Journal of Spatial Information Science
Number 30 (2025), pp. 93–116
doi:10.5311/JOSIS.2025.30.389

Research Article

Opportunities and challenges of integrating geographic information science and large language models

Nico Van de Weghe[1,2], Lars De Sloover[1,2], Anthony G. Cohn[3,4,5], Haosheng Huang[1,2], Simon Scheider[6], Renée Sieber[7,8], Sabine Timpf[9], and Christophe Claramunt[10]

[1]Department of Geography, Ghent University, Ghent, Belgium
[2]GeoAI Research Center, Ghent University, Ghent, Belgium
[3]School of Computing, University of Leeds, Leeds, The United Kingdom
[4]The Alan Turing Institute, London, The United Kingdom
[5]Department of Computer Science & Technology, Tongji University, Shanghai, China
[6]Department of Human Geography & Spatial Planning, Utrecht University, Utrecht, the Netherlands
[7]Department of Geography, McGill University, Montreal, Canada
[8]Bieler School of Environment, McGill University, Montreal, Canada
[9]Institute of Geography, Geoinformatics Group, University of Augsburg, Augsburg, Germany
[10]Naval Academy Research Institute, Brest Naval, France

Received: July 12, 2024; returned: October 29, 2024; revised: December 17, 2024; accepted: March 7, 2025.

Abstract: The integration of large language models (LLMs) with geographic information science (GIScience) represents a new frontier in interdisciplinary research that combines advanced natural language processing with sophisticated spatial data analysis. This paper explores the synergistic potential of combining the natural language understanding and generation capabilities of LLMs with the expertise of GIScience in handling complex geospatial data. By exploring the specific contributions that LLMs can offer to GIScience, such as improving data processing, analysis, and visualization, and the mutual benefits that GIScience can offer to LLMs in terms of spatial reasoning and conceptual frameworks, we outline a comprehensive framework and a research agenda for this integration. Furthermore, we address the societal and ethical implications of this convergence, highlighting the challenges of bias, misinformation, and environmental impact. Through this exploration, we aim to set the stage for innovative applications in urban planning, environmental analysis, and beyond, while emphasizing the need for responsible use of AI.

Keywords: large language models (LLMs), geographic information science (GIScience), multimodal data integration, spatial reasoning

---

1 Introduction

At a time when both data-driven and knowledge-driven insights are crucial, the integration of large language models (LLMs) and geographic information science (GIScience) represents a new area of interdisciplinary research. The integration of artificial intelligence (AI) has driven GIScience in the past by enabling sophisticated spatial analysis and modeling techniques. This trend began in the early 1990s, when the decline of traditional quantitative methods in geography sparked interest in new AI techniques and expert-based computational approaches [48]. Formal models soon followed, providing AI reasoning mechanisms and languages for processing geographic objects at a fundamental level [16, 51]. With the increasing incorporation of foundation AI models into GIScience, deep learning and machine learning models have been widely used for accurate interpolation and analysis of large geospatial datasets, replication, simulation, and prediction of environmental and urban phenomena [36].

However, the emergence of LLMs heralds a new era and offers unprecedented potential for the expansion of GIScience through improved natural language understanding, geospatial text generation, and innovative applications at the interface of language and spatiotemporal data. LLMs using transformer architectures [57] and large amounts of training data can recognize, summarize, translate, predict, and generate text and other forms of content [9]. These models are characterized by their ability to understand and generate language for general purposes, allowing them to perform a wide range of natural language processing tasks based on patterns learned from data. Despite their ability to generate coherent and contextually relevant text, LLMs still lack true understanding or semantic comprehension equivalent to human cognition [6]. Furthermore, LLMs still encounter significant limitations in knowledge-intensive and complex tasks. Problems such as hallucinations, the tendency to draw incorrect conclusions due to biases in data input, the lack of integration of new knowledge, and the traceability of information sources are major limitations that LLMs must overcome [6, 46].

As emphasized in previous work [28, 37], 'spatial is special' because spatial data requires reasoning about unique relationships (such as proximity, adjacency, and scale) that are fundamentally different from non-spatial data. GIScience involves advanced spatial data analysis and reasoning [19] to identify the underlying theories of geographic information. While LLMs operate based on statistical patterns learned from data, they lack an inherent semantic understanding of the concepts they generate. They can mimic comprehension by producing coherent and contextually relevant text, but they do not truly understand the meaning behind the text [7]. Integrating LLMs with GIScience can lead to innovative methods and insights, especially in the processing and interpretation of geospatial data.

We report on the synergy between the ability of GIScience to process complex geospatial data and the capabilities of LLMs. In exploring this integration, it is crucial to be aware of the ethical implications and potential risks, particularly in relation to AI-generated misinformation, data breaches, and potential risks for opaque decision-making, misuse, and even illegal activity. Our exploration acknowledges both the potential and challenges of combining the capabilities of GIScience and LLMs. This synergy sets the stage for further research and innovative applications in areas where geospatial data is paramount, such as urban planning, smart cities, environmental analysis, wildlife research, health research, and beyond.

In the following sections, we explore the specific contributions of LLMs to GIScience and how these models can improve the creation, processing, analysis, and visualization of geospatial data. We also discuss the mutual benefits that GIScience can provide to the development and refinement of LLMs, particularly with regard to spatial reasoning, formal models, and the integration of complex, multidimensional data. A separate section addresses the societal implications and ethical considerations of the integration of both, highlighting potential biases and the need for responsible use of AI. We then present a comprehensive framework (Figure 1) that outlines the unique contributions of LLMs and GIScience to each other, as well as the open questions that require further research. Challenges mentioned as 'societal issues' in Figure 1 are discussed in the concluding comments of this paper.

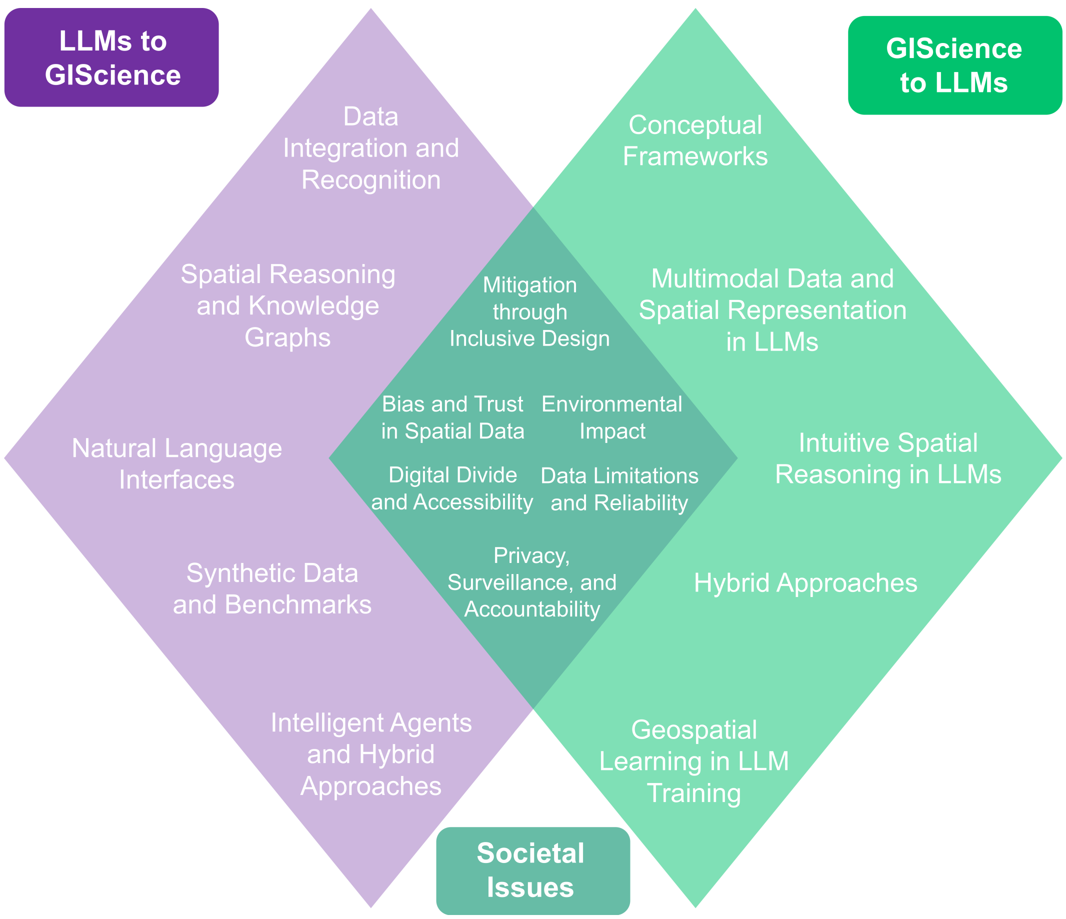

Figure 1: Framework for interdisciplinary contributions and open questions in the integration of LLMs and GIScience.

