

The Role of AI in Transforming Education in Spatial Data Science

Wu CHEN ¹, Xintao LIU ¹, Zhongyu LAI ¹

¹ Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Rapid advancements in Artificial Intelligence (AI) are revolutionising education, particularly within Geomatics fields such as land surveying, remote sensing, and Geographic Information Systems (GIS). This work explores the integration of AI tools into educational frameworks to enhance spatial understanding, data analysis, and immersive learning. We introduce **GeoAI Mentor**, an intelligent tutoring system powered by the **GeoDataGPT** engine, which integrates Large Language Models (LLMs) with dynamic knowledge graphs and Retrieval-Augmented Generation (RAG). Functioning as a 24/7 intelligent partner, the system provides personalised guidance, equipping students with essential skills in geospatial processing and predictive modelling. Pilot evaluations using Hong Kong scenarios demonstrate that **GeoAI Mentor** effectively dismantles technical barriers by achieving high-precision semantic retrieval and automating complex multi-step spatial reasoning. By bypassing the steep learning curve of software syntax, this approach reallocates cognitive resources toward higher-order analytical intent, preparing graduates for the complexities of modern geospatial industries and the demands of the digital era.

Keywords: Geomatics, Generative Artificial Intelligence, Education

1. Introduction

The digital transformation of the geoinformation industry demands professionals proficient in spatial data science (Tu et al., 2021). However, education in this field faces a significant barrier: the steep learning curve associated with professional GIS software and programming languages. Learners often struggle with complex operational details and syntax debugging, diverting their attention from the underlying logic of spatial analysis (Zhu et al., 2021). This creates a “technical barrier” where students may master software tools but lack the ability to apply spatial thinking to solve real-world problems.

Artificial Intelligence (AI), particularly Large Language Models (LLMs), offers a transformative solution by enabling natural language interaction with

complex data systems (Rudolph et al., 2023). While AI has been explored for isolated geomatics tasks, current tools often function merely as general coding assistants rather than “spatial-aware” mentors. They lack the domain-specific pedagogical frameworks necessary to guide students through the complete data-to-insight workflow, failing to effectively bridge the gap between theoretical concepts and technical implementation (Li & Ning, 2023).

To address this gap, this study introduces **GeoAI Mentor**, an intelligent tutoring system designed for the Hong Kong Common Spatial Data Infrastructure (CSDI). Built upon the **GeoDataGPT** core engine, this platform integrates a dynamic knowledge graph and adaptive learning mechanisms. By functioning as a 24/7 intelligent partner, **GeoAI Mentor** democratises access to spatial analytics, transforming spatial data science education from a tool-centric to a logic-centric model and equipping students with the critical skills needed in the digital era.

2. Methodology

This study employs a technical development and educational integration approach to construct **GeoAI Mentor**, a comprehensive intelligent tutoring system for geomatics education. The system architecture is hierarchical: it utilises **GeoDataGPT** as the underlying computational engine for spatial data processing, while integrating a Dynamic Domain Knowledge Base and an Adaptive Learning Module to fulfil pedagogical objectives.

2.1 Core AI Framework: GeoDataGPT for Spatial Intelligence

The **GeoDataGPT** framework is an AI-driven system designed to democratise geospatial data analysis and enhance learning outcomes in geoinformation education. Centred on the integration of an LLM with domain-specific knowledge augmentation, the framework translates natural language queries into executable geospatial workflows. This reduces the reliance on specialised programming languages and GIS tool proficiency.

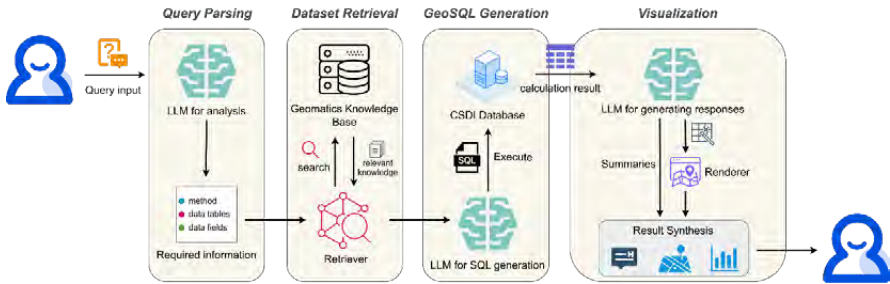
As illustrated in Figure 1, **GeoDataGPT** integrates four core modules to achieve context-aware spatial intelligence, which are operationalised through two primary functional phases: **Semantic Parsing and Knowledge Retrieval**, followed by **Computational Execution and Visualisation**.

