Spatial Information Research (2026) 34:16
https://doi.org/10.1007/s41324-026-00670-3

REVIEW

A review of large language models in geomatics: integrating multimodal data, addressing challenges, and exploring synergies

Horia Ameen[1] · Mario Soilán[1]

Received: 29 July 2025 / Revised: 21 January 2026 / Accepted: 25 January 2026 / Published online: 21 February 2026

Abstract
The exponential growth of geospatial data from satellite imagery and LiDAR to crowdsourced and sensor-based streams poses challenges for traditional geomatics methods. This review examines the transformative role of large language models (LLMs) and vision language models (VLMs) in addressing these challenges by enabling the integration, interpretation, and automation of multimodal geospatial data. The study synthesizes applications across three domains: geospatial data analysis and management, environmental and urban planning, and remote sensing. Particular attention is given to domain-specific copilots and intelligent agents that enhance automation and decision support. The review also highlights advances in multimodal data fusion, autonomous GIS systems, and context-aware mapping tools, while critically assessing limitations such as computational demands, domain adaptability, and ethical concerns. By providing a structured overview of current applications and emerging research directions, the study underscores the potential of language-driven models to augment geomatics methodologies and reshape innovation, decision-making, and policy in spatial information science.

Keywords GeoAI · Large language models (LLMs) · Vision–language models (VLMs) · Geospatial data · Multimodal models

1 Introduction

Geomatics is an interdisciplinary field dedicated to the collection, analysis, and interpretation of geospatial data, and it has advanced considerably with innovations in data visualization, acquisition, and processing. Historically rooted in cartography and photogrammetry, the field has expanded to include GIS and remote sensing, enabling the integration of diverse spatial datasets. Today, geomatics encompasses a wide range of sources like satellite imagery, aerial photography, LiDAR, drone-based photogrammetry, and radar data. With the rapid growth of communication technologies and internet infrastructure, geospatial information now extends beyond conventional maps and imagery to include unstructured streams such as social media, crowdsourced contributions, and real-time IoT sensor data.

This rapid data growth necessitates advanced data processing and interpretation tools that can handle any type of data, extract meaningful insights, and facilitate timely decision-making across applications like urban planning, environmental monitoring, and disaster response etc. Recently, large language models (LLMs) have emerged as transformative tools in artificial intelligence [1], designed to process vast amounts of text data and capable of performing complex language-based tasks, reviews the evolution of language modeling, tracing the shift from statistical models to neural networks and then to LLMs that leverage massive data and model scaling to improve performance [2]. As model sizes increase, LLMs display unique capabilities beyond those of smaller models, such as enhanced language comprehension and generation. Recent advancements, such as ChatGPT, Deepseek AI, Google Gemini, Grok etc. highlight the transformative impact of LLMs in AI. These models, initially developed for tasks in natural language processing, have shown great promise in fields requiring semantic understanding and data synthesis. The inspection of how large vision language models (LVLMs), integrated with multi-sensor data from cameras and LiDAR, enhance the understanding of traffic scenes. Traffic scene analysis is

✉ Horia Ameen
horia.ameen@uvigo.gal

Mario Soilán
msoilan@uvigo.es

1 University of Vigo, Vigo, Spain

complex due to interactions among road users, infrastructure, and traffic rules, often requiring detailed reasoning that traditional task-oriented neural networks struggle to achieve [3]. LVLMs, however, leverage vast knowledge bases to effectively process spatiotemporal information, supporting visual question-answering (VQA) for transportation research.

Emerging trends in geospatial analysis highlight the integration of traditional spatial data with human behavior sources such as social media and mobile data, alongside the growing role of GeoAI and LLMs in advancing multimodal analytics and guiding future research directions [4]. Addressing these diverse data types requires sophisticated methods capable of interpreting both types of data information, further highlighting the importance of advancements in geospatial data analytics.

A scientometric analysis based on the Web of Science database using the query term “GeoAI” for the period 2020–2025 identifies 329 publications with a total of 3,805 citations and an H-index of 33. Figure 1 shows a sharp increase in publications beginning in 2021, with a significant rise in both publication count and citations during 2023–2025. This trend reflects the rapid expansion of GeoAI research in recent years, highlighting the growing relevance of AI-driven methods in geospatial domains and reinforcing the need for a focused review of emerging approaches such as LLMs and VLMs. This review takes a comprehensive approach by analyzing LLMs capabilities tailored to geomatics, integrating diverse applications into a unified perspective. It also outlines existing limitations, offers guidance for future research, and identifies key areas where LLM advancements could drive innovation in geospatial analysis and decision-making.

The primary objectives of this review are as follows: (1) To synthesize how LLMs and VLMs are applied across key geomatics domains, including geospatial data analysis and management, environmental/urban planning, and remote sensing, with attention to both structured and multimodal data sources. (2) To evaluate how LLMs integrate with specialized geospatial frameworks and intelligent agents (e.g., GeoAgent, ChatGeoAI, PlanGPT, Tree-GPT) to enhance automation, interpretation, and decision support in geomatics workflows. (3) To examine the practical challenges of adopting LLMs in geomatics, particularly issues of data limitations, model adaptability, computational efficiency, and ethical considerations and propose directions for addressing them in future research.

This paper is structured as follows: Sect. 1 presents the background, motivation, scope, and objectives. Section 2 covers the fundamentals of LLMs, including its architecture and geomatics-relevant features. Section 3 discusses LLM applications in geo-spatial data analysis, interpretation, and decision support. Section 4 is about discussion and analysis including the challenges such as data limitations, adaptability, computational demands, ethical issues and future directions and the conclusion is provided in Sect. 5.

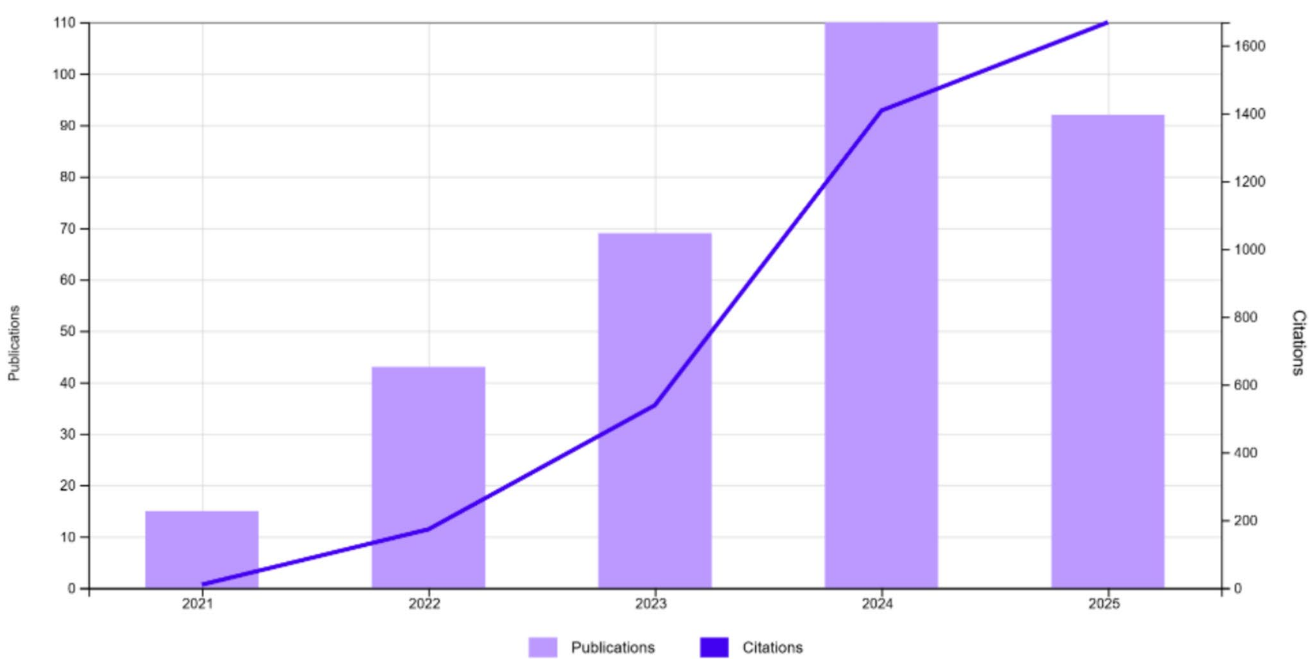


Fig. 1 Publications and citations on GeoAI (2021–2025)

2 Fundamentals of large language models

2.1 Overview of LLMs

LLMs represent a significant leap in AI, these are advanced AI models trained on vast amounts of data to understand and generate human language, LLMs utilize complex architectures such as the Transformer model. Transformers are designed to process sequential data efficiently by employing self-attention mechanisms, which allow the model to weigh the importance of each part of the input data relative to others. Unlike previous recurrent neural networks (RNNs) models e.g., LSTM and GRU, which process data step-by-step, Transformers can simultaneously consider all parts of the sequence, providing a structured framework for parallelized processing and long-range dependency management. This enables LLMs to handle long contexts and complex relationships within the data, making them highly effective for natural language understanding and generation. The relevance of Transformers in LLMs, the emphasis on architecture's ability to efficiently handle dependencies across long sequences, setting the foundation for the powerful models we see today.