2 What LLMs offer to GIScience

2.1 Data integration and recognition

LLMs present a transformative opportunity to enhance GIScience by streamlining access to geospatial data and automating complex tasks using their natural language processing (NLP) capabilities. Through advanced semantic analysis, LLMs can analyze unstructured documents and identify spatial relationships within texts, such as relative location descriptions (e.g., "next to", "north of", or "between"), supporting the extraction and inference of spatial patterns [49]. These models can automatically identify location-based information within text, such as place names, coordinates, or spatial relationships, and generate specialized code to process and analyze this data. This capability allows users to streamline tasks that would otherwise require advanced programming skills. Furthermore, LLMs simplify the handling of complex geographic datasets, making it easier to organize, query, and visualize spatial information in ways that were previously challenging for non-experts [12]. The integration of LLMs with imagery supports image-to-text generation, which enables the extraction of meaningful information from visual data [50]. Additionally, multimodal LLMs allow users to ask questions about features or labels on maps, facilitating more intuitive interaction with geographic data.

2.2 Spatial Reasoning and Knowledge Gaps

Spatial relation inference, a valuable function within GIScience, refers to the ability of models to understand and interpret spatial relationships between objects or entities in a given context. LLMs trained to identify spatial relationships—such as proximity, containment, or directional cues—can enable tasks like spatial joins, spatial indexing, and even workflow automation within GIS applications [29]. When supported by knowledge graphs that link spatial entities and relationships in a structured format, LLMs can perform more targeted searches and retrieve relevant spatial information efficiently [27,45]. This integration allows the model to consider not only the language but also the underlying spatial data connected to that language [17].

Spatial recognition tasks often involve textual and visual data. For instance, models such as CLIP (contrastive language-image pre-training) have shown promise in associating textual descriptions with corresponding images, allowing for richer spatial understanding [49]. Fine-tuning a general-purpose LLM on domain-specific datasets containing geographic and spatial information can improve its spatial reasoning capabilities. This process involves adjusting the model's parameters based on additional training data, focusing on spatial terms, geographical contexts, and common spatial reasoning tasks. Further reasoning processes can focus on specific spatial operations such as proximity analyses [35]. Moreover, LLMs can refine their understanding by processing sequential prompts to improve the accuracy of their spatial inferences. Finally integrating LLMs with embodied agents (robots or systems that physically interact with the environment) may provide real-world experiences that enrich the model's spatial reasoning, such as embodied large vision-language models (LVLMs) [15].

Latest research has demonstrated the effectiveness of LLMs in representing textual descriptions of geometry and spatial relations within GIScience [31]. The authors systematically assess how well LLMs interpret and generate spatially explicit descriptions, highlighting their strengths and weaknesses in understanding geometric concepts. They establish benchmarks for the performance of LLMs in terms of geometry and spatial relationships to facilitate comparisons between models and identify the most effective ones. The findings indicate that LLMs can support spatial reasoning processes by accurately interpreting complex spatial relationships, which is valuable for various GIScience applications. Complementary work by [8] explores how LLMs comprehend geographic data and support geospatial decision-making. Focusing on autoregressive models, it employs three experimental approaches: probing LLMs for geo-coordinates to assess geospatial knowledge, using geospatial and non-geospatial prepositions to evaluate awareness, and conducting a multidimensional scaling (MDS) experiment to test reasoning abilities in determining city locations. The results suggest that the effective synthesis of geospatial knowledge may not necessarily require larger LLMs, as commonly assumed. Instead, a federation of smaller, specialized LLMs tailored to specific geospatial tasks could provide an alternative, potentially more efficient approach. This highlights a promising direction for future research, emphasizing sophistication and task-specific optimization over sheer scale.

2.3 Natural language interfaces

Recent advances in LLMs, characterized by sophisticated natural language user interfaces and multimodality, may fundamentally change the way geospatial data is accessed, managed, analyzed, and visualized [10]. These models enable interactions through a conversational, human-like language, significantly reducing the complexity traditionally associated with analyzing geospatial data using geographic information systems (GIS). LLMs have shown the ability to extract geographic patterns and outliers from large documents and predict human trajectories from large location-based series of movements [62]. This integration streamlines analytical workflows and enriches geovisualization by enabling users to ask queries in natural language and obtain intuitive visual representations tailored to specific goals. LLMs, when extended appropriately, can also trigger geographic data processing workflows that embed specific queries and produce cartographic outputs [39]. Indeed, LLMs trained on extensive textual data, including geographic information and relationships, can often extract patterns and correlations, suggest potential solutions and strategies for problem-solving tasks [11], and facilitate dialog with non-experts [65, 68]. These advances lower the technical barriers of traditional GIS, making sophisticated geospatial data analysis accessible to a wider audience, and fostering a more inclusive environment for spatiotemporal exploration and discovery.

2.4 Synthetic data and benchmarks

LLMs also have the potential to generate synthetic datasets that simulate various spatiotemporal settings, which is valuable for hypothesis testing and model development. In the age of big data with vast volumes, velocities, and varieties of data (e.g., remotely sensed data), validation of geospatial data is a challenge. Synthetic dataset generation is a promising area, allowing LLMs to simulate diverse spatiotemporal scenarios useful for hypothesis testing, model development, and data validation, especially when real-world data collection is resource-intensive [22]. LLMs can generate synthetic datasets for GIScience by leveraging their understanding of spatial relationships and contextual data through a process known as generative modeling. This involves training the model on existing geographic datasets to learn the underlying patterns and distributions of spatial phenomena. The LLM can then create new datasets by sampling from these learned distributions while respecting certain user-defined constraints such as geographic boundaries, population densities, or environmental factors. This capability enhances the training and validation of various machine learning models in GIS, supporting predictive analytics, and improving unsupervised learning and exploratory data analysis. LLMs provide virtual assistant capabilities through natural language, opening up many possibilities for much greater interaction between GIS and end users with non-expert skills, and then new avenues for interactive learning. However, there is still a need for the development of unified frameworks that provide appropriate guidelines for synthetic data generation [40].

To evaluate the quality and reliability of these synthetic datasets and LLMs' spatial capabilities broadly, recent synthetic benchmarks have been developed to assess the spatial reasoning ability of LLMs and their ability to develop comprehensive problem-solving strategies [13,35]. These benchmarks are crucial for validating both the synthetic data generation process and the broader spatial reasoning capabilities of LLMs. However, spatial reasoning remains a challenge for LLMs because they are not embodied and humans acquire much of their spatial understanding through physical interaction with the world [34]. Combining LLMs with explicit spatial reasoners can mitigate these limitations and provide more robust methods for spatial analysis [35].

