Integrating LLM-based workflows into GIS classroom learning activities

Peter Mooney[1], Paddy Gorry[2], and Levente Juhász[3]

[1]Department of Computer Science, Maynooth University, Ireland; peter.mooney@mu.ie
[2]Hamilton Institute, Center for Research Training in Data Analytics, Maynooth University, Ireland; patrick.gorry.2015@mumail.ie
[3]Geospatial Analytics, Technology and Open Research (GATOR) Lab, Fort Lauderdale Research and Education Center, University of Florida, USA; levente.juhasz@ufl.edu

Abstract

As Generative AI (GenAI) enters the GIS classroom, educators face an important question: can Large Language Models (LLMs) autonomously generate the synthetic spatial datasets required for some educational tasks? This study evaluates seven LLMs by comparing their point pattern generated outputs against procedural generation (RADIAN) and OpenStreetMap data using Cross-L and Cross-K functions. Findings reveal that while LLMs can generate realistic semantic attributes, they largely fail to produce usable spatial geometries. Most models generate unnatural distributions statistically segregated from ground-truth data. Relying on LLMs for raw spatial coordinate generation risks exposing students to flawed spatial patterns. We advocate for a hybrid workflow that combines procedural generators for accurate geometric foundations with LLMs for rich semantic attribute generation.

Keywords: LLM, chatbot, point pattern, GIS education, spatial analysis

1. Introduction and Motivation

Educators face a delicate balancing act regarding usage of Generative AI (GenAI) in the classroom. If not carefully scaffolded, GenAI usage can mask poor or shallow student understanding of key learning concepts. However, encouraging GenAI in the classroom can dramatically accelerate higher-order learning by offloading routine cognitive work (Chiu 2024). Overall, educators are "encouraged to incorporate GenAI tools into their teaching practices while remaining vigilant about potential risks" (Law 2024). In the GIS classroom this can take the form of avoiding routine, repetitive, and time-consuming data manipulation work often associated with spatial datasets. In our recent work, we have reported on our research into enhancing the learning experiences for

Published in "Proceedings of the 1st International Conference on Geospatial Artificial Intelligence (GeoAI 2026) – Oral Presentation Papers", edited by Haosheng Huang and Nico Van de Weghe, GeoAI 2026, 3-6 June 2026, Ghent, Belgium.

This contribution underwent single-blind peer review based on the extended abstract.

students in the GIS classroom. RADIAN (**RA**n**D**om spat**I**al d**A**ta ge**N**erator) (Gorry and Mooney 2024, 2025) describes the development of a fully reproducible software-based tool for the generation of synthetic geospatial datasets for classroom usage. Generating synthetic geospatial datasets reduces the overheads of searching for suitable datasets that may not always be available or may require significant time and technical resources for instructors to merge, integrate and prepare.

Figure 1 illustrates the overall concept of this work. Most educators spend a great deal of time and energy preparing datasets for student learning tasks in the classroom. This is the baseline situation. Students then use these datasets for learning tasks and generate outputs (such as map visualizations in GIS). The baseline situation is as Cobb (2020) calls it "technical and containing very real challenges within a field that is already highly labor-intensive". With RADIAN we successfully challenged the baseline by providing students with Python-based software to generate synthetic datasets for their learning tasks in the GIS classroom. This current work seeks to move beyond Phase 1 and incorporate GenAI into the overall workflow. LLMs/chatbots could potentially be used by students to generate spatial datasets for their learning tasks. This would: remove the need for specialist software, support learning-by-doing, and iterative experimentation. It also allows students to become more familiar with prompting and GenAI interactions. In a more applied sense using GenAI in this context can help students develop critical skills such as prompt formulation, verification of outputs, and ethical use of AI all of which prepares them for real-world GIS workflows rather than obstructing them from technologies they will encounter in real-world practice. Consequently, the core objective of this research is to evaluate whether LLMs can generate geospatial datasets that would be usable in the classroom, potentially augmenting the existing functionality of RADIAN.