Figure 2 showing Transformer architecture diagram, inspired by the "Attention is All You Need" paper, illustrating the core idea of multi-headed self-attention in sequence processing. The architecture of transformers forms the foundation of LLMs [5].

Recent advancements in Transformer-based LLMs have focused on improving the ability to handle long-context prompts, where inputs may contain extensive information requiring holistic processing to generate coherent and relevant responses. Modern LLMs have significantly extended context windows compared to earlier models like GPT-3. For instance, GPT-4 Turbo supports up to 128 k tokens, enabling the processing of entire documents or extended

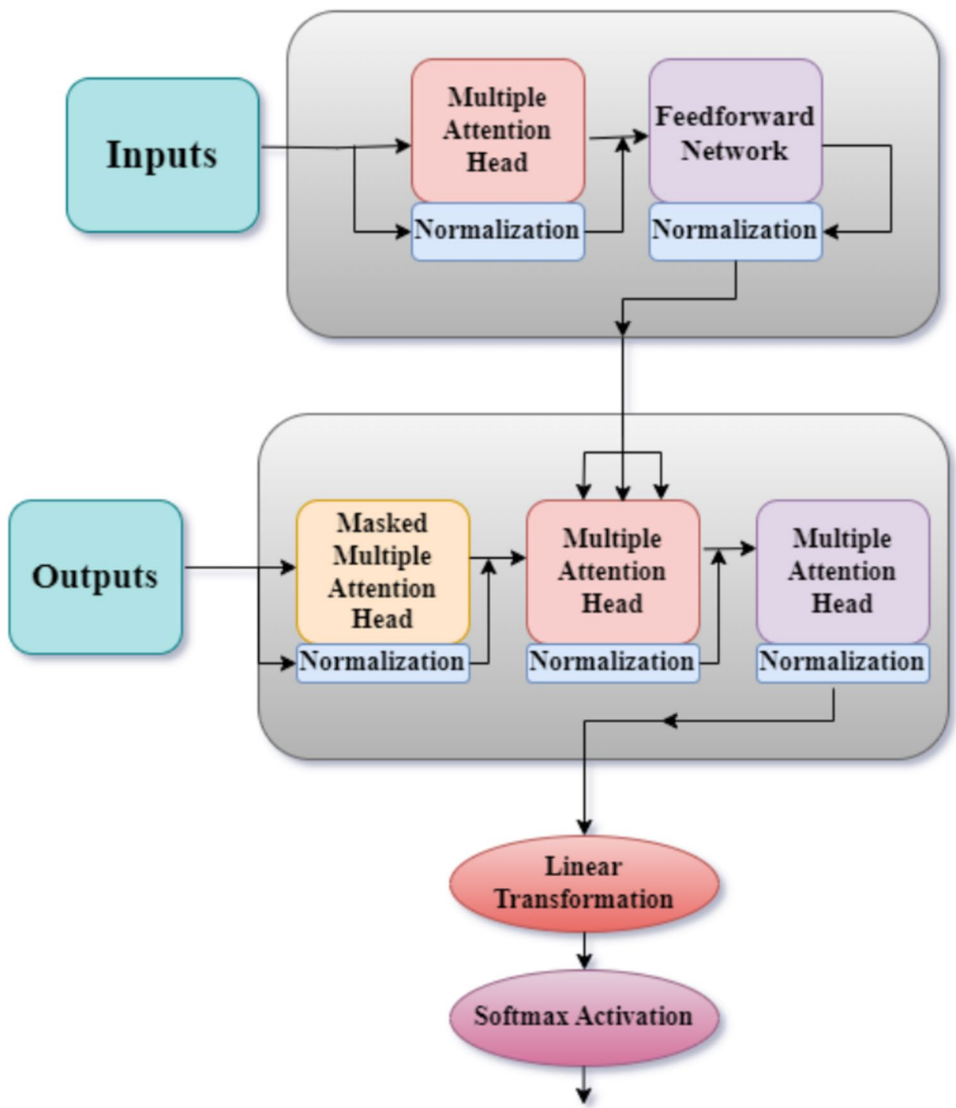


Fig. 2 Transformer architecture with layered self-attention

conversation histories in a single prompt. These models also integrate efficient memory mechanisms to maintain coherence over longer interactions [6]. As LLMs continue to evolve, their relevance transcends merely language generation, becoming integral to fields like geospatial analysis, healthcare, and engineering. Engineers and practitioners must consider features like token limits, cost efficiency, and personalization to select the most appropriate model for their specific needs. However, this architectural choice allows LLMs to grasp intricate dependencies and relationships across language data, accommodating long-range context. Recent advancements in LLMs have demonstrated their potential in causal discovery tasks, where identifying cause-and-effect relationships between variables is essential for reasoning and decision-making.

As these models grow in size and parameter count, they demonstrate emergent abilities such as zero-shot and few-shot learning [7], which were not observed in earlier, smaller models. Zero-shot learning refers to a model's ability to perform a task it has never explicitly been trained for, based solely on its understanding of the task's description. Few-shot learning involves the model being able to perform a task with only a few examples provided during training, rather than requiring large amounts of labeled data. These capabilities are significant because they allow larger models, like GPT-4, to generalize better across a variety of tasks without task-specific fine-tuning. Smaller models, such as GPT-2 or earlier transformer-based models, lacked the scale and parameter count to generalize effectively in this way, often requiring extensive task-specific training to perform well on a new task. These unique capabilities show how LLMs can be novel in advancing language understanding and generation, making them powerful tools for applications that require nuanced interpretation and contextual awareness, such as in geomatics.

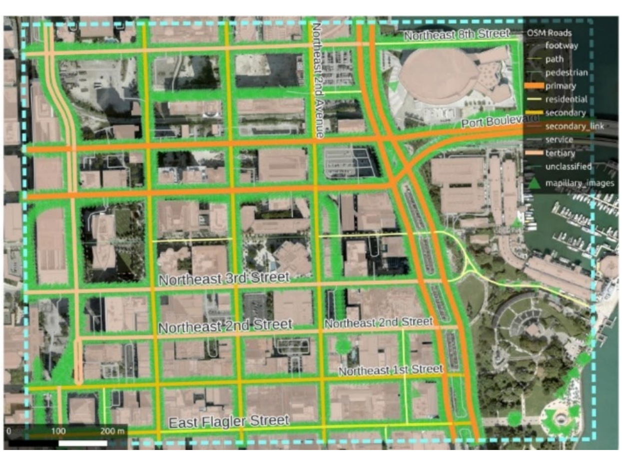


Fig. 3 Integration of textual and visual data in geomatics using VLMs for enhanced geospatial analysis [14]

2.2 Characteristics relevant to geomatics

Geomatics involves specialized terminology across sub-disciplines such as surveying, GIS, and remote sensing. Large language models (LLMs), particularly when fine-tuned, can adapt to these domain-specific vocabularies, enabling accurate interpretation of technical language. This capability is crucial for tasks such as automated documentation and geospatial data annotation. Vision–language models (VLMs) like the versatile and honest model (VHM) demonstrate this potential in remote sensing, outperforming existing approaches in scene classification, visual question answering, and visual grounding [8].

LLMs also show value in understanding spatial relationships and geographic contexts, supporting tasks such as land-use mapping [9], crop yield prediction through multimodal data fusion [10], and extracting correlations across geospatial features [11]. Domain-specific adaptations, including the spatial–temporal data pretrained language model (STD-PLM), have been developed to improve applications in environmental monitoring [12] and urban planning [13]. Furthermore, multimodal integration allows LLMs to connect visual data (e.g., satellite imagery, street-level photos) with textual descriptions, providing context-aware insights for mapping and annotation tasks [14].

Figure 3 illustrates the integration of textual and visual data in geomatics, showcasing how multimodal models can enhance data interpretation. It depicts maps and street-level photographs, emphasizing the role of VLMs in correlating visual information with descriptive insights, facilitating tasks such as an-notation and context-aware analysis in geospatial applications.

These characteristics highlight how LLMs can be tailored or adapted to geomatics tasks, allowing for more efficient analysis, interpretation, and management of complex geospatial datasets.

2.3 Traditional and classical approaches to geospatial analysis

Foundational approaches methods in geomatics encompass three primary categories: rule-based models, statistical methods, and classical machine learning. Rule-based models rely on explicit logic and deterministic queries, such as SQL-based operations in GIS tools, which are extended through spatial query languages to handle spatial operations, relationships, and graphical presentations, enabling precise retrieval and manipulation of geospatial data [15]. Statistical methods include foundational techniques like regression and Kriging, which are widely used for geospatial interpolation and analysis. Regression methods, such as linear regression and multiple regression, establish relationships between dependent and independent variables, enabling predictions based on observed data.