- **Semantic Parsing and Knowledge Retrieval:** The system utilises the reasoning capabilities of the DeepSeek-R1 LLM agent to decompose user queries into structured sub-tasks. To ensure domain-specific accuracy, a Retrieval-Augmented Generation (RAG) engine employs BGE-M3-Embedding to perform semantic searches across a geospatial

knowledge base containing over 1,000 CSDI dataset schemas and 500 annotated educational Q&A records.

- **Computational Execution and Visualisation:** By synthesising user intent with retrieved metadata, the LLM agent generates validated GeoSQL scripts. These scripts undergo iterative refinement via a rule-based validator to ensure compliance with CSDI schema constraints, such as coordinate system compatibility, before execution in a spatial database. Finally, results are transformed into multi-modal outputs, including natural language explanations and interactive spatial visualisations rendered through ArcGIS JS, to help students connect results with geospatial theories.

Figure 1: Technology architecture of the GeoDataGPT.



Note: The architecture of the GeoDataGPT framework consists of four sequential modules. The system first parses a user’s query and enriches it with a knowledge retriever. An LLM then generates an executable GeoSQL script, and a final module synthesises the database results into summaries and visualisations.

2.2 GeoAI Mentor

To enhance geomatics education, we are exploring the expansion of the **GeoDataGPT** framework into a comprehensive educational platform, “**GeoAI Mentor**”, by integrating custom teaching modules (as shown in Figure 2). It will serve as a virtual mentor for geomatics students, answering their questions about the field, guiding them through data analysis, and providing personalised learning paths. This section details the proposed educational extension functions, focusing on modules beyond the inherited AI infrastructure.

Dynamic Domain Knowledge Base

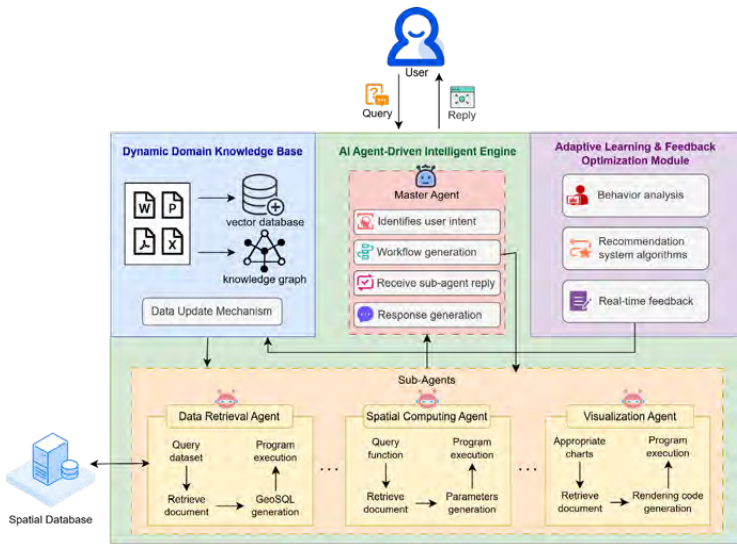
To link theoretical knowledge with practical applications, **GeoAI Mentor** introduces a semantic knowledge graph that organises geospatial concepts into an interactive, explorable network. Built using Neo4j, the graph integrates existing educational resources (i.e., textbooks, publications, lecture slides, exam papers) to link entities (e.g., “Coordinate System”) with hierarchical subtopics

(e.g., “HK1980 Grid vs. WGS84”) and cross-domain applications (e.g., “GNSS in building monitoring”). Unstructured content (e.g., research articles, textbooks) is encoded into semantic vectors using the BGE-M3 model, enabling context-aware retrieval through similarity searches. An automated update process periodically ingests the latest academic publications via APIs and incorporates educator-curated case studies, ensuring knowledge currency. This dynamic infrastructure transforms static textbooks into an evolving resource that adapts to technological advancements.