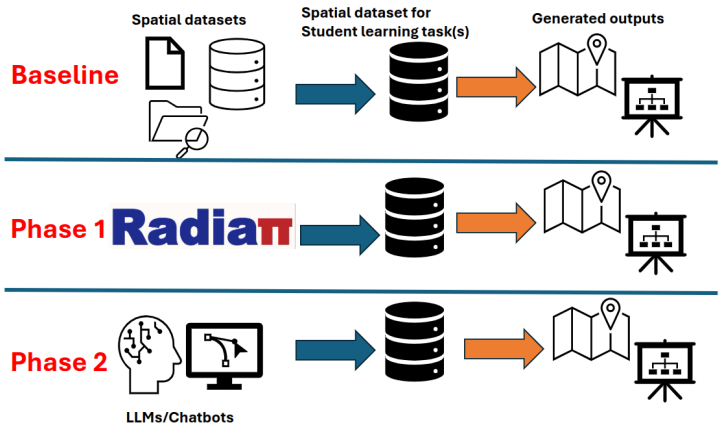

Figure 1: Integration of LLMs/Chatbots (Phase 2) into GIS classroom teaching and learning activities expanding on previous work done on development of automated approaches for Phase 1.

Previous Work

This research builds upon an existing body of work we have carried out on the emergence of GenAI in the classroom. We have investigated the capabilities of ChatGPT for taking basic GIS examinations and assessments (Mooney et al. 2023). Hochmair, Juhász, and Kemp (2024) showed that chatbots generally performed well on tasks related to spatial literacy, GIS theory, and interpreta-

tion of programming code and functions, but the chatbots revealed weaknesses in mapping, code writing, and spatial reasoning. In GIS programming classes, recent empirical evidence suggests that chatbots are increasingly serving as partial substitutes for traditional help resources, such as external websites or instructor feedback, particularly as the complexity of GIS programming tasks increases (Hochmair 2025b). Another research across geomatics sub-disciplines shows that students in advanced geospatial analysis courses tended to provide lower ratings for chatbot responses than those in quantitative measurement courses, likely due to higher performance expectations for complex spatial tasks, and the use of multimodal inputs such as image-enhanced prompts encourage exploratory behavior and more creative student engagement with course content (Hochmair 2025a). LLMs have also been evaluated as a mapping assistant for enhancing the efficiency of collaborative mapping (Juhász et al. 2023). As shown in Figure 1 our goal is to use LLMs/Chatbots to perform the task of generating synthetic spatial datasets that can be used by students in our GIS classes. Long et al. (2024) argue that given the recent advancements in LLMs the "synthetic data produced by LLMs emerges as a viable alternative or supplement to human-generated data". Wei et al. (2025) proposed a trainable method to enable LLMs to master GIS tools including understanding the functions of GIS tools. Ning et al. (2025) outlined a framework, with an LLM-driven decision core, capable of selecting data sources from a predefined list and fetching data using source-specific handbooks that document metadata and data retrieval details. Wang et al. (2024) found that the roles that LLMs and GenAI models can play in GIS can be "categorized as a purifier, converter, generator, and reasoner".

2. Methodology

The analysis framework, including the exact prompts, model versions, raw responses, and analysis scripts are available on GitHub at https://github.com/paddeaux/geoai.

2.1. Study Design and Data Collection

To evaluate the capacity of LLMs to generate usable geospatial datasets, we designed a comparative prompt framework consisting of three tasks. Each task required the model to output a valid GeoJSON `FeatureCollection` of 500 point geometries (restaurants) within the WGS84 coordinate boundaries of two distinct study areas: London and Berlin. We chose these two major capitals to leverage the so-called 'rich-get-richer' effect in LLM data representation, ensuring the evaluation utilizes well-documented urban hubs where representational consistency is highest compared to other more underrepresented areas (Dudy et al. 2025). The tasks are defined as follows:

1. **Task A (Synthetic):** Generation of fictional but plausible restaurant locations;

2. **Task B (OSM-Derived):** Retrieval of real-world restaurant locations derived from open data such as OpenStreetMap (OSM); and

3. **Task C (Random Spatial Baseline):** Generation of points following complete spatial randomness (CSR) to serve as a null spatial model.

To evaluate the LLMs geographic outputs across these tasks, we structure our spatial analysis

around three primary Research Questions (RQs). **RQ1 (Internal Consistency)** investigates whether models possess dynamic geographic representations for "synthetic" versus "real" places or if they rely on rigid internal spatial assumptions. **RQ2 (Synthetic Baseline Comparison)** evaluates how the LLMs' imagined spatial distributions compare against procedurally generated administrative baselines. Finally, **RQ3 (Ground-Truth Accuracy)** assesses the models' capability to accurately reconstruct the true point pattern distributions of real-world OSM data. Data collection was executed programmatically using an automated Python pipeline to ensure reproducibility. Models were either hosted on HiPerGator or accessed through HiPerGator as a proxy. Seven state-of-the-art models (at the time of writing) were selected: `gpt-4.1`, `gpt-5.4`, `claude-4.7-opus`, `gemini-3.1-pro`, `gpt-oss-120b`, `gpt-oss-20b`, and `llama-3.1-8b-instruct`. A validation step was enforced post-generation, which manually fixed some of the outputs with minor errors (e.g. missing closing "}" in GeoJSON output). As mentioned previously, to enhance reproducibility all of the individual model outputs, input prompts, city boundaries, data processing, and visualisations are available on GitHub.