Kriging, on the other hand, is a geostatistical interpolation method that predicts unknown values in a geographic area by leveraging the spatial autocorrelation of observed data points. Unlike regression, Kriging incorporates both the distance and the degree of variation between sample points, offering more accurate and localized pre-dictions. When integrated with GIS, Kriging not only provides detailed visualizations of spatial data but also supports climate change analysis by identifying trends over time and space, aiding in planning and resource management. These methods remain critical in geomatics for their ability to handle structured data and provide interpretable, deterministic results in environmental and geospatial applications [16]. Classical machine learning methods, such as Random Forests and support vector machines (SVMs), have been pivotal in geomatics for solving complex spatial problems, including land cover classification and clustering. For instance, SVMs with various kernel functions have been employed in GIS-based land-slide susceptibility mapping to accurately identify high-risk areas by analyzing conditioning factors like slope, soil type, and rainfall patterns [17].

Despite their contributions, these methods face inherent limitations. They rely heavily on structured data, requiring extensive data preprocessing and feature engineering, which restricts their adaptability to complex or unstructured geospatial information. Scalability is another challenge, as these algorithms may become computationally expensive with large geospatial datasets. Vision-based deep learning models, such as convolutional neural networks and transformer-based vision architectures, have significantly advanced tasks like image classification, object detection, and semantic segmentation in remote sensing. However, these models are primarily optimized for perceptual understanding and typically lack the ability to perform higher-level reasoning, contextual interpretation, or interactive querying [4].

In contrast, LLMs introduce complementary capabilities by enabling semantic reasoning, cross-domain knowledge integration, and natural language interaction with geospatial data. Rather than replacing vision-based deep learning methods, LLMs can augment existing geospatial analysis pipelines by facilitating multimodal fusion, automating complex workflows, and supporting explainable decision-making. This distinction highlights that LLMs are not a universal substitute for vision-based approaches but instead offer an additional layer of abstraction that enhances interpretability, accessibility, and reasoning in geomatics applications [11].

3 Applications of LLMs in geomatics

Section 3 delves into the growing applications of LLMs within geomatics, where their ability to process and interpret geospatial data is transforming geomatics applications. These applications can be grouped into three primary clusters: (a) geospatial data analysis and management, (b) environmental monitoring and urban planning and (c) remote sensing applications.

Each cluster represents a distinct set of tasks and techniques that leverage the power of LLMs to enhance the efficiency and accuracy of geospatial decision-making.

3.1 Geospatial data analysis and management

Geoinformatics involves the collection, processing, and interpretation of geographic data to support decision-making. The goal is to convert raw geospatial data into actionable insights, enabling effective resource management, spatial planning, and problem-solving. One of the key challenges in this field is the ability to process and analyze large volumes of complex geospatial data, which often require intricate spatial reasoning and an understanding of spatial relationships. For non-technical users, traditional geospatial analysis results such as reports, dashboards, or technical maps may be less intuitive and insufficient to answer complex questions posed by decision-makers. These outputs often lack contextual explanations, making it difficult for non-experts to derive actionable insights. To address this gap LLMs provide context-aware explanations, making geospatial insights more accessible and easier to communicate to non-experts. For example, frameworks like GeoAgent integrate LLMs with retrieval-augmented generation (RAG) techniques and Monte Carlo tree search (MCTS) algorithms [18]. The RAG technique enables the model to retrieve relevant knowledge from external geospatial databases, generating more accurate, context-driven outputs tailored to user queries. This capability ensures that decision-makers and non-technical users receive not only the data but also the interpretive insights necessary for informed decision-making.

MCTS, is a search algorithm that helps GeoAgent perform sequential reasoning and refine tasks iteratively. In the context of geospatial data analysis, MCTS enables the system to break down complex, multi-step tasks into smaller, manageable tasks. Each sub-task is executed, and the results are evaluated to determine the next steps. This iterative process allows GeoAgent to refine its output continuously, learning from feedback after each step, which is essential when dealing with data that evolves over time or requires ongoing adjustments.

Figure 4 shows how GeoAgent paves the way for automated geospatial task programming, facilitating more accessible and efficient use of geospatial data for research purposes.

In the application presented by Mansourian and Oucheikh [19], the challenge of making geospatial data analysis accessible to non-experts is addressed by using LLMs. Typically, working with tools like PyQGIS requires specialized technical knowledge, but this system allows users to interact with geospatial data using natural language and shows how Llama 2, an LLM, interprets plain language queries and generates the corresponding PyQGIS code to perform spatial data manipulation tasks. This automation significantly lowers the barrier for non-experts, allowing them to execute tasks that would typically require advanced GIS skills. One of the key innovations of this approach is the integration of geospatial ontologies, which provide a structured framework for understanding the relationships between geospatial terms and concepts. These ontologies help Llama 2 understand the context of the queries, enabling it to produce more accurate results. By embedding these ontologies into the LLM, Llama 2 can refine its interpretations, ensuring that even ambiguous or complex queries are processed correctly.

In Fig. 5 the recreated architecture is inspired by the work of (Ali Mansourian et al.) on ChatGeoAI. That supports the claim of democratizing access to geospatial analysis by demonstrating that Llama 2 can effectively translate user queries into accurate spatial data manipulation tasks without requiring technical GIS knowledge.

Traditional GIS tasks often require manual interventions, involving repetitive actions that demand expertise and substantial human resources. Recent advancements show that LLMs such as GPT-4 can be integrated into Autonomous GIS systems to address these challenges. Such systems aim to self-generate, self-organize, self-verify, self-execute, and self-grow geospatial tasks, including decision-making, data collection, data operation, and history retrieval. These capabilities enable automated collection, analysis, and visualization of spatial data, thereby reducing the need for human oversight [20].

This leads to a reduction in manual operation time and human intervention. For instance, the system can autonomously generate graphs and maps based on data inputs [21], returning accurate results without needing further human involvement, thus streamlining geospatial data analysis workflows and making them more accessible. As in Fig. 6 the workflow of LLM-Geo.

In a project report a series of structured experiments have been conducted to evaluate ChatGPT's ability to handle geospatial data analysis tasks. These tests were designed to address gaps in the literature regarding ChatGPT's capability to process geospatial file formats (e.g., GML, GeoJSON,

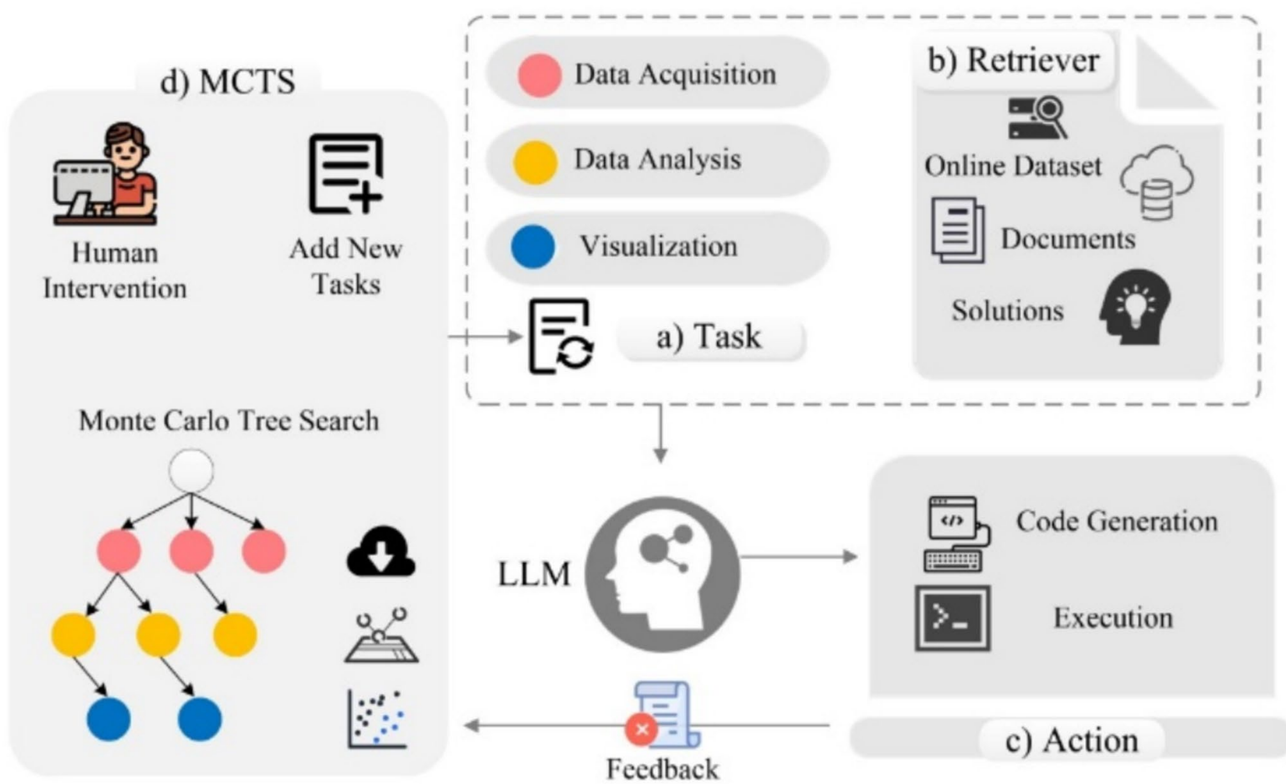


Fig. 4 A geospatial data analysis task programming agent [18]

