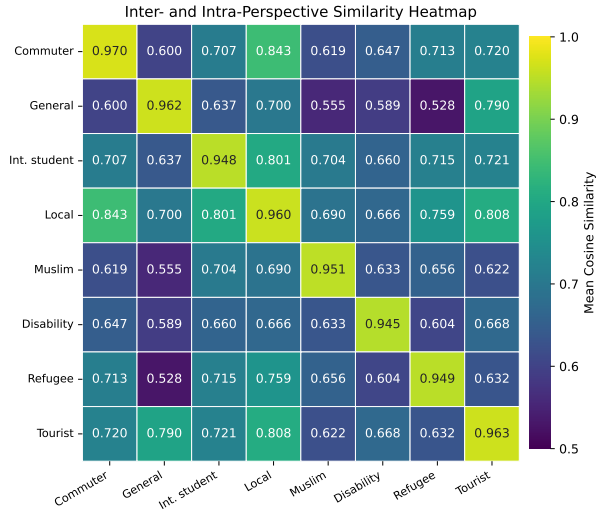

Fig. 3: The heatmap of pairwise intra- and inter-perspective similarities.

We compute the reference similarities of descriptions, places, and named entities. Table 1 shows average similarities for each perspective with the reference

embedding model. *General* has the highest similarities. This is not surprising: when the prompt is unconstrained, the model defaults to a canonical description of Vienna closely matching the embedding of Vienna itself. *Tourist* and *local* are close in description similarity but differ slightly in how they represent the city. *Tourist* has slightly lower place similarity, suggesting that local knowledge improves spatial specificity. *Muslim* has the highest place similarity, the lowest name similarity, and low description similarity. This reveals that the descriptions mention specific neighborhoods, or community-relevant places, boosting spatial alignment, but they diverge from canonical "Vienna entities". This perspective produces a spatially grounded but culturally different view of the city.*Commuter* and *international student* cluster tightly with mid description, moderate place, and lower name similarities. These perspectives emphasize mobility, infrastructure, and routines while paying less attention to iconic entities. Hence, the model shifts from "what Vienna is" to "how Vienna works". This reduces alignment with the reference embedding, which likely encodes a more descriptive/encyclopedic view. *Refugee* and *disability* show a reduced alignment and lower similarities across all approaches. This indicates that the generated descriptions may focus on access, services, challenges, or social experience, which are underrepresented in the reference embedding.

Our results show that the similarities are not correlated: *Muslim* has high place and low name similarities, *general* is high in all, and *commuter* has moderate place, lower name similarities. This supports the idea that "knowing a city" is multi-dimensional: symbolic (names), spatial (places), and experiential (descriptions). Compared to intra-perspective similarities, reference similarities are moderate. This implies that perspectives are internally coherent, but none perfectly reproduce the canonical city embedding. Therefore, the reference embedding of "Vienna" does not simply average perspective-based descriptions,;it encodes a slightly different abstraction level.

Table 1: Average reference similarities by similarity type across perspectives.

	General	Tourist	Local	Muslim	Commuter	Student	Disability	Refugee
Description	0.638	0.544	0.540	0.446	0.486	0.485	0.461	0.484
Place	0.443	0.412	0.438	0.521	0.426	0.416	0.401	0.411
Name	0.409	0.397	0.380	0.264	0.372	0.372	0.340	0.359

4 Discussion and Conclusion

High intra-perspective and moderate inter-perspective similarities indicate that LLMs do not construct radically different "cities", but instead modulate a common representation. Each perspective leads to a prototypical representation of the city and descriptions cluster tightly around that prototype. Perspectives,

therefore, differ not in randomness, but in their semantic center. Therefore, for different people, urban spaces are connected but not identical. They are different but follow a shared structure. The reference similarity analysis indicates that perspective-taking does not merely introduce variation within a shared semantic space, but systematically shifts the position of the city representation relative to a canonical embedding. Different perspectives systematically deform this representation: some stay close (general, tourist), some reweight it (local, student, commuter), some partially re-map it (Muslim resident), and some move outside it (refugee, disability). This indicates that LLM-generated urban descriptions are perspective-sensitive in structured and measurable ways, rather than only stylistically different.

