RESEARCH IN PROGRESS

EXPLORING THE FEASIBILITY OF GENERATIVE AI IN ENHANCING THE IDENTIFICATION OF SPATIAL AND REGULATORY OPPORTUNITIES USING URBAN DIGITAL TWINS

DOI
https://doi.org/10.18690/um.fov.4.2025.47

ISBN
978-961-286-998-4

ROB PETERS,[1] FABIAN KOK,[2] JOHN VAN MEERTEN,[3]
KOEN SMIT[3]

[1] Province of Utrecht, Utrecht, the Netherlands
rob.peters@provincie-utrecht.nl
[2] HU University of Applied Sciences Utrecht, Data Science Department, Utrecht, the Netherlands
fabian.kok@hu.nl
[3] HU University of Applied Sciences Utrecht, Digital Ethics, Utrecht, the Netherlands
john.vanmeerten@hu.nl, koen.smit@hu.nl

This study explores the feasibility of using generative AI to enhance spatial and legislative opportunity finding within urban digital twins. Digital Twins (DTs) integrate real-time and historical data to provide comprehensive views of built environments, potentially aiding different disciplines involved in spatial planning practices. However, the complexity of legal frameworks and their visualization in DTs remains a challenge. Leveraging advancements in large language models (LLMs), this research investigates how multimodal AI can interpret complex legislative data to improve spatial planning. The study employs a Design Science Research (DSR) methodology, focusing on tuning existing LLMs with spatial content. Key findings include the successful generation of Geography Markup Language (GML) code, enhancing interoperability with spatial planning tools, and the iterative design process that improved the model's performance. Preliminary results indicate that a multimodal approach, including text, images, and GML code, significantly enhances the model's capability. Future research will focus on improving data quality, expanding multimodal capabilities, and evaluating real-world applications. This study contributes to the development of transparent, contestable, and explainable AI solutions for spatial planning.

Keywords:
generative AI,
LLMs,
spatial planning,
feasibility,
digital twins

University of Maribor Press

1 Introduction

The development of Digital Twins (DTs) for spatial planning offers potential advantages for policymakers, council members, GIS specialists, and other related disciplines. They enable the integration of real-time data, historical data, and predictive analytics to create a comprehensive view of built environments. Despite the fact that there is no consensus on a standardized or formal definition of a DT, most research views a DT as a cyber-physical system that shares the concepts of a physical entity, a virtual model and connections between both (Liu et al., 2022). In the context of this study, we adhere to the following definition of a DT (Boschert & Rosen, 2016): 'The DT is not one complete model of the physical product, but a set of linked operation data artefacts and (simulation) models, which are of suitable granularity for their intended purpose and stakeholders and evolve throughout the product life-cycle.' Yet, this definition does not fully comply with a true DT in the sense that it should support bi-directional communication (Fuller et al., 2020; Liu et al., 2022). Given this definition, there are still areas for improvement, one of which is the complexity of the underlying legal fundament and its visualization in a DT. Recent advancements in AI, especially the accessibility and potential multimodality of large language models (LLMs), present an opportunity to enhance spatial planning by interpreting complex legislative frameworks through various modalities. Multimodal AI models can process and integrate data from different sources, such as text, images, and spatial data, to provide more accurate and context-aware insights (Huang & Zhang, 2024; Ji & Gao, 2023; Sangeetha et al., 2024).

The research problem addressed in this paper is the necessity of explorative opportunity finding, particularly legislative opportunity finding, in urban spatial planning. This involves navigating numerous conflicting policy ambitions within limited space, such as parking, housing, mobility, recreation, and adhering to legislative norms like sound levels, heat stress, water safety, and emissions (Binnenlandsbestuur, 2011; Boer et al., 2007). The primary users and testers of this research are urban and provincial policy planners. They aim to identify specific areas for planning complex structures, such as windmills and solar panel fields, for which they currently use traditional GIS overlay methods to explore and pinpoint opportunity zones. To study the potential of LLM-technology in this context, we address the following research question in this paper: To what extent can a large language model, tuned with spatial content, assist actors in answering the "where can I do what?" question in urban planning?

