International Cartographic Association
Association Cartographique Internationale

Large Language Model-Based Classification and Mapping of Personality Traits from Geo-Social Media

Ridwan Wahed [a], Alexander Dunkel [b], Dirk Burghardt [a,*]

[a] *Institute of Cartraphy, Technische Universität Dresden, Germany; firstname.lastname@tu-dresden.de,*
[b] *Leibniz Institute of Ecological Urban and Regional Devlopment, a.lastname@ioer.de*

* Corresponding author

Keywords: personality mapping, geo-social media, Large Language Models, big five personality traits, spatial reliability maps

Abstract:

The rapid development of large language models (LLMs) has opened new opportunities for analyzing complex personality traits from unstructured social media text. Traditional personality assessment typically relies on resource-intensive surveys with limited spatial and temporal coverage, whereas the large volumes of geotagged social media data offer the potential for scalable and near real-time personality mapping. However, the effectiveness of LLMs for inferring nuanced psychological traits from geo-social media content and for visualizing their spatial distribution remains largely unexplored. Therefore, the research addresses three key research questions concerning the use of large language models (LLMs) for personality analysis from geo-social media data. First, it examines to what extent LLMs are effective and suitable for classifying complex human traits, such as the Big Five personality dimensions, from geotagged social media text, considering aspects such as accuracy, interpretability, transparency, and computational and energy efficiency. Second, it investigates which spatial patterns of the Big Five personality traits can be identified across Germany based on LLM-derived classifications. Third, it compares the resulting LLM-based personality maps with existing survey-based personality maps to explore how both approaches differ and what their similarities and differences reveal about their respective methodological strengths and limitations.

Large Language Models (LLMs) have significantly advanced the classification of personality traits from social media text. Studies such as Christian et al. (2021) show that combining pretrained transformer models can achieve high accuracy in predicting the Big Five traits. Research by Peters and Matz (2024) demonstrates that GPT-based zero-shot approaches can infer personality traits with moderate correlations to self-reported scores, though demographic biases may occur. Further work by Davidson and Chae (2025) and Liu and Shi (2025) highlights methodological improvements in LLM-based text classification, including prompt optimization and efficient model training strategies. In parallel, spatial studies by Ebert et al. (2022b) demonstrate that techniques such as distance weighting and spatial smoothing enable the cartographic representation of regional personality patterns.

The analyses in this research are based on geotagged social media dataset comprising approximately 70 million posts. This dataset was collected across four prominent platforms Twitter, Instagram, Flickr, and iNaturalist as described in Dunkel et al. (2023) and organized according to a common Location-Based Social Networks (LBSN) data structure[1]. The methodology follows a three-stage approach, including data preprocessing, LLM-based personality classification, and spatial visual analysis using heatmaps and uncertainty visualization. After language processing and quality filtering to remove bot-generated content and spam, the dataset was refined to 8.5 million high-quality posts from 70,000 users. The open-source LLM DeepSeek-R1-Distill-Qwen-32B was employed with optimized parameters (temperature = 0.1) and a zero-shot chain-of-thought prompting strategy to classify personality traits. The resulting classifications were spatially aggregated using distance-weighted smoothing techniques and visualized through continuous heatmaps combined with spatial reliability maps to capture both patterns and uncertainty (see Figure 1).

Results from the LLM-based classification showed optimal performance with structured prompting that included explicit scale definitions and justification requirements, achieving F1 scores of 0.85 and an accuracy of 0.85 for conscientiousness. Aggregating text at the user level proved superior to post-level classification, as it preserved the full distribution of personality values while avoiding central tendency bias.