2.5 Intelligent agents and hybrid approaches

Beyond synthetic data creation, integrating LMMs into GIScience holds the potential to streamline research processes and foster innovative thinking. As a result, researchers might explore new forms of problem-solving and in-depth hypothesis formulation and testing as technical complications are reduced. For example, the integration of LLMs into research workflows has been shown to enhance hypothesis generation, leveraging the ability of machine learning algorithms to identify patterns beyond human recognition. These models can generate novel, high-quality hypotheses that are rated as comparable to, or even exceeding, those produced by human researchers in terms of clarity, originality, and impact [4,67].

Additionally, AI tools can reduce the cost and time associated with exploring combinatorial spaces for innovation, which significantly enhances the efficiency of hypothesis testing and development [1]. While there are concerns about over-reliance on these technologies potentially stifling creativity, structured approaches to hypothesis generation, such as using machine learning to reveal novel insights from high-dimensional datasets, demonstrate how these tools can complement human ingenuity [41]. This interplay between human creativity and AI-driven innovation could herald a dramatic shift in the pace and scope of advances in GIScience. Furthermore, the integration of LLMs into intelligent agents—autonomous systems capable of perceiving their environment and making decisions to achieve specific goals [30]—could significantly advance GIScience. Intelligent agents can contribute to GIScience by improving our understanding of how spatial relationships and temporal dynamics are structured and analyzed. By learning complex geospatial concepts, intelligent agents can help develop more sophisticated models for spatial analysis. This integration can lead to deeper insights into the nature of geographic information and the principles that govern space and time, and provide a more robust framework for spatial reasoning and analysis. Such systems can use sensors to collect data and feed it back to the language model, allowing it to learn from physical interactions, similar to how humans acquire spatial understanding through experience [34]. To effectively integrate LLMs with spatial models as spatial agents, several best practices and potential roadblocks must be considered. First, ensuring that the datasets used for training both LLMs and spatial models are consistent in terms of terminology and geographic references can significantly enhance performance. Additionally, employing multimodal training strategies that combine textual and spatial data allows LLMs to learn a more comprehensive model. Leveraging knowledge graphs to connect spatial entities and their relationships can also improve spatial reasoning capabilities. Challenges include data quality and availability, as incomplete datasets can hinder model performance. The complexity of spatial reasoning itself poses another barrier; LLMs, as recent studies [13] demonstrate, continue to struggle with tasks involving intricate spatial relationships, such as reasoning about cardinal directions. Despite improvements in performance, these limitations highlight the challenges LLMs face in achieving robust spatial reasoning comparable to human intuition. Additionally, computational resource demands for training integrated models can be substantial, necessitating algorithm optimization. Specific tasks that may enhance a spatial agent's capabilities in GIScience are still to be explored, as are the ways that LLMs could be used to improve the interpretability of spatial models.

However, it is worth considering whether pure LLMs are already obsolete in the rapidly evolving field of AI. Recent advances suggest that newer models and hybrid approaches may soon surpass LLMs in capability and efficiency, especially for more complex reasoning tasks. Retrieval-augmented generation (RAG), for example, combines the power of LLMs with a retrieval system to provide more accurate and contextualized information [18]. The integration of knowledge graphs into RAG systems can further enhance their ability to retrieve relevant facts and relationships for geospatial reasoning tasks [17]. By representing geospatial entities, attributes, and relationships in a structured knowledge graph, RAG systems can perform more targeted searches and provide LLMs with cleaner, more relevant context. Small language models (SLMs) aim to improve efficiency by focusing on critical aspects of data processing [26]. Hybrid AI approaches, which integrate knowledge-driven and data-driven methods, provide a more comprehensive understanding by combining logical reasoning with pattern recognition. In addition, models such as joint-embedding predictive architecture (JEPA) and its variants (e.g., I-JEPA, V-JEPA) are being investigated for their potential to better understand and predict complex systems [3]. These advances highlight a shift towards more specialized and efficient AI systems that could redefine the landscape of geospatial analysis and beyond.

3 What GIScience offers to LLMs

3.1 Conceptual framework

GIScience, with its inherently multimodal nature and its specialized analytical and conceptual frameworks [42], has much to contribute to the development of LLMs. While LLMs excel in processing unstructured textual data, GIScience brings unique reasoning capabilities, thematic diversity, and domain-specific methodologies that can significantly enhance LLM performance and expand their applicability. GIScience specializes in managing complex geographical phenomena, spatial, and spatiotemporal data relationships, including hierarchies, proximity, and network connectivity.

These concepts provide a robust foundation for improving the reasoning capabilities of LLMs. By embedding GIScience-inspired algorithms and models, LLMs can better interpret spatial hierarchies, such as spatial hierarchies (e.g., administrative subdivisions of countries or regions), networks (e.g., transportation or utility networks), (dis)similarity measures (e.g., Euclidean or network distance), neighborhood structures and proximity relationships. GIScience can support LLMs in understanding phenomena that vary over different spatial scales, bridging gaps between global and local perspectives to offer more context-aware interpretations, a critical aspect often overlooked in text-based systems. This integration allows LLMs to contextualize relationships across scales, such as distinguishing between local, regional, and global phenomena, which is essential for applications in urban planning, disaster management, or climate modeling GIScience's focus on spatial topology can also help LLMs avoid errors in spatial reasoning by enforcing rules about adjacency, containment, and connectivity.

3.2 Multimodal data and spatial representation in LLMs

Although modern LLMs are increasingly multimodal, integrating images, audio, and text, GIScience brings a unique form of multimodality through spatial data types (e.g., vector, raster, and spatiotemporal datasets). These formats provide structured representations of geographic phenomena that LLMs can use to build richer semantic models and a sound representation of real-world phenomena. Vector data, characterized by points, lines, and polygons, captures discrete objects such as landmarks, roads, and parcels. Raster data provides a pixel-based approach to representing continuous surfaces such as elevation and temperature gradients. GIScience also supports event-based models, allowing LLMs to reason about dynamic phenomena (e.g., natural disasters, urban growth) by incorporating temporal changes and spatial relationships into their predictions. Integrating these capabilities can enable LLMs to process, predict, and describe complex interactions between space and time more effectively. This spatiotemporal perspective offers LLMs a structured approach to integrating dynamic datasets, enabling these models to comprehend and predict changes across multiple spatial and temporal scales.

One of the challenges with LLMs is their tendency to hallucinate, generating plausible-sounding but incorrect outputs. GIScience offers well-established analytical models and validation techniques that can mitigate biases in the training data by providing external benchmarks for spatial reasoning. For example, spatial analysis methods, such as clustering, interpolation, or network analysis, can serve as a check against the interpretations generated by LLMs, ensuring their outputs are consistent with real-world spatial patterns. They can limit hallucinations where the AI 'imagines' patterns that reflect biases in the training data. Additionally, GIScience's focus on uncertainty modeling can enhance LLM outputs by helping them explicitly quantify and communicate the confidence levels of spatial predictions, addressing a critical gap in current LLM capabilities.

3.3 Intuitive spatial reasoning in LLMs

Alternative intuitive modeling approaches stemming from naive geography may offer new opportunities, as LLMs provide valuable support for interpreting informal and vague descriptions of geographic concepts. Naive geography principles, when paired with the generative capabilities of LLMs, could enhance their ability to intuitively process and simulate human-like spatial reasoning. For example, this synergy could enable LLMs to propose plausible geographic scenarios or alternatives based on vague descriptions, such as "a village near a river in a mountainous area," helping to generate more context-aware analyses. Furthermore, integrating folk geographic knowledge with LLMs could democratize spatial analysis by allowing non-experts to interact with complex geospatial tools through natural language interfaces.