Adaptive Learning and Feedback Optimisation

GeoAI Mentor employs a machine learning-driven system to optimise individual learning trajectories. By analysing interaction patterns such as query frequency and task completion rates, the platform identifies specific knowledge gaps to construct detailed student profiles. A Long Short-Term Memory (LSTM) network is utilised to predict learning trajectories, allowing for dynamic resource allocation and content delivery strategies that align with Cognitive Load Theory. This creates a closed-loop optimisation process where real-time student feedback fine-tunes the system’s recommendations.

Figure 2: Technology architecture of the GeoAI Mentor.



Note: GeoAI Mentor consists of three functional modules. The AI Agent-Driven Intelligent Engine, also known as GeoDataGPT, is the core of the system. The Dynamic Knowledge Base stores mapping domain knowledge and provides a reliable knowledge reference for the LLM. The Adaptive Learning and Feedback Optimisation Module provides personalised learning routes and optimises the knowledge base based on student feedback.

3. Results and Discussion

To assess the pedagogical efficacy of **GeoAI Mentor** prior to large-scale classroom deployment, we conducted a systematic pilot evaluation using real-world scenarios from the Hong Kong CSDI. This assessment focuses on how the system resolves three foundational learning barriers in geomatics education: data discovery, analytical execution, and complex spatial reasoning.

3.1 Evaluation of Educational Outcomes and Barrier Mitigation

Mitigating the Data Discovery Barrier

The first formidable obstacle in spatial data science education is the “technical barrier” of identifying appropriate datasets among thousands of options. In our pilot test regarding “road network density”, the system successfully bypassed the limitations of traditional keyword-based searches by demonstrating high precision in semantic retrieval. Unlike standard tools, the knowledge augmentation engine mapped the user’s high-level intent to specific CSDI schemas, identifying the ‘Road Network’ and ‘District Boundary’ datasets. Crucially, by providing a validity assessment (see Figure 3a) and advising on data limitations, the system functions as a digital scaffold, mimicking an expert instructor’s guidance to ensure students understand data quality before analysis.

Bypassing Technical Operations to Accelerate Insight

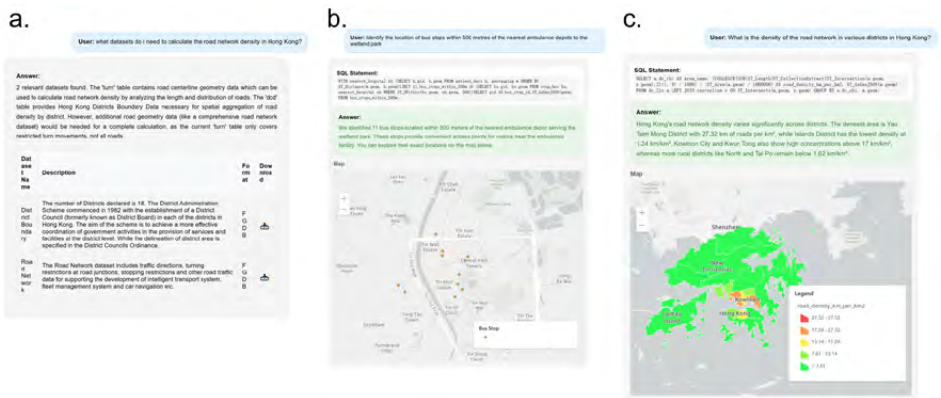
Moving to quantitative analysis, the system effectively abstracted complex technical operations that typically divert student attention from spatial logic. When tasked with calculating road density across districts, the system automated the spatial aggregation and normalisation process (km/km^2), bypassing the steep learning curve of GIS software and programming languages. As shown in Figure 3b, the output was not merely a data table but an instantly rendered choropleth map. This “Code-to-Visualisation” capability significantly reduced the “time-to-insight” and enabled learners to immediately allocate their cognitive resources to identifying and interpreting spatial patterns (e.g., the high density in Yau Tsim Mong) rather than getting bogged down in procedural debugging.

Scaffolding Multi-step Spatial Reasoning

The most significant finding is the system’s ability to structure chained logic for novices, a skill that typically requires years of practice. The scenario “finding bus stops within 500m of ambulance depots near a wetland park” (Figure 3c) requires a complex sequential workflow. By utilising Common Table Expressions, the system successfully translated a high-level cognitive problem into a structured computational workflow. This confirms that **GeoAI Mentor**

provides the necessary ‘Chain-of-Thought’ scaffolding, allowing students to master complex spatial reasoning.