The where-can-I-do-what capability was expressed as an important need in 2004 to decrease the administrative burden for SME's seeking permission for business activities, which was later described again by the Ministry of Internal Affairs (Binnenlands Bestuur, 2011). It has a long history of rule-based capability research (Peters & van Engers, 2004) and we investigate the additional power of generative AI to enhance this capability. Given the aim to develop solutions for governmental legislative processes, transparency, contestability, and explainability are crucial (Dittmar, 2024; Liao & Vaughan, 2024). This paper outlines the initial configuration of our solution, preliminary findings, and the iterative design challenges encountered.

2 Background and related work

Large language models have the potential to aid spatial planning combined into Digital Twins. Recent research suggests that fine-tuned LLMs can support urban design by integrating city-specific data. For example, specialized models like CityGPT use the "CityInstruction" dataset, which incorporates geographical information, allowing them to tackle urban planning tasks effectively. These models can even perform at a level comparable to proprietary solutions on city-focused benchmarks (Bentley et al., 2024; Choi & Yoon, 2025; Feng et al., 2024). Another example is UrbanLLM, a customized model designed for city management. It breaks down complex planning questions into smaller tasks and coordinates specialized tools to handle them. This approach allows UrbanLLM to perform significantly better than general LLMs, like the GPT series and LLaMA, when tackling complex urban planning challenges (Jiang et al., 2024).

Multi-agent systems are also becoming more popular. In one study, researchers combined LLM ‘agents’ with a city's digital platform and knowledge base to answer questions about urban policies and services. The system efficiently directs queries to GIS data, legal documents, and AI models, significantly reducing planning time from days to just a few hours while maintaining high accuracy (Kalyuzhnaya et al., 2025). Furthermore, a study by Ji and Gao (2023) examined how LLMs, such as GPT-2 and BERT, process spatial data. By encoding geometries in the Well-Known Text

(WKT) format, these models were tested on identifying geometry types and spatial relations. Results showed they maintained geometry types and captured spatial relations (73% accuracy) but struggled with numerical estimates and object retrieval.

Taking these recent developments into account, we argue that further research into the application of domain-specific LLMs in geospatial tasks is needed, as the current knowledge base is rather immature and few studies focus on applying the fundamental knowledge on LLMs and DT's for spatial planning.

3 Research method

This study employs Design Science Research (DSR) methodology, involving multiple iterations of the model. DSR is a research paradigm that focuses on the creation and evaluation of artifacts designed to solve identified problems. This study, in its current phase, can best be positioned in the relevance cycle (Hevner et al., 2004, figure 2). The design choices in this study are guided by the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and CODIO (Code Goed Openbaar Bestuur) implementation guidelines (Government of the Netherlands, 2022; Meijer, 2023). For tuning of the LLM model we used the content of all Dutch spatial plans 2012-2022 of all cities at www.ruimelijkeplannen.nl. To test the feasibility of the research approach and the capabilities of the LLM, we investigated three opportunity exploration maps (wind energy, solar power field and forest planting), see e.g., (Provincie Utrecht, 2024), together with their policy makers and technical designers to set the capability benchmark. This focus group of policy practitioners were familiar with the legalistic issues, the policy ambitions, the area constraints and the (end-)user needs such as from city planners. By directly including the GIS designers in the focus group who had built the traditional application we had access to the work process and content variables.

Key design decisions included:

- Using a large dataset of spatial content: The assumption was that a larger dataset would improve the model's ability to generalize and provide accurate insights. However, the downside was that the content was outdated, which posed challenges for data cleaning and preprocessing.

- Employing open-source LLM technology: For reasons of transparency, only open-source tools were used. This choice limited the processing power available, as commercial models and tools generally offer better performance.
- Focusing on tuning rather than training: Due to resource constraints, the study focused on tuning existing language models rather than training new ones from scratch. This approach was expected to limit the feasibility study's scope but was necessary given the available resources.
- Triangulation approach: To improve tuning efficiency, a triangulation approach was used, involving spatial, textual, and pictorial data. The assumption was that consistent data organization would enhance the tuning process. However, the quality of the pre-training data was less than expected, which posed additional challenges.