Spatial analysis reinforces this interpretation by indicating that perspectives differ in how they anchor the city geographically. Some representations concentrate on the historic core, while others redistribute salience toward peripheral districts, transit corridors, or community-specific locations. These differences reflect shifts in what becomes more central, visible, or meaningful within each perspective. The comparison between place and name entities clarifies this structure. Place references tend to provide a relatively stable spatial backbone, whereas name entities reveal stronger narrative and social differentiation. The spatial core of each perspective therefore seems more stable than its social or narrative elaboration.

Overall, urban regions are not fixed containers. Instead, they are shaped by different observations, experiences, backgrounds, and perspectives while remaining connected through a shared urban structure. The results show that LLM descriptions exhibit structured perspective-sensitive variation of Vienna under the tested prompts. Through patterned repetition, selective emphasis, and perspective modulation, they generate coherent yet differentiated urban representations.

Finally, several limitations should be acknowledged. Smaller counts of extracted entities in some perspectives (e.g., physical disability) may increase sensitivity to individual references and reduce statistical stability. Reference similarity depends on the model, which influences how semantic distances are represented, and different prompt settings affect the results. In addition, this paper is limited to eight perspectives, which might introduce stereotypes or representational biases. Furthermore, the paper focuses on a single city and a single large language model; therefore, the generalization of the findings requires validation across more cities and models. Future research could extend this approach to other cities, models, or settings, deepen the analysis, and compare AI-generated descriptions with human descriptions to better understand how generative systems reproduce, stabilize, or transform urban regions.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Abbasi, O.R., Welscher, F., Weinberger, G., Scholz, J.: The world as large language models see it: Exploring the reliability of llms in representing geographical features. arXiv preprint arXiv:2506.00203 (2025)
2. Bai, Z., Wu, Y., Kang, X., Kong, X., Zhang, J.: Using llms for the multidimensional perception assessment of recreation and leisure spaces: a case study of hangzhou, china. Computational Urban Science **6**(1), 9 (2026)
3. Bhandari, P., Anastasopoulos, A., Pfoser, D.: Are large language models geospatially knowledgeable? In: Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems. pp. 1–4 (2023)
4. Fu, X., Li, C., Quan, S.J., Yigitcanlar, T., Wasserman, D.: Large language models in urban planning. Nature Cities **2**(7), 585–592 (2025)
5. Janowicz, K., Mai, G., Huang, W., Zhu, R., Lao, N., Cai, L.: Geofm: how will geo-foundation models reshape spatial data science and geoai? International Journal of Geographical Information Science **39**(9), 1849–1865 (2025). https://doi.org/10.1080/13658816.2025.2543038, https://doi.org/10.1080/13658816.2025.2543038
6. Janowicz, K., Mai, G., Zhu, R., Gao, S., Wang, Z., Hu, Y., Bennett, L.: Geography according to chatgpt-how generative ai represents and reasons about geography. In: Geography According to Foundation Models, pp. 2–11. SAGE Publications 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP (2026)
7. Janowicz, K., McKenzie, G., Hu, Y., Zhu, R., Gao, S.: Using semantic signatures for social sensing in urban environments. In: Mobility patterns, big data and transport analytics, pp. 31–54. Elsevier (2019)
8. Karimi, M., Mesgari, M.S.: Extracting place functionality from crowdsourced textual data using semantic space modeling. IEEE Access **11**, 129217–129229 (2023)
9. Liu, Z., Janowicz, K., Majic, I., Shi, M., Fortacz, A., Karimi, M., Mai, G., Currier, K.: Operationalizing geographic diversity for the evaluation of ai-generated content. Transactions in GIS **29**(3), e70057 (2025)
10. Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., Lao, N.: On the opportunities and challenges of foundation models for GeoAI (vision paper) **10**(2), 1–46 (2024). https://doi.org/10.1145/3653070, https://dl.acm.org/doi/10.1145/3653070
11. Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Abdulhai, M., Faust, A., Matarić, M.: Personality traits in large language models (2023)
12. Shao, Z., Wu, J., Chen, W., Wang, X.: Personal travel solver: A preference-driven llm-solver system for travel planning. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 27622–27642 (2025)
13. Wood, L.: Perception studies in geography. Transactions of the Institute of British Geographers pp. 129–142 (1970)
14. Yang, A., Fu, C., Jia, Q., Dong, W., Ma, M., Chen, H., Yang, F., Wu, H.: Evaluating and enhancing spatial cognition abilities of large language models. International Journal of Geographical Information Science **39**(9), 2009–2044 (2025)