GIScience provides analytical models that are essential for checking the validity of the interpretations of spatial data made by the LLM. These models can ensure that the conclusions drawn from the spatial analysis are reliable and robust. They can limit hallucinations where the AI 'imagines' patterns that reflect biases in the training data. In addition, GIScience provides a conceptual framework for defining analytical purposes that helps LLMs understand the context and goals of spatial analysis.

An important contribution of GIScience to LLMs is the provision of a 'spatial language'. This language includes terminologies, concepts, and contexts specific to spatial information [32] and enriches LLMs with spatially relevant vocabulary and expressions for features and relationships between one feature and another. An emerging field called Geographic Question Answering (GeoQA) provides methods for answering complex geographic questions that go beyond simple fact retrieval [43]. While GeoQA research builds upon developments in spatial knowledge graphs within GeoAI [29], both knowledge graphs and LLMs contribute distinct yet complementary approaches to spatial reasoning. Knowledge graphs provide structured representations of spatial relationships through explicit connections that can support both symbolic reasoning and embedding-based methods. These graph-based approaches enable explicit representation of spatial relationships, while LLMs learn to encode spatial semantic information through their training on large text corpora. The intersection of these approaches suggests potential benefits: knowledge graphs offer structured, verifiable spatial relationships, while LLMs provide flexible natural language understanding and generation capabilities.

3.4 Hybrid approaches

GIScience provides methodologies that can complement LLMs in hybrid AI approaches, such as integrating rule-based spatial reasoning with data-driven techniques. For example, spatial agents informed by GIScience can serve as intermediaries, processing geospatial data and feeding structured inputs into LLMs. This collaboration allows LLMs to leverage the strengths of GIScience without requiring exhaustive retraining on specialized datasets. However, the effective integration of these approaches for GIScience applications remains an active area of research requiring rigorous empirical evaluation to validate their comparative and combined effectiveness in specific use cases [52]. GIScience enables LLMs to integrate spatial querying and reasoning, topological modeling, and spatial statistics to promote natural language explanations that are spatially relevant [61]. GIScience can help LLMs achieve more sophisticated geospatial reasoning, allowing them to resolve ambiguities in spatial queries, such as identifying spatially similar regions or inferring implicit spatial relationships. This can also support the identification of previously unnoticed spatial patterns. This integration paves the way for interdisciplinary collaboration, bringing together experts from different fields such as urban planning, environmental science, and public health. These fields require specialized knowledge and reasoning that can be embedded into LLM training processes. Such collaborations could lead to innovative solutions to complex spatial problems that utilize the analytical capabilities of LLMs.

3.5 Geospatial learning in LLM training

By incorporating GIScience datasets and methodologies, LLMs can move beyond generic language-based training paradigms to include domain-specific geospatial learning. Training LLMs on synthetic geospatial datasets generated through GIScience principles enables them to generalize better to real-world scenarios, particularly in underrepresented regions or phenomena where data scarcity is a challenge. Furthermore, GIScience methodologies for generating synthetic spatial datasets can enhance the diversity and robustness of training data, improving model performance and resilience.

By integrating GIScience concepts, reasoning capabilities, and thematic diversity, LLMs can gain a more nuanced understanding of the spatial and temporal dimensions of the world. This synergy opens up new possibilities for hybrid systems that combine the strengths of GIScience and LLMs, advancing both fields and enabling applications that were previously unattainable. GIScience's unique contribution lies in its ability to bridge the gap between abstract language-based reasoning and the concrete, spatially explicit nature of real-world phenomena, making it an indispensable partner in the evolution of LLMs. Ultimately, this collaboration underscores GIScience's unique role as a bridge between abstract language-based reasoning and the concrete, spatially explicit nature of real-world phenomena, making it an indispensable partner in the evolution of LLMs.

4 Societal issues

Any integration of GIScience and LLMs should include a robust consideration of societal issues. While the incorporation of LLMs into GIScience or the enhancement of LLMs with GIScience could offer significant advances, they also pose ethical, social, and environmental challenges. One of the main concerns with LLMs is the potential bias or lack of trust in their outputs. These models are trained on large datasets that can contain biases as inputs, leading to the reinforcement of stereotypes and unfair treatment of certain groups. Additionally, spatial data often lacks context, which can lead to misinterpretations and skewed conclusions. When integrated into GIScience, these biases can manifest themselves in location-based services and analyses, leading to discriminatory practices. For example, when an LLM is used to generate content about different neighborhoods, it can perpetuate existing social biases, for instance, linking minorities to crime or characterizing safe neighborhoods as white. Algorithms can direct police to "over police" areas or suggest the denial of public services to the poor. The integration of LLMs and GIScience also can affect the trustworthiness of the information generated. LLMs can produce plausible but incorrect output, leading to the spread of misinformation. It can overemphasize crime in a community or underemphasize the impacts of climate change on a particular country. Misinformation can include inaccurate maps or flawed spatial analysis, which can mislead decision-makers and the public. Li et al. [37] argue that "Interpretability ... makes a GeoAI model transparent and inherently more trustworthy." That's debatable since it conflates social acceptance with interpretability and implies that all audiences can be satisfied with transparency alone or singular interpretations or explanations [2]. More importantly, trustworthiness in GeoAI should account for the unique characteristics of geospatial data that span multiple spatial, temporal, and thematic dimensions, although one could employ insights from Volunteered Geographic Information [20], in elevating the role of the contributor (as opposed to the contribution) and thus assessing the trustworthiness of a model. For example, how well does the modeler know the location they are modeling or trust the spatial output of an LLM?

To mitigate some biases, one could use diverse and representative datasets, conduct bias audits, and engage impacted groups in the model development process. Appropriate bias mitigation could include (1) expanding the distribution of geographic data to new examples, (2) removing inappropriate geographic data and generating new data, or (3) identifying hidden data and integrating it effectively. One could attempt to quantify the differences in a model's behavior across various demographic groups [14]. Equally or even more importantly, actively involving affected communities can help to build more equitable geospatial solutions [53]. As LLMs now incorporate visual capabilities, their evaluation should consider prior knowledge of the specific tasks they are capable of performing, or at least a reference for doing so (e.g., landmarks identification, image classification and segmentation, change detection), and the diversity of geographic data sources to be processed [52]. Incorporating human expertise and participatory approaches in the validation process is crucial to ensure the accuracy, reliability, and trustworthiness of models dealing with geographic data [53]. This implies active design processes, iterative interactions with local people, and collaborative crowdsourcing platforms to gather feedback and validations. This should ensure that the insights derived from LLMs are both meaningful and unbiased, providing valuable support for decision-making in various geographic applications.

The training of LLMs requires considerable computing resources, which leads to significant energy consumption and environmental impact. Integration with GIScience, which also involves extensive data processing, exacerbates this issue. The carbon footprint during training and inference of these models is a growing concern, necessitating the development of more energy-efficient algorithms and sustainable practices in AI research [55].

The use of advanced technologies, such as LLMs and the applications of GIScience, can widen the digital divide. Access to these tools and the benefits they provide can be unevenly distributed due to the high costs and technical expertise required. This can limit the ability of marginalized communities to use these technologies, exacerbating existing inequalities. It is crucial to make a concerted effort to democratize access to these technologies and ensure that their benefits are distributed more equitably across different population groups. People can be sensitized to the differential access to LLMs depending on location—especially as LLMs become increasingly costly—and to the geopolitical economics of LLMs that concentrate wealth in the Global North. Integrating LLMs with GIScience also requires a focus on expertise and education to ensure effective use. Training for GIS professionals and LLM developers is essential to understanding geospatial data nuances and the capabilities of LLMs. Interdisciplinary collaboration between geography, data science, and AI experts can enhance application robustness. For example, human geography contains thousands of articles on spatial bias and integrating the lived experience of non-experts. Democratizing tools and providing access to LLM-powered GIS applications empower a broader user base, generating better models and fostering inclusive participation in geospatial decision-making. Although LLMs are trained on large datasets, the diversity and complexity of geospatial data and queries in GIScience far exceed their training data [8]. This discrepancy can lead to errors and unreliable answers when LLMs encounter new or unexpected geospatial queries. For example, an LLM may have difficulty accurately interpreting complex spatial relationships or temporal dynamics that were not well represented in its training data. This limitation can affect the reliability of geospatial analyses and decision-making processes that rely on LLMs. To address this issue, continuous updates to training data and models are required, as well as the development of specialized techniques to effectively handle queries that lie outside the domain of geospatial data.