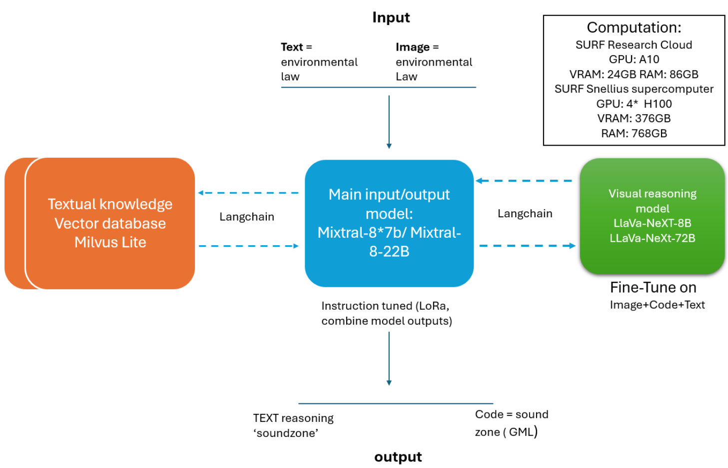

Figure 1: Architectural overview of the AI4Spatial solution

The feasibility study involved several steps, including:

- Testing picture recognition: Initial tests involved using LLava for picture recognition, such as identifying landmarks like the Utrecht Dom tower.
- Increasing content filtering efforts: Efforts were made to filter and clean the legacy GML-coded (Geography Markup Language) spatial plans to improve the quality of the input data.

- Optimizing configuration for multimodal cross-referencing: The configuration was optimized to handle cross-referencing between different types of data, such as text, images, and GML code.
- Upgrading configuration performance: The performance of the configuration was upgraded using the Dutch SURF Snellius supercomputer platform.
- Shifting from textual retrieval to GML code generation: The focus shifted from generating legal text to generating GML code, which is more useful for spatial planning tools and visualizations.
- Mixtral Model upgrade to Pixtral: The model was upgraded to Pixtral, eliminating the need for LLava and improving the handling of multimodal content.
- Testing with specific norms: Tests were conducted with specific norms, such as sound levels, to evaluate the model's ability to generate meaningful GML code.
- Prompt engineering: Prompt engineering techniques were used to filter the output and improve the relevance of the generated content. Examples were the inclusion of names of locations such as Utrecht or objects such as the Dom Tower in Utrecht in the prompt.

3 Preliminary results

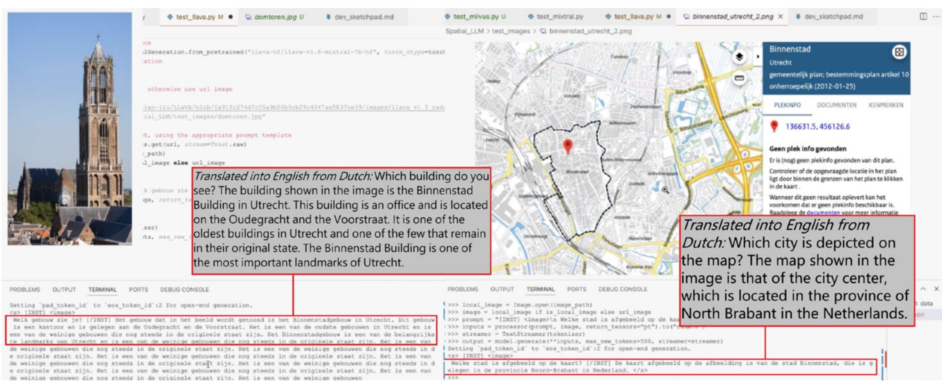

Figure 2: Erroneous results (hallucination) of the first iteration, where the computer does not recognize the right province (Brabant instead of Utrecht and the model labels the object as 'binnenstad' (inner city), rather than the Dom Tower

The initial setup faced challenges, particularly with text-based retrieval depending heavily on direct content feeds. The model generated answers to prompts, but the quality of the output was limited by the quality of the input data. Subsequent iterations focused on generating meaningful GML code rather than generating legal text, enhancing interoperability with existing spatial planning tools.

Key Findings

- The first iteration generated answers to prompts but required significant content filtering. The reliance on direct pre-filtered content feeds in the prompts limits the model's ability to generalize and provide accurate insights.
- The second iteration demonstrated the feasibility of generating GML code, improving integration with spatial planning tools. The model was able to generate GML code that represented meaningful contours for specific norms, such as sound levels in housing and business areas, but it was yet unable to put those polygons into the context of a spatial plan.
- The third iteration (ongoing) involves semi-automated content cleaning and a method of image-binding, enhancing the model's capability to handle multimodal content. The use of image binding allowed for better cross-referencing between different types of data (code, image and juridical text), improving the overall performance of the model.