2.2. Analysis Methods

Our analysis employed a combination of descriptive spatial statistics and second-order point pattern functions. For baseline descriptive metrics, we calculated model adherence (success in formatting and parameter compliance) as the number of instructions followed (see Github), standard distance to measure the physical spread of the generated points, the Nearest Neighbor Index (NNI), and the Variance-to-Mean Ratio (VMR) via Quadrat Analysis to evaluate broader clustering behavior. To evaluate attribute agreement in Task B, we calculated an *OSM Grounding Score*, defined simply as the proportion of names in the generated dataset that matched real-world OpenStreetMap POI names in that city. To analyze the spatial distributions across varying geographic distances, we utilized two primary bivariate point pattern methods. First, a bivariate version of Besag (1977)'s L-function (Cross-L) was used to evaluate internal model consistency. By comparing the spatial clustering of Task A outputs against Task B outputs, this method detects the degree to which a model's generated spatial coordinates remain static, regardless of whether the prompt requests synthetic or real-world data. Second, Ripley (1977) Cross-K function was used to test ground-truth accuracy by characterizing the spatial dependence between the LLM-generated points and the true RADIAN points. For both functions, we employed Monte Carlo random labeling with 99 permutations to randomly assign points to each category, establishing 95% confidence envelopes to test for statistical significance.

3. Results

Our results highlight a contrast between an LLM's ability to generate realistic-sounding semantic data and its capacity to reason about two-dimensional geographical space. As visualized in the spatial distributions (Figure 2) and detailed in the descriptive metrics (Tables 1 and 2), only `claude-4.7-opus` provided structurally reasonable spatial data, and only for Tasks A and B. Furthermore, we observed that high formatting adherence scores can be misleading regarding spatial quality. For example, while `llama-3.1-8b-instruct` successfully followed formatting instructions to generate 500 points in Task A (Table 1), it placed all points directly on top of each other, resulting in a Standard Distance and NNI of 0.00. To test our core hypotheses regarding

Model	Adherence	Points (n)	Density (pts/km²)	Within Boundary (%)	NNI	Quadrat VMR	Std Distance (km)
`claude-4.7-opus`	5/6	552	0.35	100.00	0.22	34.08	4.67
`gemini-3.1-pro`	6/6	500	0.31	100.00	0.78	0.18	7.17
`gpt-5.4`	5/6	500	0.31	92.80	1.32	1.21	13.99
`gpt-oss-120b`	6/6	500	0.31	100.00	0.00	NaN	NaN
`llama-3.1-8b`	5/6	498	0.31	100.00	0.00	495.98	0.00
RADIAN	5/6	500	0.31	100.00	0.74	15.51	10.69

Table 1: Task A (Synthetic Data Generation)

Model	Adherence	Points (n)	Density (pts/km²)	Within Boundary (%)	NNI	Quadrat VMR	Std Distance (km)	OSM Grounding (%)	Unique Names
`claude-4.7-opus`	5/7	540	0.34	100.00	0.23	92.52	4.93	34.30	540
`gemini-3.1-pro`	7/7	500	0.31	100.00	0.78	0.17	8.14	96.00	50
`gpt-5.4`	5/7	610	0.38	100.00	0.04	30.61	3.68	25.70	610
`gpt-oss-120b`	5/7	500	0.31	54.00	2.10	0.18	22.37	19.80	500
`gpt-oss-20b`	5/7	480	0.30	56.20	2.06	0.17	22.04	95.00	20
`llama-3.1-8b`	6/7	585	0.37	100.00	0.02	580.97	0.51	100.00	3
RADIAN	6/7	500	0.31	100.00	0.60	31.37	9.83	98.00	471

Table 2: Task B (OSM-Derived)

these spatial failures across the higher-performing models, we framed our quantitative point pattern analysis around the three primary research questions described above.