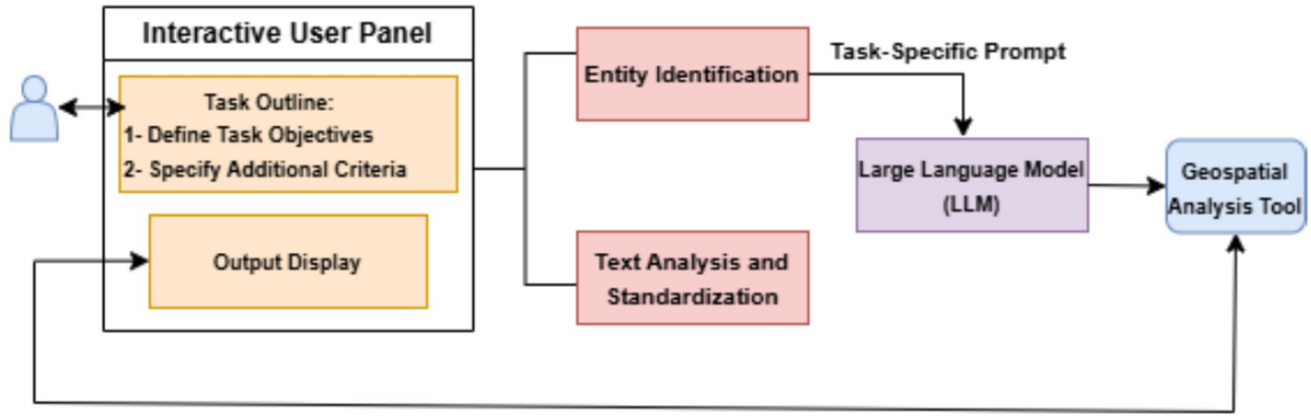


Fig. 5 System architecture for ChatGeoAI [19]

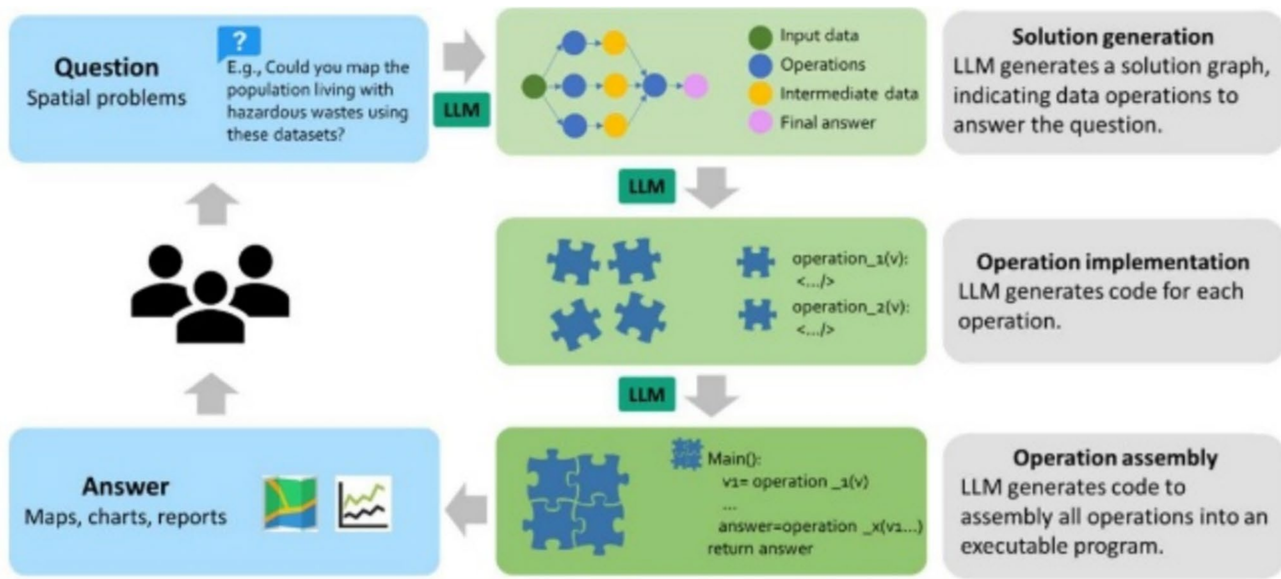


Fig. 6 Workflow of Geo-LLM [20]

and Shapefiles) and its ability to interact with external APIs. The experiments, namely Test 1: Geospatial Data Analysis Using File Formats, Test 2: Comparing File Upload and API Calling in ChatGPT-4, and Test 3: Using ChatGPT and LangChain for API Calls, determine whether ChatGPT could be effectively utilized for automating GIS tasks, such as data manipulation, spatial queries, and visualization [22].

Building on the advancements demonstrated in the studies cited, LLMs are significantly transforming geospatial data analysis and management. However, there are several frameworks with similar names, such as GeoAgent, Geo-LLM, GeoLLM-Engine, and LLM-Geo, each addressing distinct aspects of geospatial data processing. These tools often aim to automate workflows, enhance accessibility for non-experts, and improve efficiency, but there is limited comparative analysis to evaluate their performance on common benchmarks. For instance, while GeoAgent excels in task decomposition and iterative refinement, Geo-LLM emphasizes autonomous GIS task execution. Similarly, LLM-Geo focuses on to the next generation of AI-powered GIS [23], GeoX address geometric understanding and reasoning tasks by bridging the gap between diagrams, symbols, and language through specialized pre-training and alignment techniques [24]. And GeoLLM-Engine targets multimodal integration for spatial analysis and for building geospatial copilots [25]. Unfortunately, no standardized benchmarks or datasets (e.g., MMLU, TruthfulQA) have been used to compare these tools comprehensively, making it challenging to assess their relative strengths or identify the best practices. This lack of standardization highlights an important gap in the field and provides an opportunity for future re-search to conduct systematic evaluations of these frameworks using shared metrics and benchmarks. Such efforts could significantly enhance the scientific community's understanding of how LLMs contribute to geospatial data analysis and management.

3.2 Environmental and urban planning applications

LLMs are enhancing environmental and urban planning by integrating geospatial technologies to address sustainability, monitoring, and development challenges.

One of the significant uses of LLMs is in flood risk perception, where the integration of a flood knowledge-constrained LLM with GIS plays a crucial role in providing accurate and personalized flood risk information. A system has been developed that leverages LLMs and GIS to communicate flood risks through natural language, making complex geospatial data more accessible to the public. The system not only generates precise information based on flood knowledge graphs but also adapts its responses to users' cognitive levels, helping to narrow the cognitive disparities that typically hinder flood risk comprehension across different demographics. This is particularly beneficial in emergency scenarios, where clear communication can aid in quicker and more informed decision-making. Notably, the integration of voice-based feedback and textual answers made the system accessible to a wide range of users, ensuring that people with no GIS expertise could easily interact with the system and receive accurate, understandable flood risk information [26].

In the context of smart city optimization, the integration of LLMs and IoT plays a pivotal role in enhancing urban information and communications technology (ICT) processes. LLMs enable real-time decision-making by interpreting complex datasets, including traffic flow and energy consumption, in natural language. This capability allows city managers to receive more nuanced and context-aware insights from IoT-generated data. For instance, LLMs can be used to optimize traffic management systems by providing real-time analysis and suggesting adjustments to traffic lights based on the evolving situation. Similarly. The integration of LLMs in smart city applications also helps overcome challenges like data confidentiality, as these models can be constrained with privacy-preserving protocols, enabling them to process sensitive information securely [27].

LLMs also play a crucial role in urban planning tasks and planning evaluation. For example, PlanGPT, a custom fine-tuned LLM, has been developed to assist in text generation and document evaluation for urban planning. This model helps planners generate proposals, evaluate urban designs, and optimize space usage, ultimately improving the quality and efficiency of urban development processes [13].

For monitoring water pollution via image analysis, YOLOv5 combined with the OpenAI API is used to identify pollutants and provide contextual alerts based on image analysis. This approach aids in the real-time monitoring of water quality, ensuring rapid responses to contamination events and supporting environmental protection initiatives [12]. Furthermore, LLMs have been employed in crop disease detection and farmer guidance. By combining MobileNetV2, ResNet50, and LLMs, deep learning models achieve high accuracy (97–99%) in detecting crop diseases from a dataset of over 39,000 images. This Integrated LLM provides tailored recommendations for crop disease prevention and profit optimization for farmers via a mobile application. Hence, helps farmers detect diseases early, guiding them on proper interventions to prevent crop damage and increase yields [28].