The integration of GISience and LLMs raises additional societal issues, particularly in the areas of privacy and surveillance, transparency and accountability, and influence on public opinion and behavior [56]. While concerns about privacy and surveillance predate LLMs, the integration of these technologies amplifies risks in unique ways. By processing large amounts of location-based data, people's movements can be closely tracked, leading to privacy concerns and potential infringements of civil liberties infringements through advanced surveillance systems. LLMs, with their ability to facilitate access via natural language queries, make it easier for users to extract sensitive insights from geospatial data, potentially lowering the barrier for misuse or intrusive analysis. Ethical issues arise from the use of personal data without informed consent and the risk of misuse for malicious purposes such as disinformation campaigns or unauthorized surveillance [5]. The complexity of LLM algorithms and the way information is represented across multiple layers of neural architectures makes decision-making processes opaque, undermining trust and complicating accountability for errors or negative outcomes. In addition, the potentially widespread dissemination of information through LLMs, amplified by easy access to LLMs such as ChatGPT for non-technical users, combined with geospatial data can significantly influence public opinion and behavior, raising ethical questions about accuracy, objectivity, manipulation, and consent.

5 Discussion

Generalizability and reproducibility of LLM-supported spatial research in GIScience are critical but nuanced concepts that deserve careful consideration. Generalizability refers to the ability of LLMs to effectively interpret, model, and analyze spatial data across different geographic datasets, while reproducibility ensures that spatial research results can be consistently validated under similar conditions. These fundamental requirements become even more complex in the modern AI landscape.

However, as highlighted by La Malfa et al. [33], reproducibility faces significant challenges in the context of proprietary Language-Models-as-a-Service (LMaaS). These challenges are particularly relevant for GIScience research, where geographic data is often dynamic and highly contextualized. According to Goodchild & Li [21] and Nichols et al. [47], reproducibility should not be viewed as a binary outcome, but as an evolving process where results may exhibit 'weak reproducibility', i.e., varying degrees of reproducibility in different spatial and temporal contexts. This perspective acknowledges that reproducibility in spatial research spans a spectrum, with results potentially more replicable in nearby areas than in more distant areas, consistent with the principle of spatial dependence.

This inherent geographic complexity is further complicated by the technical challenges of working with commercial LLM services. Proprietary cloud-based LLMs frequently modify, deprecate, or remove services without notice, complicating the replication of georeferenced experiments. This evolving nature of LMS not only hinders the validation of geospatial research results but also compromises their robustness and long-term integration into decision-making frameworks that rely on spatial analysis and evaluation.

The persistence of challenges related to generalizability, such as overfitting to specific datasets, can lead to models that perform poorly when applied to new or varied spatial contexts. In geospatial contexts, this can manifest itself in models misinterpreting spatial patterns or failing to account for variations in the local environment or cultural background. Similarly, the lack of reproducibility due to variations in model training or data preprocessing can hinder the validation of results and methods in GeoAI.

These problems are further amplified by the principle of spatial heterogeneity, which states that study outcomes are inherently variable depending on the spatial and temporal bounds of the data. In the context of LLMs, this suggests that replicability must account for spatial variation in language model predictions and parameter tuning across different spatial datasets [38]. The concept of weak replicability is crucial here, as models retain generalizable structures but adjust their results based on location-specific data. Metrics of replicability could be explored to assess the stability of LLM-derived findings in nearby and distant regions, supporting the notion that replicability is a variable and not a binary property when spatial heterogeneity is taken into account.

These issues may impede the adoption of LLMs in decision-making processes, and potentially limit progress towards meaningful interoperability of LLMs and GIScience. To overcome the challenges of generalizability and reproducibility in LLM-supported spatial research, several directions can be pursued: develop spatially-aware standardized methods for training and testing models; use diverse geographic datasets that represent varying spatial scales and regions to enhance model adaptability; acknowledge the diversity of spatial data sources, including institutional datasets and increasingly heterogeneous, crowdsourced data that may vary in quality but provide valuable contextual information, promote transparency through open spatial data and code sharing; implement robust cross-validation techniques to ensure unbiased results; establish benchmarking frameworks specifically tailored to geospatial tasks, such as spatial prediction, routing, or spatial feature extraction; and foster a culture of continuous evaluation and adaptation to keep models updated with new spatial data, including institutional and crowdsourced resources.

Leveraging open LLMs is particularly desirable for GIScience applications, as they allow for greater transparency and flexibility compared to proprietary spatial models. However, even open LLMs often lack full transparency unless they also release their training data, as seen in efforts like those of AllenAI [54]. Without access to training data that covers diverse geographic contexts, critical questions about spatial biases, limitations, and model generalization capabilities remain unresolved.

Focusing on these strategies, including advocating for more transparent open LLM initiatives, can enhance the reliability of LLM-supported spatial research and contribute positively to the sound integration of GIScience and LLMs.

While LLMs demonstrate promise in zero-shot or few-shot scenarios for text-based tasks, they often underperform in complex multimodal tasks, highlighting the need for tailored approaches for different geospatial domains [44]. Furthermore, developing a multimodal LLM capable of reasoning across various geospatial data types through effective alignments remains a major challenge, alongside unique risks related to ethical considerations, data privacy, and potential biases in geospatial decision-making.

Decades ago, the combination of CAD and databases led to the development of GIS. At that time, three approaches competed with each other: the integration of CAD systems with databases (e.g., AutoCAD with attributes), the extension of databases with spatial functions (e.g., Oracle with spatial features) and the development of new GIS systems from scratch (e.g., ArcGIS, QGIS). Today we are seeing a similar development in GIS and LLMs, which may expand into what could be called "Natural GIS", "Generative GIS", or "LLMGIS". This new development points to a competition between augmenting existing GIS with LLMs (e.g., ArcGIS with AI assistants), integrating LLMs with existing tools (e.g., ChatGPT with Python and shapefiles), and creating entirely new systems from scratch.

Development and educational research on LLM-based innovations are still in their early stages, and while automated learning processes could be favored when dealing with geographic concepts and applications, there is a need to integrate ethical issues for successful integration into learning programs [63]. This is particularly relevant in the context of GIScience, as its rapidly growing development in the era of major environmental and societal challenges and the increasing need for open science requires a rethinking of its educational foundations. In addition, the continuous evolution of supporting technologies, from novel interfaces to location-based services, is an important driver that will play an essential role in the development of LLM and the integration of GIScience and should be taken into account.

We note that although we have focused here on LLMs, there is increasing interest in Small Language Models (SLMs), which are trained on much smaller, typically domain-specific and highly curated datasets, and which are then experts in particular domains of focus. These models have the advantage of being less costly to train (and run) and less likely to hallucinate given the nature of their training data. Their narrow nature can be mitigated by a community of SLMs, each focused on specific domains of expertise and perhaps integrated via a more general purpose language model that can help route the prompts to the appropriate SLMs.