Addressing **RQ1 (Internal Consistency)**, we evaluated whether each LLM utilizes different geographic representations for "synthetic" versus "real" places or if it relies on a spatial template. Using a bivariate Cross-L function to compare `claude-4.7-opus`'s Task A outputs against its Task B outputs, Figure 3 shows that the observed Cross-$L(r)$ curve runs within the simulated simulated the 95% Monte Carlo confidence envelope of the mean across all distance thresholds. This internal consistency indicates that the spatial patterns of Task A and Task B are statistically indistinguishable. However, none of the other models produced usable datasets consistently for these tasks. Examples in this results are provided for London for brevity.

For **RQ2 (Synthetic Baseline Comparison)**, we examined how an LLM-imagined synthetic city compares to a procedurally generated synthetic city by evaluating `claude-4.7-opus`'s Task A outputs against the RADIAN Task A baseline using the Cross-L function. Visually (Figure 2 `claude-4.7-opus` generated points more clustered than RADIAN ($NNI = 0.22$ vs. 0.74). The statistical envelope confirms this deviation (Figure 3) as the observed Cross-$L(r)$ curve significantly differs from the simulated mean. Finally, for **RQ3 (Ground-Truth Accuracy)**, we evaluated whether the LLMs could accurately reconstruct the true spatial distribution of real-world OpenStreetMap data and understand micro-level geographic context, but without giving them a tool to do so. Using Ripley's Cross-K function with 95% Monte Carlo random labeling envelopes, we tested `claude-4.7-opus` and `gpt-5.4`'s Task B outputs against the true 500-sample OSM dataset generated via RADIAN. At the macro-scale (0-3000m), we tested the null hypothesis that the LLM and OSM points share identical underlying distributions. As shown in Figure 3, the observed K for all models fell below the 95% envelope, suggesting segregation and indicating that LLM-generated points and true OSM points are significantly less co-located than originally expected.

Figure 2: Generated outputs for London Tasks A, B and C from RADIAN, `claude-4.7-opus` and `gpt-5.4`. All other outputs of other models are available in the GitHub repository

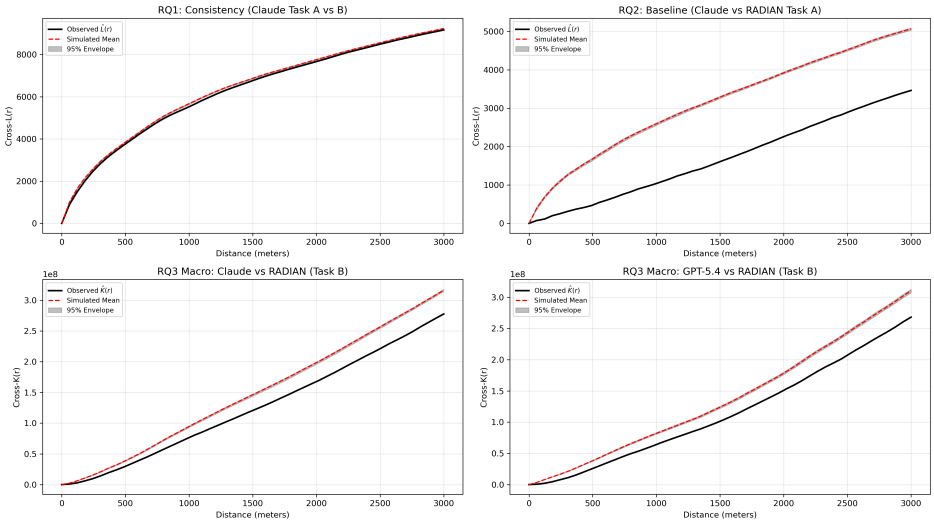

Figure 3: Calculated Cross $L(r)$ and $K(r)$ functions for London

4. Conclusions

Our work highlights a critical distinction between an LLM's capacity for semantic generation and its capability for generating usable geospatial datasets, at least with a zero-shot approach. While these modern models demonstrate a high degree of formatting adherence and can seamlessly generate realistic attribute data, this apparent competence often masks severe spatial failures. In this study, only `claude-4.7-opus`, a flagship model at the time of writing, was able to generate usable datasets for both Tasks A and B. As evidenced by our point pattern analyses, other models frequently suffer from isolated coordinate memorization, and a failure to replicate natural geographical distributions, even when the prompt instructions are followed. Consequently, there are significant pedagogical implications for integrating raw LLMs into the GIS classroom for these types of tasks. Relying purely on chatbots to generate raw spatial datasets risks exposing students to fundamentally flawed, statistically segregated, point patterns that do not reflect real-world geographic characteristics. While LLMs show promise as educational assistants and "converters" of data, procedural generators like RADIAN remain essential for establishing accurate, statistically robust geometric foundations. Future workflows in GIS education should consider a hybrid approach: utilizing robust procedural software for spatial coordinate generation, while leveraging the semantic capabilities of LLMs to populate those true geometries with rich, realistic attribute data.