The integration of LLMs into environmental and urban planning applications shows their ability to bridge data-driven decision-making with accessibility for diverse stakeholders. This capability becomes even more significant when applied to remote sensing, where the analysis of vast amounts of satellite, UAV, and LiDAR data is critical for addressing large-scale environmental and geospatial challenges. Despite these promising developments, several challenges remain in the adoption of LLM-based systems for environmental and urban planning. Many of the existing approaches rely heavily on large, high-quality, and domain-specific datasets, which limit their generalizability across regions with sparse or heterogeneous data. The integration of LLMs with IoT and remote sensing systems also introduces issues related to data noise, uncertainty propagation, and real-time scalability. Furthermore, privacy, security, and ethical concerns become particularly critical in urban contexts, where sensitive personal and infrastructural data are often involved. The lack of standardized benchmarks and evaluation protocols for assessing LLM-driven planning and environmental monitoring systems further complicates objective performance comparisons. Future research should focus on developing robust multimodal fusion frameworks, interpretable decision-support mechanisms, and privacy-preserving architectures [27].

The following section delves into how LLMs are driving innovation in remote sensing by integrating multimodal data analysis and improving geospatial intelligence workflows.

3.3 LLMs applications in remote sensing

Remote sensing, a vital technique within geomatics, utilizes satellite or airborne sensors to capture data about the Earth's surface. It provides critical information for applications such as mapping, environmental monitoring, and disaster management. Recent advancements in cross domain AI models, such as GPT for language understanding, SAM for image segmentation, and CLIP for cross-modal alignment are increasingly transforming remote sensing by enabling more accurate data interpretation, automated image analysis, and informed decision-making.

The segment anything model (SAM) [29] and updated version SAM-2 [30], by Meta AI, is a cutting-edge model designed for zero-shot segmentation of images, capable of identifying and segmenting objects in visual data without prior training on specific datasets. However, SAM 2 is designed as a unified model capable of handling promptable segmentation tasks across images as well as videos, offering improved accuracy and real-time performance. SAM can identify regions of interest in satellite imagery, it is used to efficiently extract cage mariculture areas from remote sensing imagery, but its performance for raft mariculture detection remains limited, requiring manual interpretation [31]. SAM's ability to segment features like water bodies, vegetation, and urban areas directly from satellite or aerial imagery makes it particularly promising for remote sensing applications.

The contrastive language–image pretraining (CLIP), developed by OpenAI, pioneered the alignment of textual and visual data using contrastive learning, enabling robust understanding of image-text relationships [32]. Its applications in re-mote sensing include land-use classification, change detection, and object labeling in aerial imagery. Models like RemoteCLIP extend CLIP's capabilities specifically for geospatial domains, enabling tasks such as zero-shot classification, image-text retrieval, and object counting through domain-aligned embeddings [33].

CLIP can label these regions based on textual descriptions; remote sensing multilingual CLIP is a model by fine-tuning CLIP with a self-supervised approach that aligns local and global image representations while leveraging machine-translated captions in multiple languages. It enhances cross-modal and multilingual image-text retrieval, as well as zero-shot classification, achieving state-of-the-art performance in vision-language tasks [34].

These models, leveraging large-scale Transformer networks, enable multi-modal tasks such as scene classification, object detection, land-use classification, change detection, and video object tracking [35]. They integrate visual and language data for more efficient interpretation and decision-making, offering substantial improvements in accuracy, efficiency, and adaptability to complex Remote Sensing tasks [36]. While the vast volume of remote sensing data presents storage and access challenges, foundation models (FMs) contribute by optimizing data indexing, retrieval, and automated processing. However, integrating FMs with advanced data compression and management strategies remains essential for efficiently handling the extensive datasets produced by Earth observation satellites [37].

In the multi-sensor image comprehension domain, EarthGPT, a multimodal LLM, has been developed to enhance remote sensing image understanding. By combining various sensor data, EarthGPT employs a cross-modal mutual comprehension approach to integrate visual perception and language understanding effectively. The model demonstrates significant advancements across multiple remote sensing tasks. For instance, in scene classification on the NWPU-RESISC45 dataset, EarthGPT outperformed state-of-the-art specialist models such as CSDS and T-CNN, with accuracy improvement of 0.25% and 0.79%, respectively. In zero-shot classification, EarthGPT achieved notable Top 1 accuracy improvements on the CLRS dataset (25.61%) and the NaSC-TG2 dataset (38.63%) compared to competing multimodal LLMs like Qwen-VL-Chat and Sphinx. Furthermore, in image captioning on the NWPU-Caption dataset, EarthGPT demonstrated significant improvements across all evaluation metrics, with gains of 17.7% (BLEU4), 10.5% (METEOR), and 66.2% (CIDEr-D), underscoring its robust captioning capabilities. In visual question answering (VQA), EarthGPT surpassed state-of-the-art models by 11.09% (CRSVQA) and achieved an average accuracy of 72.06% in zero-shot VQA, with improvements of 25.60% over models like MiniGPTV2. For visual grounding tasks, EarthGPT enhanced precision and intersection-over-union (IoU) metrics, with increases of 3.13% (cIoU) and 1.30% (mIoU) compared to other specialist models. These results highlight EarthGPT's superior ability to handle diverse remote sensing tasks, bridging the domain gap between natural and remote sensing imagery while setting a new benchmark for multimodal image under-standing in the remote sensing domain [38].

ChatGPT has also been integrated with remote sensing models to support a range of interpretation tasks. In this system, users can submit remote sensing images with queries, and the model generates responses based on task-specific planning and execution. The correctness of task planning was evaluated across multiple models, including GPT-3.5-turbo, GPT-4, and GPT-3. Among these, GPT-4 achieved the highest correctness (97.2%), particularly excelling in scene classification, land use classification, and object detection. This significantly outperformed GPT-3.5-turbo (95.4%) and GPT-3 (89.7%), demonstrating GPT-4's superior ability to align remote sensing queries with task-specific plans. These results highlight how LLMs can simplify and enhance remote sensing analysis by providing highly accurate task planning, thereby improving automation and accessibility for diverse users [39].

Another notable development is Tree-GPT [40], a modular LLM expert system designed for forest remote sensing image understanding. By incorporating domain knowledge bases and toolchains, Tree-GPT enhances LLM capabilities to analyze forestry data and generate code for calculating tree structural parameters from remote sensing images. Experimental results using high-resolution data from Shenzhen, China (2018), including RGB orthoimages (0.025 m resolution) and LiDAR data (100 points/m^2), demonstrate that Tree-GPT achieves segmentation accuracy comparable to manual annotations, with the segment anything model (SAM) often outperforming manual methods in tree segmentation tasks. To refine segmentation, Tree-GPT employs a bipartite graph matching algorithm using RetinaNet (target detection) and ResNet50 (backbone network) with DeepForest pre-trained weights. Beyond segmentation, Tree-GPT excels in data visualization and interactive analysis, enabling tasks such as retrieving specific tree parameters (e.g., tallest tree), generating scatter plots, visualizing tree growth diagrams, and creating box plots grouped by height categories. For statistical learning, Tree-GPT estimates Gaussian distribution parameters with RMSE as a confidence measure, supporting more precise ecological modeling. Its domain knowledge base, structured into 4,000-token blocks and indexed using Facebook AI similarity search (FAISS) algorithm, ensures efficient information retrieval. Moreover, the integration of foundation models (FMs) and multimodal large language models (LLMs) is transforming remote sensing by enhancing automation, accuracy, and multimodal analysis. SAM and CLIP enable precise segmentation and alignment of visual and textual data, while ChatGPT and EarthGPT excel in diverse tasks like scene classification, object detection, and image captioning, often surpassing state-of-the-art models.

Collectively, these advancements underscore the transformative role of LLMs and FMs in remote sensing, driving efficiency and innovation across applications. Additionally, the comprehensive Table 1 provided in this section underscores the versatility of LLMs across a wide range of remote sensing and geomatics domains, showcasing their impact on geospatial data analysis, environmental and urban planning and remote sensing tasks. However, these advancements are accompanied by significant challenges, including issues related to data quality, scalability, and