6 Towards a research agenda

The potential convergence of LLMs and GIScience offers a promising avenue for advancing geospatial data analysis and natural language processing when reasoning about geospatial texts, as well as improving interaction with non-experts and citizens. This paper explored the mutual benefits of integrating the linguistic skills of LLMs with the expertise of GIScience in spatial data, highlighting their potential to transform the processing, analysis, and visualization of geospatial data. However, this integration also brings challenges, including the risk of bias, misinformation, and environmental impact. By addressing these issues and promoting responsible use of AI, the synergy between LLMs and GIScience can lead to innovative solutions in various fields (e.g., urban planning) and improve our ability to understand and interact with complex spatiotemporal phenomena. The search for a research agenda should aim to enhance GIScience through the integration of LLMs by focusing on several key areas. A primary challenge lies in exploring how to integrate LLMs with external reasoners, especially given their well-known limitations in complex reasoning, including spatial reasoning. The focus is on improving the NLP capabilities of LLMs to analyze unstructured geospatial documents and extract complex spatial relationships in different contexts. Key future directions include establishing benchmarks for evaluating their performance across scales in both space and time, which remains essential for ensuring the reliability and scalability of their applications in GIScience.

This includes the development of prompting techniques specifically tailored to GIScience use-cases, such a the integration of domain-specific knowledge into the prompts to improve inference accuracy. For example, recent research on geoentity-type constrained knowledge graph embeddings has shown promise in improving the inference of spatial relationships [24]. Promising advancements include the use of graph-based structures that capture spatial relationships more effectively, the inclusion of semantic attributes to enrich spatial context, the identification of toponyms from large textual sources in social media, and the classification of geographic entities based on their spatial, temporal, and functional characteristics. LLMs can be improved for GIScience by integrating geospatial alignments where spatial relationships, coordinates, and topological structures are embedded directly into the training process of the model. This alignment enables LLMs to better capture spatial dependencies, reason about geographic entities, and provide more accurate insights for location-based tasks.

The fusion of LLMs with knowledge graphs and ontologies offers a significant opportunity to enhance spatial reasoning by anchoring text-based insights in a geospatial knowledge framework [29]. Additionally, the use of multimodal LLMs that can process text, images, and geospatial vector data simultaneously presents a compelling avenue for analyzing hybrid datasets commonly encountered in GIScience. Further exploration is required to address specific challenges, such as ambiguity in spatial language, representing dynamic geospatial phenomena, and ensuring that models can generalize across regions with different geographic, social, and cultural contexts. Future research should also investigate methods for integrating fine-grained temporal analyses that enable LLMs to discern patterns over time and generate predictions aligned with dynamic spatial processes.

LLMs should generate spatial code by leveraging geospatial libraries, such as GeoPandas and PostGIS while ensuring accurate handling of spatial data types. Establishing open datasets and benchmarks for spatial NLP tasks will be instrumental in accelerating research and fostering collaborations between the NLP and GIScience communities. Developing a multimodal LLM capable of reasoning across various geospatial data types through effective alignments remains a major challenge in enhancing the capabilities of LLMs to process text, imagery, and geographic data, improving spatial understanding and visualization.

The agenda should include developing frameworks that combine LLMs with knowledge graphs for better spatial reasoning while carefully considering the role of synthetic datasets in GIScience research. If well-designed, such datasets can provide controlled environments for model training, evaluation, and stress testing beyond typical real-world distributions. However, their generation must be guided by rigorous methodological frameworks that address potential biases and ensure a meaningful representation of diverse spatial phenomena. Importantly, synthetic data creation requires a priori knowledge, whether through data selection, resampling, or theoretical modeling. This process relies on the expertise from GIScientists rather than solely on machine learning. LLMs, lacking awareness of modeling purposes, cannot assess biases without this theoretical grounding. Particular attention must also be paid to preventing the reinforcement of existing societal biases in spatial data, such as demographic stereotypes linked to specific geographic areas. Effective synthetic data generation depends on deliberate bias mitigation strategies rather than merely replicating real-world distributions. This requires unified frameworks that provide appropriate guidelines for synthetic data generation [40].

To achieve effective and robust integration of LLMs and spatial knowledge graphs, a comprehensive, neuro-symbolic approach is essential, especially given the inherent challenges LLMs face in spatial reasoning [59]. First, training LLMs should explicitly incorporate spatial knowledge graph representations to provide them with inherent geospatial understanding. This can be effectively achieved by embedding entity-relationship triples from geospatial knowledge datasets directly into the model architecture, or through targeted fine-tuning steps. Graph Neural Networks (GNNs) offer a powerful technique for encoding spatial relationships from knowledge graphs into vector embeddings. These embeddings can then be seamlessly integrated into the input layers of the LLM, enabling the model to inherently capture, process, and reason with spatial relationships [66]. Furthermore, the semantic richness of knowledge graphs greatly improves the understanding, querying, and analysis of geospatial information, particularly when dealing with Geospatial Big Data (GBD). The integration of GBD with knowledge graphs, as highlighted by Wu et al. [60], provides semantic benefits for spatial reasoning, entity disambiguation, and fusion of multi-source data. This integration facilitates semantic modeling by extracting entities, attributes, and relationships from GBD and mapping them to knowledge graph concepts, enabling intelligent querying through natural language and complex statements, and enhancing geospatial reasoning by uncovering implicit spatial relationships and patterns [60].

Second, attention mechanisms within LLMs should be strategically adapted to prioritize spatial relationships. This may mean incorporating spatial biases into attention layers, allowing the model to weigh connections between entities based on their spatial proximity, topological relationships, or even directional relationships within the knowledge graph [58]. This is particularly relevant in geospatial contexts where spatial adjacency, connectivity, and even geometric features are crucial for reasoning, as explored in recent studies on spatially explicit machine learning and GeoAI [42], and further enhanced by incorporating geometric features in knowledge graph embeddings, which improves prediction accuracy for both geo-entities and spatial relationships [24]. In addition, constraining knowledge graph embeddings by geoentity-types enables more accurate prediction of spatial relations in natural language, effectively capturing both spatial and semantic relationships [25].

At the processing level, dynamic querying of knowledge graphs during inference is not only beneficial but essential. This allows LLMs to access and utilize up-to-date spatial facts and reason about evolving spatial patterns in real time. Advanced methods for dynamic knowledge graph querying, such as Retrieval-Augmented Generation (RAG), can empower LLMs to retrieve highly relevant spatial information on demand, effectively augmenting their reasoning process and enabling them to handle the dynamic nature of geographic information [69]. Beyond real-time analysis, LLMs can also be used to systematically extract and visualize spatial relationships from historical narratives by using knowledge graphs to map the intricate relationships between places and entities, providing new insights for the exploration of environmental data descriptions [23]. This enhanced neuro-symbolic integration strategy, which includes advanced training, geometrically and semantically informed embeddings, and dynamic inference enhancements, can enable LLMs to generate novel and in-depth geospatial insights by effectively synthesizing spatial knowledge from diverse and dynamic sources. Ultimately, this integrated approach aims to bridge the inherent gap between traditional, rule-based GIS-based spatial reasoning and the emerging field of cutting-edge AI-driven geospatial analysis, paving the way for the development of more sophisticated, adaptable, and insightful applications across the spectrum of GIScience.

Additionally, we advocate for user-centric designs that exploit LLMs to create tools that make GIScience technologies available to non-experts, with tailored approaches for different geospatial domains, fostering interdisciplinary collaboration to tackle complex spatial problems. Addressing challenges related to data quality, the complexity of spatial reasoning, and the computational requirements of integration is crucial. Finally, soundly recognizing the importance of fairness and accountability, among other elements of ethics, bolsters the acceptance of the models. The overall agenda should encourage the exploration of LLMs' roles in enhancing advanced spatial operations and developing autonomous systems capable of learning spatial concepts through interactions with their environment. Finally, blending GIScience with LLMs can contribute to the emerging concept of a World Model [64] that seamlessly integrates geospatial knowledge, linguistic, and conceptual insights. This aligns closely with the earlier discussed synergy between LLMs and GIScience, enhancing the model's ability to reason about, predict, and understand complex spatial relationships and geographical phenomena, while also improving geospatial analysis and interaction with spatiotemporal data.