Acknowledgments

This work has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI 18/CRT/6049.

References

Besag, Julian. 1977. "Contribution to the Discussion on Dr. Ripley's Paper." *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (2): 193–195.

Chiu, Thomas KF. 2024. "The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney." *Interactive learning environments* 32 (10): 6187–6203.

Cobb, Casey D. 2020. "Geospatial analysis: A new window into educational equity, access, and opportunity." *Review of Research in Education* 44 (1): 97–129.

Dudy, Shiran, Thulasi Tholeti, Resmi Ramachandranpillai, Muhammad Ali, Toby Jia-Jun Li, and Ricardo Baeza-Yates. 2025. "Unequal Opportunities: Examining the Bias in Geographical Recommendations by Large Language Models." In *Proceedings of the 30th International Conference on Intelligent User Interfaces,* 1499–1516. IUI '25. New York, NY, USA, March. https://doi.org/10.1145/3708359.3712111.

Gorry, Paddy, and Peter Mooney. 2024. "A software tool for generating synthetic spatial data for GIS-classroom usage." *AGILE: GIScience Series* 5:26. https://doi.org/10.5194/agile-giss-5-26-2024. https://agile-giss.copernicus.org/articles/5/26/2024/.

Gorry, Paddy, and Peter Mooney. 2025. "RADIAN – A tool for generating synthetic spatial data for use in teaching and learning." *Cartography and Geographic Information Science* 52 (3): 314–330. https://doi.org/10.1080/15230406.2024.2377981.

Hochmair, Hartwig H. 2025a. "Student chatbot use and perceptions in a course assignment: Comparing two geomatics courses." *Geomatica* 77, no. 2 (December): 100079. ISSN: 1195-1036. https://doi.org/10.1016/j.geomat.2025.100079.

———. 2025b. "Use and Effectiveness of Chatbots as Support Tools in GIS Programming Course Assignments" [in en]. *ISPRS International Journal of Geo-Information* 14, no. 4 (April): 156. ISSN: 2220-9964. https://doi.org/10.3390/ijgi14040156.

Hochmair, Hartwig H., Levente Juhász, and Takoda Kemp. 2024. "Correctness Comparison of ChatGPT-4, Gemini, Claude-3, and Copilot for Spatial Tasks" [in en]. *Transactions in GIS* 28 (7): 2219–2231. https://doi.org/10.1111/tgis.13233.

Juhász, Levente, Peter Mooney, Hartwig H Hochmair, and Boyuan Guan. 2023. "ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs." *arXiv preprint arXiv:2306.03204.*

Law, Locky. 2024. "Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review." *Computers and Education Open* 6:100174.

Long, Lin, Rui Wang, Ruixuan Xiao, Junbo Zhao, Xiao Ding, Gang Chen, and Haobo Wang. 2024. "On LLMs-driven synthetic data generation, curation, and evaluation: A survey." In *Findings of the Association for Computational Linguistics: ACL 2024,* 11065–11082.

Mooney, Peter, Wencong Cui, Boyuan Guan, and Levente Juhász. 2023. "Towards Understanding the Geospatial Skills of ChatGPT: Taking a Geographic Information Systems (GIS) Exam." In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery,* 85–94. GeoAI '23. New York, NY, USA, November. https://doi.org/10.1145/3615886.3627745.

Ning, Huan, Zhenlong Li, Temitope Akinboyewa, and M Naser Lessani. 2025. "An autonomous GIS agent framework for geospatial data retrieval." *International Journal of Digital Earth* 18 (1): 2458688.

Ripley, B. D. 1977. "Modelling Spatial Patterns." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (2): 172–212.

Wang, Siqin, Tao Hu, Huang Xiao, Yun Li, Ce Zhang, Huan Ning, Rui Zhu, Zhenlong Li, and Xinyue Ye. 2024. "GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review." *International Journal of Digital Earth* 17 (1): 2353122.

Wei, Cheng, Yifan Zhang, Xinru Zhao, Ziyi Zeng, Zhiyun Wang, Jianfeng Lin, Qingfeng Guan, and Wenhao Yu. 2025. "GeoTool-GPT: a trainable method for facilitating Large Language Models to master GIS tools." *International Journal of Geographical Information Science* 39 (4): 707–731.