Figure 3: Demonstration of GeoAI Mentor capabilities.



Note: The user’s request in the blue block was input into the system, which deconstructed the complex query and generated the feedback in the grey block. This output includes the sophisticated SQL for the chained analysis, a text summary of the findings, and an interactive map visualising the results.

3.2 Pedagogical Impact and Limitations

The progression from data retrieval to complex reasoning illustrates a substantial reduction in extraneous cognitive load (Sweller, 2010). By automating the procedural “how”, the system reallocates student working memory to the “why”, aligning with Cognitive Load Theory by focusing on germane load. This pilot success validates a shift toward prioritising logical spatial reasoning over rote software proficiency (Bearman et al., 2016; Zhu et al., 2021). Students are challenged to move beyond the “Remembering” level of Bloom’s Taxonomy, focusing instead on the “Analyzing and Evaluating” levels by articulating logical spatial questions (Anderson et al., 2001). However, reliance on AI introduces new challenges. The accuracy of the system is strictly bound by the quality of the CSDI metadata; ambiguous data descriptions can lead to retrieval failures. Furthermore, the risk of “automation bias” necessitates that “AI verification” becomes a core competency in the modern geomatics curriculum, transforming students into critical auditors of AI-generated spatial data (Mollick & Mollick, 2023).

4. Implications

The emergence of **GeoAI Mentor** signifies more than just a technological milestone; it heralds a fundamental paradigm shift in the pedagogical landscape of spatial data science by dismantling the technical barrier that has long teth-

ered students to the minutiae of software syntax and procedural operations. In this redefined educational ecosystem, the role of the learner is radically transformed: students are no longer passive consumers of GIS tools but are empowered as architects of inquiry who utilise AI as a cognitive scaffold to engage in higher-order problem-solving and “dialogic geospatial analysis”.

This transition necessitates a comprehensive overhaul of our training directions, prioritising new literacies such as AI verification, cartographic ethics, and the ability to critically synthesise AI-generated outputs with grounded geographical reality. Consequently, the societal demand for professional talent is shifting away from the traditional GIS technician toward the spatial data strategist—individuals who possess the “Spatial-AI Hybridity” required to navigate complex data infrastructures and address the wicked problems of the 21st century.

Ultimately, the democratisation of spatial intelligence through such mentors ensures that geographic analysis becomes a universal language, empowering a more informed citizenry to leverage spatial data as a strategic asset for social and environmental resilience.

Ethical Declaration Statement

The project has been approved by the Department of Land Surveying and Geo-Informatics of The Hong Kong Polytechnic University (Reference Number: 88JT).

Generative AI Policy

During the preparation of this work, the authors used “DeepL” and “ChatGPT” to improve the writing style. After utilising the tool/service, the authors thoroughly reviewed and edited the content as necessary and assumed full responsibility for the publication’s content.

References

- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives* (Complete ed.). Longman.
- Bearman, N., Jones, N., André, I., Cachinho, H. A., & DeMers, M. (2016). The future role of GIS education in creating critical spatial thinkers. *Journal of Geography in Higher Education*, 40(3), 394–406. <https://doi.org/10.1080/03098265.2016.1144729>
- Li, Z., & Ning, H. (2023). Autonomous GIS: the next-generation AI-powered

- GIS. *International Journal of Digital Earth*, 16(2), 4668–4686. <https://doi.org/10.1080/17538947.2023.2278895>
- Mollick, E., & Mollick, L. (2023). *Assigning AI: Seven Approaches for Students, with Prompts*. <https://doi.org/10.48550/arxiv.2306.10052>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, 22(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Tu, W., Xia, J., Wang, C., Lu, M., & Yue, Y. (2021). Exploration and Practise of Education Reform of Spatial Computing and Analysis Courses Towards Smart City. *Journal of Geomatics*, 47(S1), 14–17. <https://doi.org/10.14188/j.2095-6045.2021911>
- Zhu, A.-X., Zhao, F.-H., Liang, P., & Qin, C.-Z. (2021). Next generation of GIS: must be easy. *Annals of GIS*, 27(1), 71–86. <https://doi.org/10.1080/19475683.2020.1766563>