Technical Findings

The second iteration of the feasibility test provided proof of the concept of GML code generation. This outcome would enable better feeds of spatial planning content and better integration with spatial planning tools such as Tygron or ESRI and environmental permit processing tools. Figure 3 illustrates the GML coding capability as output of the large language model configuration.

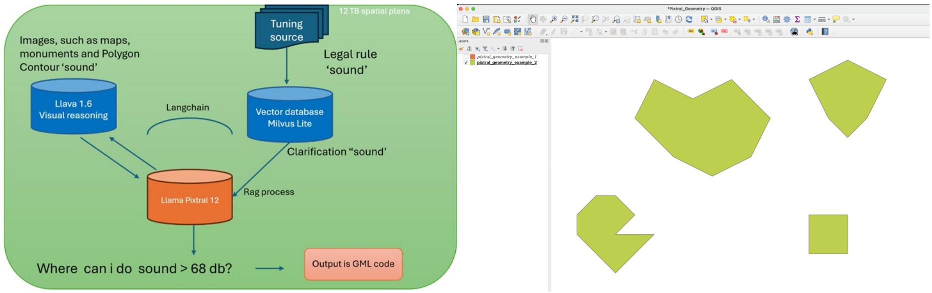

Figure 3: Example of the creation of the 'sound' polygon in GML code

During the testing of the second configuration, it was discovered that the Pixtral-based setup was not capable of handling tri-modal embedding of content triplets, such as satellite image/legal text/GML code or polygon/legal text/GML code, without significant additional scripting. Therefore, two additional techniques were applied: semi-automated content cleaning (iteration two) and image binding (iteration three).

4 Conclusions and Future Research

The feasibility study is ongoing, with promising results so far. However, several design challenges were encountered, such as the initial focus on text-based standards and the need for a multimodal approach. Future research will continue to refine the model and the use case process description such as 'finding windmill space' or 'solar panel farming area' or 'power net congestion optimization' with domain experts. We will also test with access to substantial geodata of the province of Utrecht (Open Geo Data, 2024) that was used to create the environmental legislative framework for this region (Omgevingsvisie Provincie Utrecht, 2024).

Key Conclusions

- Initial Setup Challenges: The initial setup failed to deliver the expected results due to the focus on text-based standards rather than geo-based standards. The shift to GML code generation improved the model's performance.
- Multimodal Approach: The second setup required a multimodal approach, which was not fully supported by the initial models. The upgrade to Pixtral

and the use of image binding improved the model's ability to handle multimodal content.

- Iterative Design Process: The iterative design process allowed for continuous improvement of the model, addressing the challenges encountered in each iteration. The focus on generating meaningful GML code rather than legal text enhanced the model's relevance for spatial planning applications.

Future Research Directions

- Enhancing Data Quality: Improving the quality of the input data through better content filtering and preprocessing techniques. This will involve collaboration with planning experts to ensure the relevance and accuracy of the data.
- Expanding Multimodal Capabilities: Further developing the model's ability to handle multimodal content, including the integration of additional data types such as satellite imagery and real-time sensor data.
- Improving Explainability and Transparency: Enhancing the explainability and transparency of the model's outputs to ensure trust and acceptance by policymakers and other stakeholders. This will involve developing methods for generating transparent justifications for the model's decisions.
- Scaling Up the Model: Addressing the computational constraints of open-source models by exploring ways to scale up the model's capabilities. This may involve leveraging cloud computing resources and optimizing the model's architecture for better performance.
- Evaluating Real-World Applications: Conducting pilot studies and real-world evaluations of the model's performance in various urban planning scenarios. This will provide valuable feedback for further refinement and validation of the model.

Acknowledgements

We would like to thank Luc de Horde and Bregtje Rijerman of the province of Utrecht for their help with this project. This project has been made possible by the Dutch Responsible Applied Artificial Intelligence programme.