Table 1 Application of LLMs in geomatics

Domain	Application	Technique used	Key feature	Citation
Geospatial data analysis and management	Geospatial data (data acquisition, analysis, visualization)	GeoAgent (RAG, MCTS, code interpreter)	Uses MCTS and RAG for complex data processing, establishing benchmarks for LLMs in geospatial analysis	[18]
	Geospatial analysis via natural language interface	ChatGeoAI (Llama 2 with PyQGIS)	Convert natural language queries into PyQGIS code, enhancing accuracy through geospatial ontologies	[19]
	Development of autonomous GIS systems	LLM-Geo (GPT-4)	Self-generating, organizing, verifying, executing, and growing spatial data for GIS tasks, automating map creation	[20]
	Handling GIS tasks and data	GPT-4, ChatGPT Code Interpreter	Evaluates the effectiveness of ChatGPT for autonomous GIS analysis	[22]
Environmental and urban planning applications	Flood risk perception	Flood knowledge constrained LLM integrated with GIS	Combines LLM with GIS to deliver accurate flood risk information via natural language	[26]
	Smart city optimization	LLMs, NLP, Deep Learning, Federated Learning (FL), IoT, Blockchain	The integration of emerging technologies to enhance ICT processes, address data confidentiality challenges, and drive innovation in sustainable smart cities	[27]
	Urban planning tasks, planning evaluation	PlanGPT (custom LLM)	Uses fine-tuned LLMs for text generation and document evaluation in urban planning	[13]
	Crop disease detection and farmer guidance	MobileNetV2, ResNet50, LLMs	Utilizes UAV imagery combined with CNN-based techniques to detect diseases and pests in crops	[28]
	Monitoring water pollution via image analysis	YOLOv5, ChatGPT API	Combines YOLOv5 and ChatGPT for pollutant identification and contextual alerting	[12]
Remote sensing applications	Remote sensing interpretation	ChatGPT, Visual Models	Uses ChatGPT to connect various AI-based remote sensing models to solve complex interpretation tasks	[39]
	Multi-sensor image comprehension in RS	EarthGPT (Multimodal LLM)	Integrates multi-sensor data and employs a cross-modal mutual comprehension approach for remote sensing tasks	[38]
	High-resolution remote sensing analysis	(GeoPixel) Pixel-grounding Multimodal LLM	First end-to-end RS-LMM with pixel-level grounding for 4 K imagery; dynamic image partitioning; GeoPixelD dataset	[41]
	Unified remote sensing vision-language tasks	SkyEyeGPT (LLM with RS instruction tuning)	Handles open-ended RS vision-language tasks including image captioning, VQA, visual grounding, and UAV video captioning with superior performance compared to GPT-4 V	[42]
	Forest remote sensing analysis	Tree-GPT (LLM Expert System)	Combines LLMs with image understanding modules to analyze forestry data and generate code for analysis	[40]
	Visual image analysis for RS	Visual ChatGPT	Combines ChatGPT's capabilities with visual computation for remote sensing image analysis and description	[36]

computational demands, which necessitate further exploration and innovation.

To address these challenges, several promising directions have emerged. Data quality issues can be mitigated through domain-adaptive pretraining, self-supervised learning, and synthetic data generation, which reduce dependence on large-scale manual annotations. Weakly supervised and noise-robust learning strategies also help improve generalization across heterogeneous remote sensing datasets. Scalability constraints can be alleviated using hierarchical modeling, patch-based processing, and distributed inference frameworks that enable efficient handling of large satellite and UAV image archives. Furthermore, computational demands can be reduced through model compression, knowledge distillation, and parameter-efficient fine-tuning techniques, making deployment more feasible on resource-constrained platforms [44].

4 Discussion

A paradigm shift in the analysis, interpretation, and management of geospatial data is marked by the incorporation of LLMs. They have greatly enhanced geospatial data processing via the integration of different kinds of data. Unlike conventional GIS tools that rely on manually curated datasets, models such as GeoAgent and LLM-Geo intro-duce automation, reducing the complexity of geospatial task execution. GeoAgent, for example, improves spatial data processing by integrating Monte Carlo tree search (MCTS) for decision optimization, making it more efficient than rule-based geospatial processing techniques. Similarly, LLM-Geo automates GIS tasks such as map generation and decision-making, which traditionally required extensive human intervention. Beyond automation, LLMs enhance geospatial interpretation and reasoning. Models like EarthGPT and RemoteCLIP demonstrate superior scene classification, image captioning, and retrieval capabilities by leveraging multimodal learning. These improvements indicate that LLMs are not merely aiding geospatial workflows but redefining the way geospatial data is analyzed.

Despite these advantages, LLM performance varies significantly depending on the application. While Tree-GPT achieves segmentation accuracy comparable to manual and notations, it still requires iterative refinement for optimal results. Additionally, models like SAM, though powerful in zero-shot segmentation, struggle with certain remote sensing tasks such as raft mariculture detection, necessitating manual adjustments. This suggests that while LLMs and vision-language models provide substantial efficiency gains, they have not fully replaced traditional remote sensing workflows.

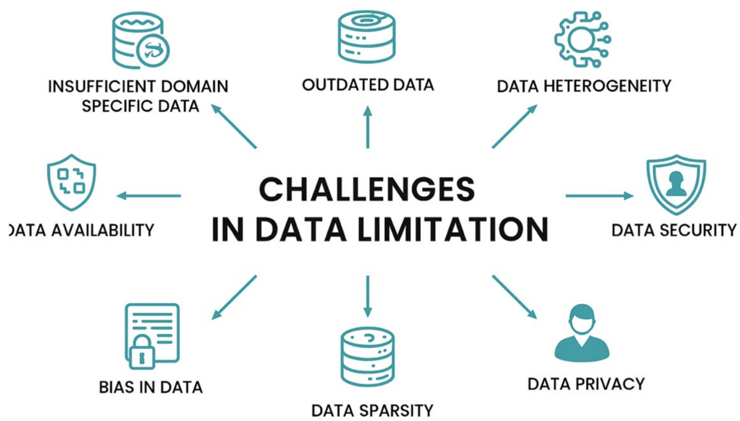
Advancements in natural language processing, powered by state-of-the-art models like GPT-4o and Gemini 2.5, are enhancing access to geospatial tools by enabling interaction with datasets through natural language queries. These improvements lower the entry barrier for non-experts, expanding the user base and fostering innovation. Such capabilities align closely with the paper's objectives of improving geospatial data integration and advancing decision-making processes through LLM-driven methodologies.

This review sets out to achieve three pivotal objectives to provide a comprehensive exploration of LLM capabilities in harmonizing diverse data formats structured geo-spatial data, such as satellite imagery, and unstructured sources, like textual survey responses or sensor data. Such integration addresses a long-standing gap in geomatics by enabling seamless synthesis of disparate data types for a unified analysis. By highlighting targeted strategies, such as fine-tuning and instruction tuning, it lays a pathway for overcoming these impediments, ensuring the scalable application of LLMs in tasks like urban planning and environmental monitoring. And the ability of LLMs to process natural language queries, as demonstrated in ChatGeoAI, reduces reliance on domain expertise, enabling non-specialists to interact with geospatial datasets, a crucial step toward democratizing geospatial analysis. This synergy not only enriches the interpretative depth of geospatial datasets but also streamlines automation of traditionally labor-intensive processes, thus broadening the scope of applications from disaster response to resource optimization. However, implementing LLMs for geospatial tasks involves overcoming significant challenges. These challenges begin with data limitations, such as the lack of diverse, domain-specific datasets and issues like sparsity, heterogeneity, and outdated information. Further, model adaptability is a critical barrier, as LLMs often struggle to adjust to geospatial tasks without losing general capabilities. Moreover, computational requirements and ethical considerations demand careful attention to ensure responsible and equitable use of LLMs in geospatial analysis. Further explores these intertwined challenges in detail, offering insights into potential solutions.

4.1 Data limitations

One of the key challenges in applying LLMs to geospatial tasks is the limitation of data in diverse and dynamic geospatial environments. The effective use of LLMs for geospatial analysis depends heavily on the quality and quantity of training data. Figure 7 illustrates the key challenges in data limitations that affect the implementation of LLMs. It starts with the lack of domain-specific data, which limits the availability and diversity of relevant datasets. This leads to further challenges such as data sparsity, heterogeneity, and

Fig. 7 Challenges in data limitations for LLMs



outdated data, all of which hinder the model's ability to generalize effectively.

As LLMs scale, the available stock of public human text data is expected to reach its limit by 2028–2032. Beyond this, training LLMs on additional data could be unsustainable without new approaches such as synthetic data generation, transfer learning, or improving data efficiency. Further work should focus on creating large-scale, domain-specific geospatial benchmarks, leveraging synthetic data generation, self-supervised learning, and multimodal data fusion (e.g., text, imagery, GIS layers).

4.2 Model adaptability

The adaptability of LLMs to new domains and tasks is a critical challenge, referring to a model's ability to adjust and perform effectively across different tasks, domains, or environments without requiring extensive retraining.

LLMs rely heavily on next-token prediction, making it difficult to generate con-text-aware and accurate responses, especially in specialized domains. Fine-tuning often produces generic outputs, sacrificing depth and relevance to avoid inappropriate content [43]. One critical issue is the limited influence developers have over LLM functionalities, particularly as models scale. The emergent nature of capabilities in LLMs makes it difficult for developers to anticipate or control all aspects of their behavior. For instance, while GPT-4 demonstrated many desired abilities envisioned by its creators, it also exhibited undesirable behaviors, such as generating instructions for harmful activities, necessitating extensive post-deployment mitigation efforts. This shows the inherent unpredictability in adapting models for diverse, real-world applications. Moreover, LLM adaptability is further challenged by their susceptibility to hallucinations [44], where models generate false yet plausible outputs. Although larger models tend to improve factual accuracy, failure modes persist, especially in cases where simplistic mitigation strategies fail silently, leading to false confidence in the model's reliability. Effective adaptation to different domains and tasks requires innovative strategies that leverage LLMs' inherent capabilities while minimizing undesirable outputs. The unpredictability of how users may elicit novel and unforeseen behaviors from LLMs compounds these challenges. Future research should focus on improving the adaptability of LLMs through parameter-efficient fine-tuning methods, retrieval-augmented generation, and tool-integrated systems that enable interaction with geospatial databases and GIS platforms.