7 Conclusion

Researchers are currently evaluating the impact of LLMs on various scientific fields and, in particular, investigating how these AI systems can revolutionize approaches to GIScience. The potential of LLMs to provide innovative methods for analyzing geographic phenomena is a topic of great interest. However, it is crucial to recognize the difference between GIScience as an established scientific discipline and LLMs as advanced computational and algorithmic tools. This raises an initial food for thought: We are essentially comparing a scientific field with a technological tool, which leads us to question the implications of this comparison. Furthermore, it is worth reflecting on the historical parallels between GIScience and the current state of LLMs. Initially, GIScience emerged as a computational paradigm that focused primarily on the integration, processing, and analysis of cartographic data. Similarly, LLMs are currently perceived primarily as computational resources rather than a scientific discipline in their own right. This brings us to a second key question: could LLMs develop into a scientific field in their own right?

Consideration of this development requires a deeper examination of the nature of LLMs and their potential evolution within the scientific landscape. Although LLMs fundamentally operate as computational tools, their capabilities go beyond traditional computational functions. Unlike previous narrow AI systems, they have the ability to comprehend, generate, and manipulate language on virtually any topic and at a sophisticated level, enabling them to take on complex tasks previously reserved exclusively for human cognition. As LLMs evolve and diversify their applications, they could gradually outgrow their current status as mere tools. Their profound impact on various scientific fields, including GIScience, suggests the emergence of a new interdisciplinary field at the interface between artificial intelligence and domain-specific sciences. This hypothetical field would not only harness the computational power of LLMs but also integrate domain-specific knowledge and methods to foster a symbiotic relationship between AI technology and scientific research. While LLMs currently serve as powerful computational devices, their potential to evolve into scientific domains in their own right offers intriguing possibilities for the future of scientific research and discovery. Exploring these possibilities requires a nuanced understanding of the dynamic interplay between technology and scientific research and paves the way for innovative approaches to knowledge creation and exploration in the years to come.

Author contributions

NVdW, CC, and LDS developed the framework, drafted the original manuscript, coordinated its organization, and oversaw the revision process. AGC, HH, SS, RS, and ST contributed to the refinement of ideas and participated in the review and editing of the manuscript. All authors have read and approved the final version.

Data availability statement

No data was used for the work described in this article. The figure presented in this manuscript is available upon reasonable request to the corresponding author.

Acknowledgments

AGC gratefully acknowledges financial support from: the Fundamental Research priority area of The Alan Turing Institute; the Special Funds of Tongji University, Shanghai, for the Sino-German Cooperation 2.0 Strategy; the Economic and Social Research Council (ESRC) under grant ES/W003473/1.

References

[1] AGRAWAL, A., MCHALE, J., AND OETTL, A. Artificial intelligence and scientific discovery: A model of prioritized search. *Research Policy 53*, 5 (2024), 104989. doi:10.1016/j.respol.2024.104989.

[2] ARRIETA, A. B., DÍAZ-RODRÍGUEZ, N., DEL SER, J., BENNETOT, A., TABIK, S., BARBADO, A., GARCÍA, S., GIL-LÓPEZ, S., MOLINA, D., BENJAMINS, R., ET AL. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion 58* (2020), 82–115. doi:10.1016/j.inffus.2019.12.012.

[3] ASSRAN, M., QUENTIN, D., ISHAN, M., BOJANOWSKI, P., CHEUNG, B., BALLAS, N., BACH, F., EHSAN, R. M., SCHROECKER, Y., BROCK, A., ET AL. Self-supervised learning from images with a joint-embedding predictive architecture. *arXiv preprint arXiv:2301.08243* (2023).

[4] BANKER, S., CHATTERJEE, P., MISHRA, H., MOREWEDGE, C. K., AND SOCHER, R. Machine-assisted social psychology hypothesis generation. *American Psychologist 79*, 6 (2024), 789. doi:10.1037/amp0001156.

[5] BARMAN, D., GUO, Z., AND CONLAN, O. The dark side of language models: Exploring the potential of LLMs in multimedia disinformation generation and dissemination. *Machine Learning with Applications 16* (2024), 100545. doi:10.1016/j.mlwa.2024.100545.

[6] BENDER, E. M., GEBRU, T., MCMILLAN-MAJOR, A., AND SHMITCHELL, S. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), pp. 610–623. doi:10.1145/3442188.3445922.

[7] BENDER, E. M., AND KOLLER, A. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020), pp. 5185–5198. doi:10.18653/v1/2020.acl-main.463.

[8] BHANDARI, P., ANASTASOPOULOS, A., AND PFOSER, D. Are large language models geospatially knowledgeable? In *Proceedings of the International Conference on Advances in Geographic Information Systems* (2023). doi:10.1145/3589132.3625573.

[9] BOMMASANI, R., HUDSON, D. A., ADELI, E., ALTMAN, R., ARORA, S., VON ARX, S., BERNSTEIN, M. S., BOHG, J., BOSSELUT, A., BRUNSKILL, E., ET AL. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2022).

[10] BROWN, T. B., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G., ASKELL, A., ET AL. Language models are few-shot learners. *Advances in Neural Information Processing Systems 33* (2020).

[11] CHEN, J., XU, W., CAO, H., ZHANG, S., FAN, K., HUANG, J., AND WANG, S. Multimodal road network generation based on large language model. *arXiv preprint arXiv:2404.06227* (2024).

[12] CHEN, M., TWOREK, J., JUN, H., YUAN, Q., PINTO, H. P. D. O., KAPLAN, J., EDWARDS, H., BURDA, Y., JOSEPH, N., BROCKMAN, G., ET AL. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).

[13] COHN, A. G., AND BLACKWELL, R. E. Evaluating the ability of large language models to reason about cardinal directions. *arXiv preprint arXiv:2406.16528* (2024).

[14] CZARNOWSKA, P., VYAS, Y., AND SHAH, K. Quantifying social biases in NLP: A generalization and empirical comparison of extrinsic fairness metrics. *Transactions of the Association for Computational Linguistics 9* (2021), 1249–1267. doi:10.1162/tacl_a_00425.

[15] DU, M., WU, B., LI, Z., HUANG, X., AND WEI, Z. Embspatial-bench: Benchmarking spatial understanding for embodied tasks with large vision-language models. *arXiv preprint arXiv:2406.05756* (2024).

[16] EGENHOFER, M. J., AND FRANZOSA, R. D. Point-set topological spatial relations. *International Journal of Geographical Information Systems 5*, 2 (1992), 161–174. doi:10.1080/02693799208901902.

[17] FENG, Y., DING, L., AND XIAO, G. GeoQAMap – Geographic Question Answering with Maps Leveraging LLM and Open Knowledge Base. In *12th International Conference on Geographic Information Science (GIScience 2023)* (Dagstuhl, Germany, 2023), R. Beecham, J. A. Long, D. Smith, Q. Zhao, and S. Wise, Eds., vol. 277 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 28:1–28:7. doi:10.4230/LIPIcs.GIScience.2023.28.

[18] GAO, Y., XIONG, Y., GAO, X., HE, T., LI, F., YAN, B., XU, R., AND ZHANG, Z. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997* (2024).

[19] GOODCHILD, M. F. Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, 1 (2010), 3–20. doi:10.5311/JOSIS.2010.1.2.

[20] GOODCHILD, M. F., AND LI, L. Assuring the quality of volunteered geographic information. *Spatial Statistics 1* (2012), 110–120. doi:10.1016/j.spasta.2012.03.002.

[21] GOODCHILD, M. F., AND LI, W. Replication across space and time must be weak in the social and environmental sciences. *Proceedings of the National Academy of Sciences 118*, 35 (2021), e2015759118. doi:10.1073/pnas.2015759118.

[22] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A., AND BENGIO, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems* (2014), vol. 27.