References

Bentley, P. J., Lim, S. L., Mathur, R., & Narang, S. (2024). Automated Real-World Sustainability Data Generation from Images of Buildings.

Binnenlandsbestuur. (2011). Geozet viewer zet overheidsbekendmakingen op kaart. https://www.binnenlandsbestuur.nl/digitaal/overheidsbekendmakingen-op-interactieve-kaart-door-geozet-viewer

Boer, A., van Engers, T., Peters, R., & Winkels, R. (2007). Separating law from Geography in GIS-based eGovernment services. Artificial Intelligence and Law, 15(1), 49–76. https://doi.org/10.1007/s10506-007-9042-4

Boschert, S., & Rosen, R. (2016). Digital Twin—The Simulation Aspect. In Mechatronic Futures (pp. 59–74). Springer International Publishing. https://doi.org/10.1007/978-3-319-32156-1_5

Choi, S., & Yoon, S. (2025). AI Agent-Based Intelligent Urban Digital Twin (I-UDT): Concept, Methodology, and Case Studies. Smart Cities, 8(1), 28. https://doi.org/10.3390/smartcities8010028

Dittmar, L. (2024). The Explainability Challenge of Generative AI and LLMs - OCEG. https://www.oceg.org/the-explainability-challenge-of-generative-ai-and-llms/

Feng, J., Du, Y., Liu, T., Guo, S., Lin, Y., & Li, Y. (2024). CityGPT: Empowering Urban Spatial Cognition of Large Language Models.

Fuller, A., Fan, Z., Day, C., & Barlowc C. (2020). Digital Twin: Enabling Technologies, Challengesand Open Research. IEEE Access, 8, 952–971.

Kalyuzhnaya, A., Mityagin, S., Lutsenko, E., Getmanov, A., Aksenkin, Y., Fatkhiev, K., Fedorin, K., Nikitin, N. O., Chichkova, N., Vorona, V., & Boukhanovsky, A. (2025). LLM Agents for Smart City Management: Enhancing Decision Support Through Multi-Agent AI Systems. Smart Cities 2025, Vol. 8, Page 19, 8(1), 19. https://doi.org/10.3390/SMARTCITIES8010019

Government of the Netherlands. (2022). Impact Assessment Fundamental Rights and Algorithms . https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. Design Science in IS Research MIS Quarterly, 28(1), 75.

Huang, J., & Zhang, J. (2024). A Survey on Evaluation of Multimodal Large Language Models.

Ji, Y., & Gao, S. (2023). Evaluating the Effectiveness of Large Language Models in Representing Textual Descriptions of Geometry and Spatial Relations.

Jiang, Y., Chao, Q., Chen, Y., Li, X., Liu, S., & Cong, G. (2024). UrbanLLM: Autonomous Urban Activity Planning and Management with Large Language Models. Findings of the Association for Computational Linguistics: EMNLP 2024, 1810–1825. https://doi.org/10.18653/V1/2024.FINDINGS-EMNLP.98

Liao, Q. V., & Vaughan, J. W. (2024). AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. Harvard Data Science Review, (Special Issue 5). https://doi.org/10.1162/99608F92.8036D03B

Liu, Y. K., Ong, S. K., & Nee, A. Y. C. (2022). State-of-the-art survey on digital twin implementations. Advances in Manufacturing, 10(1), 1–23.

Meijer, A. (2023). Code Digitale Overheid kan vertrouwen in goed bestuur waarborgen - Nieuws - Universiteit Utrecht. https://www.uu.nl/nieuws/code-digitale-overheid-kan-vertrouwen-in-goed-bestuur-waarborgen

Omgevingsvisie provincie Utrecht. (2024). Plannenviewer Provincie Utrecht. https://ruimtelijkeplannen.provincie-utrecht.nl/NL.IMRO.9926.2020OWVISIE-VA01

Open Geo Data. (2024). Geo-Point Utrecht. https://geo-point.provincie-utrecht.nl/pages/open-data

Sangeetha, S. K. B., Mathivanan, S. K., Rajadurai, H., Cho, J., & Easwaramoorthy, S. V. (2024). A multi-modal geospatial–temporal LSTM based deep learning framework for predictive modeling of urban mobility patterns. Scientific Reports, 14(1), 31579. https://doi.org/10.1038/s41598-024-74237-3