4.3 Computational requirements

The development and deployment of LLMs face significant computational challenges, particularly related to their reliance on GPUs, CPUs, and other specialized hardware. These issues hinder scalability, accessibility, and efficiency in both training and inference processes. Specifically, these challenges include:

Hardware dependency: Training LLMs like GPT versions require high-performance GPUs or TPUs capable of processing massive amounts of data and executing complex computations [44]. However, such hardware is expensive, often costing thousands of dollars per unit, limiting accessibility for smaller organizations. While CPUs are more common and cost-effective, their processing capabilities fall short for LLM workloads, particularly for tasks requiring real-time responses or parallel processing of large datasets.

High memory demand: Training LLMs demands substantial memory resources, with requirements escalating from gigabytes to terabytes as model parameters increase from millions to trillions. Such demands often require clusters of GPUs, driving up costs and complicated setups. Inference, though less resource-intensive, still demands substantial memory, especially for handling large batch sizes or complex prompts [2].

Energy inefficiency: The extensive use of power-intensive hardware results in significant energy consumption. Training large models on GPU clusters consumes terawatt-hours of electricity, contributing to environmental concerns. Millions of dollars are spent on computing hours and energy usage for pre-training LLMs, which results in significant financial and environmental expenses. These models, known as Red AI, use enormous amounts of computing power to attain cutting-edge performance [45].

Scalability and deployment challenges: Scaling LLMs across distributed GPU/TPU clusters is non-trivial due to limitations in interconnect speeds, resource allocation, and load balancing. Real-time deployment in edge environments is nearly impossible due to the reliance on cloud-based GPU servers, increasing latency and operational costs [46]. Highlights scalability challenges for LLMs, including limited context windows, resource-intensive training, and the complexities of managing vast datasets.

Efforts to mitigate these challenges associated with LLMs focus on improving efficiency, scalability, and accessibility. Model compression techniques such as pruning, quantization, and distillation aim to reduce model size, allowing deployment on less powerful GPUs or CPUs. Simultaneously, hardware innovations, including the development of AI-specific chips like NVIDIA's H100 GPUs and Google's TPUs, are designed to optimize computational workloads for LLMs. Distributed training optimization seeks to enhance parallel processing and inter-GPU communication algorithms, reducing overhead and increasing training efficiency. Additionally, cloud-based solutions provide scalable GPU resources to organizations lacking access to advanced hardware, though they come with cost and privacy concerns.

In the context of geomatics, these computational challenges are further amplified by the nature of geospatial data. Applications often involve high-resolution satellite imagery, LiDAR point clouds, hyperspectral data, and long spatiotemporal sequences, all of which significantly increase memory usage and processing time compared to text-only tasks. Moreover, real-time or near-real-time applications, such as disaster response, traffic monitoring, and environmental surveillance, impose strict latency constraints that are difficult to meet with large LLM-based systems.

Ultimately, addressing these challenges requires a combination of advances in model architecture and hardware design, ensuring that LLMs are more efficient, cost-effective, and widely accessible.

4.4 Ethical considerations

Geospatial data, especially when tied to individual location histories, property details, or sensor measurements, inherently raises concerns about privacy, data protection, and anonymity. These issues become more critical as geo-data often contains sensitive information that could be misused if not managed responsibly. However, LLMs introduce additional complexities. Therefore, integrating LLMs into geomatics requires robust ethical frameworks focused on informed consent, transparency, and responsible data management. The potential for LLMs to disseminate misinformation in social, and personal domains, such as influencing elections or facilitating scams [47]. Moreover, the "black box" nature of LLMs, referred to lack of transparency, complicates accountability and makes it challenging to understand or rectify their outputs [48]. Economically, the automation capabilities of LLMs threaten job displacement in sectors like customer service and content creation, necessitating workforce retraining. In educational contexts, reliance on LLMs can lead to human skill degradation, as students may depend on AI-generated content, thereby undermining critical thinking and learning processes [49].

Looking forward, the trends in LLM research suggest several promising avenues for geomatics like the integration of LLMs with domain-specific models. For instance, integrating remote sensing-specific language models trained on datasets from NASA, ESA, and open-source GIS platforms could enhance LLM comprehension of spatial relationships and geospatial semantics. One of the limitations in the current research landscape is the lack of standardized evaluation metrics for geospatial LLMs. While NLP models benefit from widely recognized benchmarks such as MMLU, Truth-fulQA, and SuperGLUE, geospatial models lack a comprehensive framework for comparing performance across different applications. Establishing standardized benchmarks for tasks such as land-use classification, remote sensing object detection, and geospatial reasoning will enable more meaningful comparisons and guide future model improvements. The establishment of transparent, unbiased, and privacy-preserving frameworks is essential for the responsible use of LLMs in geomatics. As LLMs continue to redefine geospatial analysis, it is imperative for geomatics professionals to stay abreast of these developments. In addition, future research must also focus on reducing hallucinations in geospatial LLMs, which remain a major barrier to reliability and real-world deployment. This can be achieved through the development of high-quality, domain-specific training datasets, improved grounding mechanisms, and multimodal supervision using authoritative geospatial sources.

Understanding the capabilities and limitations of these tools will be crucial for leveraging their potential effectively. Collaboration across AI, geomatics, and other scientific disciplines will foster innovation, enabling LLMs to tackle increasingly complex geospatial challenges.

5 Conclusion

This review has comprehensively explored the integration of LLMs with geospatial technologies, demonstrating how these approaches extend beyond language processing to enhance geospatial data integration, interpretation, and automation. Applications span remote sensing analytics, GIS task automation, urban and environmental planning, and intelligent geospatial assistants, with recent advances such as GeoAgent, Tree-GPT, RemoteCLIP, and GeoPixel illustrating the accelerating convergence of language, vision, and spatial intelligence. At the same time, critical challenges remain. The field requires domain-specific multimodal datasets, standardized evaluation frameworks, and interpretable models that can support reliable decision-making in high-stakes geospatial contexts. Addressing these limitations will be essential to ensure both technical robustness and ethical responsibility. However, LLMs hold immense promises in driving sustainable development and environmental monitoring in the AI-driven era.

Acknowledgements This work has been funded by grant RYC2021-033560-I funded by MCIN/AEI/https://doi.org/10.13039/501100011033 and by European Union NextGenerationEU/PRTR and by grant ED431F 2024/02 funded by Xunta de Galicia, Spain-GAIN. Funding for open access charge: Universidade de Vigo/CISUG.

Author contributions Conceptualization, M.S.; methodology, M.S.; formal analysis, H.A.; investigation, H.A.; resources, M.S.; data curation, H.A.; writing—original draft preparation, H.A.; writing—review and editing, M.S.; visualization, H.A.; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript."

Funding Funding for open access publishing: Universidade de Vigo / CISUG. This work has been supported by the Spanish Ministry of Science, Innovation and Universities through Grant RYC2021-033560-I funded by MCIN/AEI/https://doi.org/10.13039/501100011033 and by European Union Next Generation EU/PRTR, and by grant ED431F 2024/02 funded by Xunta de Galicia, Spain-GAIN.

Data availability No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Declarations

Conflict of interest The authors declare that they have no competing interests, financial or non-financial, directly or indirectly related to the work submitted for publication.

References

1. Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., Mian, A. (2024). A comprehensive overview of large language models. https://arxiv.org/abs/2307.06435
2. Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y. (2024). A survey of large language models. https://arxiv.org/abs/2303.18223
3. Jain, S., Thapa, S., Chen, K.-T., Abbott, A. L., & Sarkar, A. (2024). Semantic understanding of traffic scenes with large vision language models. *IEEE Intelligent Vehicles Symposium (IV), 2024*, 1580–1587. https://doi.org/10.1109/IV55156.2024.10588373
4. Hochmair, H. H., Navratil, G., & Huang, H. (2023). Perspectives on advanced technologies in spatial data collection and analysis. *Geographies, 3*(4), Article 4. https://doi.org/10.3390/geographies3040037
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In: *Advances in Neural Information Processing Systems, 30 (NeurIPS 2017)*. Curran Associates, Inc. https://arxiv.org/abs/1706.03762
6. Huang, Y., Xu, J., Lai, J., Jiang, Z., Chen, T., Li, Z., Yao, Y., Ma, X., Yang, L., Chen, H., Li, S (2024). *Advancing transformer architecture in long-context large language models: A comprehensive survey*. https://arxiv.org/abs/2311.12351
7. Meshkin, H., Zirkle, J., Arabidarrehdor, G., Chaturbedi, A., Chakravartula, S., Mann, J., Thrasher, B., & Li, Z. (2024). Harnessing large language models' zero-shot and few-shot learning capabilities for regulatory research. *Briefings in Bioinformatics, 25*(5), Article bbae354. https://doi.org/10.1093/bib/bbae354
8. Pang, C., Weng, X., Wu, J., Li, J., Liu, Y., Sun, J., Li, W., Wang, S., Feng, L., Xia, G.S., He, C. (2024). *VHM: Versatile and honest vision language model for remote sensing image analysis*. http://arxiv.org/abs/2403.20213
9. Wu, H., Li, Y., Lin, A., Fan, H., Fan, K., Xie, J., & Luo, W. (2024). A review of crowdsourced geographic information for land-use and land-cover mapping: Current progress and challenges. *International Journal of Geographical Information Science, 38*(11), 2183–2215. https://doi.org/10.1080/13658816.2024.2379468
10. Tan, K. (2024). *Large language models for crop yield prediction*. Research Square. https://doi.org/10.21203/rs.3.rs-4750823/v1
11. Zhang, Y., Wang, Z., He, Z., Li, J., Mai, G., Lin, J., Wei, C., & Yu, W. (2024). BB-GeoGPT: A framework for learning a large language model for geographic information science. *Information Processing & Management, 61*(5), Article 103808. https://doi.org/10.1016/j.ipm.2024.103808
12. Samuel, D. J., Sermet, Y., Cwiertny, D., & Demir, I. (2024). Integrating vision-based AI and large language models for real-time