[1] https://lbsn.vgiscience.org/lbsntransform/docs/

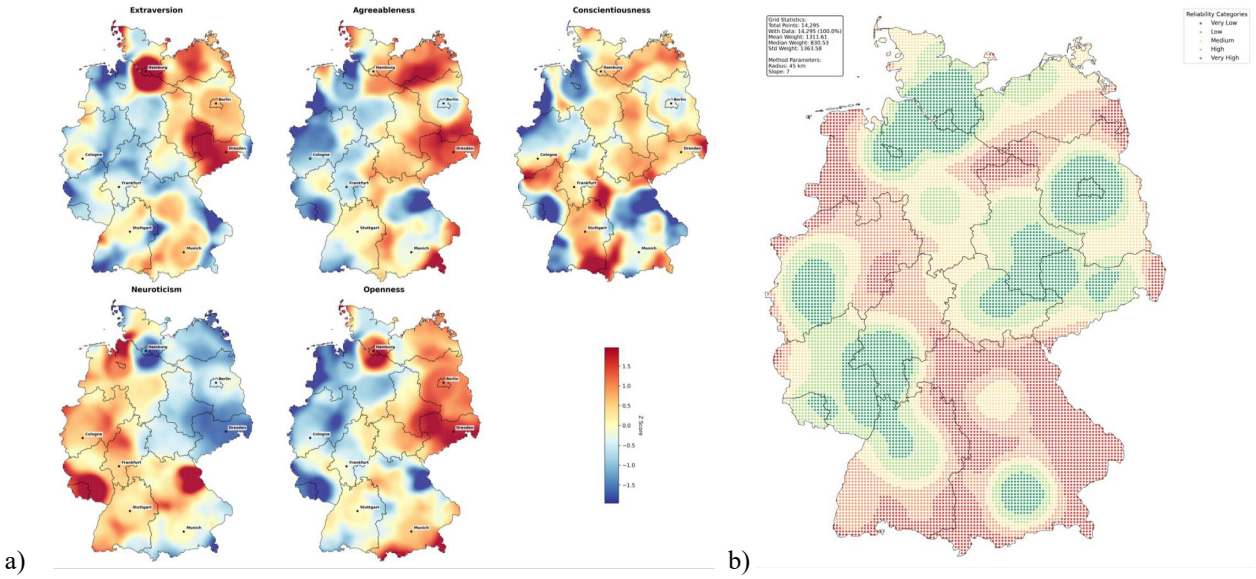

Figure 1. a) Spatial Distribution of Big Five Personality Traits across Germany; b) Spatial Reliability Map

Spatial analysis showed distinct personality clustering patterns that extended beyond administrative boundaries. Overall, 42.1% of the grid cells contained user data, and reliability was directly correlated with urban density. A comparison with existing survey-based personality maps (Ebert et al. 2022a) highlighted both notable consistencies and systematic deviations. The spatial distribution of conscientiousness closely matched traditional findings, particularly in southern German regions. Significant differences were observed for neuroticism: LLM-based maps indicated higher values in western urban centers, whereas survey data suggested high emotional stability in the same areas. These discrepancies likely reflect fundamental differences between behavior expressed on social media and self-reported, stable personality traits. This suggests that LLM-based approaches may capture the contextual expression of personality rather than underlying dispositional traits.

This study shows that LLM-based personality mapping represents a viable and complementary alternative to traditional survey methods, capturing distinct and informative aspects of personality expression across geographic regions. The approach provides a reproducible framework for LLM-based geo-social media analysis of personality while emphasizing the multifaceted nature of measuring human traits. The results advance both computational geographic psychology and the cartographic representation of complex psychological constructs.

References

Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021a). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. Journal of Big Data, 8(1), 68. https://doi.org/10.1186/s40537-021-00459-1

Davidson, T., & Chae, Y. (2025). Large Language Models for Text Classification: From Zero-Shot Learning to Instruction-Tuning. OSF. https://doi.org/10.31235/osf.io/sthwk_v2

Dunkel, A., Hartmann, M., Hauthal, E., Burghardt, D., & Purves, R. (2023). From sunrise to sunset: Exploring landscape preference through global reactions to ephemeral events captured in georeferenced social media. PLOS ONE, 18, e0280423. https://doi.org/10.1371/journal.pone.0280423

Ebert, T., Gebauer, Jochen. E., Brenner, T., Bleidorn, W., Gosling, S. D., Potter, J., & Rentfrow, P. J. (2022a). Are Regional Differences in Psychological Characteristics and Their Correlates Robust? Applying Spatial-Analysis Techniques to Examine Regional Variation in Personality. Perspectives on Psychological Science, 17(2), 407–441. https://doi.org/10.1177/1745691621998326

Ebert, T., Mewes, L., Götz, F. M., & Brenner, T. (2022b). Effective Maps, Easily Done: Visualizing Geo-Psychological Differences Using Distance Weights. Advances in Methods and Practices in Psychological Science, 5(3), 25152459221101816. https://doi.org/10.1177/25152459221101816

Liu, M., & Shi, G. (2025). Enhancing LLM-Based Text Classification in Political Science: Automatic Prompt Optimization and Dynamic Exemplar Selection for Few-Shot Learning. https://arxiv.org/html/2409.01466v2

Peters, H., & Matz, S. C. (2024). Large language models can infer psychological dispositions of social media users. PNAS Nexus, 3(6), pgae231. https://doi.org/10.1093/pnasnexus/pgae231