water pollution surveillance. *Water Environment Research, 96*(8), Article e11092. https://doi.org/10.1002/wer.11092
13. Zhu, H., Zhang, W., Huang, N., Li, B., Niu, L., Fan, Z., Lun, T., Tao, Y., Su, J., Gong, Z., Fang, C. (2024). *PlanGPT: Enhancing urban planning with tailored language model and efficient retrieval*. https://arxiv.org/abs/2402.19273
14. Juhász, L., Mooney, P., Hochmair, H. H., & Guan, B. (2024). ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs. https://arxiv.org/abs/2306.03204
15. Egenhofer, M. J. (1994). Spatial SQL: A query and presentation language. *IEEE Transactions on Knowledge and Data Engineering, 6*(1), 86–95. https://doi.org/10.1109/69.273029
16. Fragkoulis, M., Spinellis, D., & Louridas, P. (2015). An interactive SQL relational interface for querying main-memory data structures. *Computing, 97*(12), 1141–1164. https://doi.org/10.1007/s00607-015-0452-y
17. Kamran, K. V., Feizizadeh, B., Khorrami, B., & Ebadi, Y. (2021). A comparative approach of support vector machine kernel functions for GIS-based landslide susceptibility mapping. *Applied Geomatics, 13*(4), 837–851. https://doi.org/10.1007/s12518-021-00393-0
18. Chen, Y., Wang, W., Lobry, S., & Kurtz, C. (2024). *An LLM agent for automatic geospatial data analysis*. https://arxiv.org/abs/2410.18792
19. Mansourian, A., & Oucheikh, R. (2024). ChatGeoAI: Enabling geospatial analysis for public through natural language, with large language models. *ISPRS International Journal of Geo-Information, 13*(10), Article 10. https://doi.org/10.3390/ijgi13100348
20. Li, Z., & Ning, H. (2023). *Autonomous GIS: The next-generation AI-powered GIS*. https://arxiv.org/abs/2305.06453
21. Deguchi, H., Shibata, K., & Taguchi, S. (2024). Language to map: Topological map generation from natural language path instructions. http://arxiv.org/abs/2403.10008
22. Holm, K. O. M. (2023). *LLMs—The death of GIS analysis? An investigation into using large language models for GIS data analysis* (Specialization project, Department of Geomatics, Norwegian University of Science and Technology). NTNU. Retrieved from https://kartai.no/wp-content/uploads/2024/01/Holm-2023-LLMs-The-Death-of-GIS-Analysis.pdf
23. 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
24. Xia, R., Li, M., Ye, H., Wu, W., Zhou, H., Yuan, J., Peng, T., Cai, X., Yan, X., Wang, B., He, C. (2025). *GeoX: Geometric problem solving through unified formalized vision-language pre-training*. https://arxiv.org/abs/2412.11863
25. Singh, S., Fore, M., & Stamoulis, D. (2024). GeoLLM-Engine: A realistic environment for building geospatial copilots. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2024)* (pp. 585–594). IEEE. https://openaccess.thecvf.com/content/CVPR2024W/EarthVision/html/Singh_GeoLLM-Engine_A_Realistic_Environment_for_Building_Geospatial_Copilots_CVPRW_2024_paper.html
26. Zhu, J., Dang, P., Cao, Y., Lai, J., Guo, Y., Wang, P., & Li, W. (2024). A flood knowledge-constrained large language model interactable with GIS: Enhancing public risk perception of floods. *International Journal of Geographical Information Science, 38*(4), 603–625. https://doi.org/10.1080/13658816.2024.2306167
27. Ullah, A., Qi, G., Hussain, S., Ullah, I., & Ali, Z. (2024). *The role of LLMs in sustainable smart cities: Applications, challenges, and future directions*. https://arxiv.org/abs/2402.14596
28. Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., Xu, J. L., & He, Y. (2024). Intelligent agriculture: Deep learning in UAV-based remote sensing imagery for crop diseases and pests detection. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2024.1435016
29. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P. (2023). *Segment anything*. https://arxiv.org/abs/2304.02643
30. Ravi, N., Gabeur V, Hu YT, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E (2024). *SAM 2: Segment anything in images and videos*. https://arxiv.org/abs/2408.00714
31. Ren, Y., Yang, X., Wang, Z., Yu, G., Liu, Y., Liu, X., Meng, D., Zhang, Q., & Yu, G. (2023). Segment anything model (SAM) assisted remote sensing supervision for mariculture—using Liaoning Province, China as an example. *Remote Sensing, 15*(24), Article 24. https://doi.org/10.3390/rs15245781
32. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. (2021). *Learning transferable visual models from natural language supervision*. https://arxiv.org/abs/2103.00020
33. Liu, F., Chen, D., Guan, Z., Zhou, X., Zhu, J., Ye, Q., Fu, L., & Zhou, J. (2024). RemoteCLIP: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–16. https://doi.org/10.1109/TGRS.2024.3390838
34. Silva, J. D., Magalhães, J., Tuia, D., & Martins, B. (2024). Multilingual vision-language pre-training for the remote sensing domain. In: *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24)* (pp. 220–232). Association for Computing Machinery. https://doi.org/10.1145/3678717.3691318
35. Zhang, Z., Wang, C., Song, J., & Xu, Y. (2022). Object tracking based on satellite videos: A literature review. *Remote Sensing, 14*(15), Article 15. https://doi.org/10.3390/rs14153674
36. Osco, L. P., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., & Marcato Junior, J. (2023). The potential of Visual ChatGPT for remote sensing. *Remote Sensing, 15*(13), Article 13. https://doi.org/10.3390/rs15133232
37. Xu, C., Du, X., Yan, Z., & Fan, X. (2020). ScienceEarth: A big data platform for remote sensing data processing. *Remote Sensing, 12*(4), Article 4. https://doi.org/10.3390/rs12040607
38. Zhang, W., Cai, M., Zhang, T., Zhuang, Y., & Mao, X. (2024). EarthGPT: A universal multimodal large language model for multisensor image comprehension in remote sensing domain. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–20. https://doi.org/10.1109/TGRS.2024.3409624
39. Guo, H., Su, X., Wu, C., Du, B., Zhang, L., & Li, D. (2024). *Remote sensing ChatGPT: Solving remote sensing tasks with ChatGPT and visual models*. https://arxiv.org/abs/2401.09083
40. Du, S., Tang, S., Wang, W., Li, X., & Guo, R. (2023). *Tree-GPT: Modular large language model expert system for forest remote sensing image understanding and interactive analysis*. https://arxiv.org/abs/2310.04698
41. Shabbir, A., Zumri, M., Bennamoun, M., Khan, F. S., & Khan, S. (2025). *GeoPixel: Pixel grounding large multimodal model in remote sensing*. https://arxiv.org/abs/2501.13925
42. Zhan, Y., Xiong, Z., & Yuan, Y. (2024). *SkyEyeGPT: Unifying remote sensing vision-language tasks via instruction tuning with large language model*. https://arxiv.org/abs/2401.09712
43. Browning, J. (2024). Getting it right: The limits of fine-tuning large language models. *Ethics and Information Technology, 26*(2), Article 36. https://doi.org/10.1007/s10676-024-09779-1
44. Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences, 14*(5), Article 5. https://doi.org/10.3390/app14052074
45. Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2024). Large language models: A comprehensive survey of its applications, challenges,

limitations, and future prospects. *Preprints*. https://doi.org/10.36227/techrxiv.23589741.v6
46. Jin, H., Huang, L., Cai, H., Yan, J., Li, B., & Chen, H. (2024). *From LLMs to LLM-based agents for software engineering: A survey of current, challenges and future*. https://arxiv.org/abs/2408.02479
47. Tokayev, K.-J. (2023). Ethical implications of large language models: A multidimensional exploration of societal, economic, and technical concerns. *International Journal of Social Analysis, 8*(9), Article 9.
48. Liyanage, U. P., & Ranaweera, N. D. (2023). Ethical considerations and potential risks in the deployment of large language models in diverse societal contexts. *Journal of Computational Social Dynamics, 8*(11), Article 11.
49. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370