[23] HARIS, E., COHN, A. G., AND STELL, J. G. Exploring spatial representations in the historical lake district texts with llm-based relation extraction. *arXiv preprint arXiv:2406.14336* (2024).

[24] HU, L., LI, W., XU, J., AND ZHU, Y. GeoEntity-type constrained knowledge graph embedding for predicting natural-language spatial relations. *International Journal of Geographical Information Science* (2024), 1–24. doi:10.1080/13658816.2024.2326057.

[25] HU, L., LI, W., AND ZHU, Y. Geometric feature enhanced knowledge graph embedding and spatial reasoning. In *Proceedings of the International Workshop on AI for Geographic Knowledge Discovery* (2024), pp. 50–53. doi:10.1145/3639365.3640719.

[26] HU, S., YUGE, T., XU, H., HU, M., LIU, P., WANG, X., HUANG, Q., LIU, X., QI, Y., QIN, J., ET AL. Unveiling the potential of small language models with scalable training strategies. *arXiv preprint arXiv:2404.06395* (2024).

[27] IBRAHIM, N., ABOULELA, S., IBRAHIM, A., MORAD, A., EL-RAMLY, M., ABDULRAHIM, H., AND GAMAL, Y. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): Models, evaluation metrics, benchmarks, and challenges. *Discover Artificial Intelligence 4*, 1 (2024), 76. doi:10.1007/s44163-024-00422-y.

[28] JANOWICZ, K. Philosophical foundations of GeoAI: Exploring sustainability, diversity, and bias in GeoAI and spatial data science. In *Handbook of Geospatial Artificial Intelligence*. Springer, 2023, pp. 26–42. doi:10.1007/978-3-031-26103-8_2.

[29] JANOWICZ, K., GAO, S., MCKENZIE, G., HU, Y., AND BHADURI, B. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science 34*, 4 (2020), 625–636. doi:10.1080/13658816.2019.1684500.

[30] JENNINGS, N. R., SYCARA, K., AND WOOLDRIDGE, M. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems 1* (1998), 7–38. doi:10.1023/A:1010090405266.

[31] JI, Y., AND GAO, S. Evaluating the Effectiveness of Large Language Models in Representing Textual Descriptions of Geometry and Spatial Relations. In *12th International Conference on Geographic Information Science (GIScience 2023)* (Dagstuhl, Germany, 2023), R. Beecham, J. A. Long, D. Smith, Q. Zhao, and S. Wise, Eds., vol. 277 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 43:1–43:6. doi:10.4230/LIPIcs.GIScience.2023.43.

[32] KUHN, W. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science 26*, 12 (2012), 2267–2276. doi:10.1080/13658816.2012.722637.

[33] LA MALFA, E., PETROV, A., FRIEDER, S., OREN, I., WANG, A., PAPERNOT, N., HASHIMOTO, T., BLITZER, J., DE SA, C., RAFFEL, C., ET AL. Language-models-as-a-service: Overview of a new paradigm and its challenges. *Journal of Artificial Intelligence Research 80* (2024), 1497–1523. doi:10.1613/jair.1.15305.

[34] LAKE, B. M., SALAKHUTDINOV, R., AND TENENBAUM, J. B. Human-level concept learning through probabilistic program induction. *Science 350*, 6266 (2015), 1332–1338. doi:10.1126/science.aab3050.

[35] LI, F., HOGG, D. C., AND COHN, A. G. Advancing spatial reasoning in large language models: An in-depth evaluation and enhancement using the stepgame benchmark. *Proceedings of the AAAI Conference on Artificial Intelligence 38*, 17 (2024), 18500–18507. doi:10.1609/aaai.v38i17.28870.

[36] LI, W. GeoAI: Where machine learning and big data converge in GIScience. *Journal of Spatial Information Science 20* (2020), 71–77. doi:10.5311/JOSIS.2020.20.666.

[37] LI, W., ARUNDEL, S., GAO, S., GOODCHILD, M. F., HU, Y., JANOWICZ, K., KARIMZADEH, M., MAVROMMATI, G., MARTIN, Y., SCHUBERT, E., ET AL. GeoAI for science and the science of GeoAI. *Journal of Spatial Information Science*, 29 (2024), 1–17. doi:10.5311/JOSIS.2024.29.242.

[38] LI, W., HSU, C.-Y., WANG, S., AND KEDRON, P. GeoAI reproducibility and replicability: A computational and spatial perspective. *Annals of the American Association of Geographers 114*, 9 (2024), 2085–2103. doi:10.1080/24694452.2023.2268992.

[39] LI, Z., AND NING, H. Autonomous GIS: The next-generation AI-powered GIS. *International Journal of Digital Earth 16*, 2 (2023), 4668–4686. doi:10.1080/17538947.2023.2277892.

[40] LONG, L., WANG, R., XIAO, X., FU, H., HUNG, H., CHEN, I., SHI, W., YUAN, Y., LIANG, P. P., LIAO, Q., ET AL. On LLMs-driven synthetic data generation, curation, and evaluation: A survey. *arXiv preprint arXiv:2406.15126* (2024).

[41] LUDWIG, J., AND MULLAINATHAN, S. Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics 139*, 2 (2024), 751–827. doi:10.1093/qje/qjad051.

[42] MAI, G., HU, Y., GAO, S., JANOWICZ, K., YANG, B., AND CAI, L. Symbolic and subsymbolic GeoAI: Geospatial knowledge graphs and spatially explicit machine learning. *Transactions in GIS 26*, 8 (2022), 3118–3124. doi:10.1111/tgis.12997.

[43] MAI, G., JANOWICZ, K., ZHU, R., CAI, L., AND LAO, N. Geographic question answering: Challenges, uniqueness, classification, and future directions. *AGILE: GIScience Series 2* (2021), 8. doi:10.5194/agile-giss-2-8-2021.

[44] MAI, Y., HUANG, W., SUN, L., SCHUBERT, E., RAUBAL, M., ZHANG, Z., HE, Y., SUN, Z., AND LI, C. On the opportunities and challenges of foundation models for geoai. *ACM Transactions on Spatial Algorithms and Systems 10*, 2 (2024), 1–46. doi:10.1145/3636196.

[45] MANVI, R., KHANNA, S., MAI, G., YAO, X., AND JANOWICZ, K. GeoLLM: Extracting geospatial knowledge from large language models. *arXiv preprint arXiv:2310.06213* (2023).

[46] MARCUS, G. The next decade in AI: Four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177* (2020).

[47] NICHOLS, J. D., OLI, M. K., KENDALL, W. L., AND BOOMER, G. S. A better approach for dealing with reproducibility and replicability in science. *Proceedings of the National Academy of Sciences 118*, 7 (2021), e2100769118. doi:10.1073/pnas.2100769118.

[48] OPENSHAW, S. A spatial analysis research agenda. *Regional Research Laboratory Initiative Discussion Paper 3* (1990).

[49] RADFORD, A., KIM, J. W., HALLACY, C., RAMESH, A., GOH, G., AGARWAL, S., SASTRY, G., ASKELL, A., MISHKIN, P., CLARK, J., ET AL. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (2021), pp. 8748–8763.

[50] RAMESH, A., PAVLOV, M., GOH, G., GRAY, S., VOSS, C., RADFORD, A., CHEN, M., AND SUTSKEVER, I. Zero-shot text-to-image generation. In *Proceedings of the 38th International Conference on Machine Learning* (2021), pp. 8821–8831.

[51] RANDELL, D. A., CUI, Z., AND COHN, A. G. A spatial logic based on regions and connection. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning* (1992), pp. 165–176.

[52] ROBERTS, J., LÜDDECKE, T., SHEIKH, R., ATZMON, Y., SANH, V., SCIALOM, T., COHEN, U., GUPTA, D., HAMPTON, G., CHITTA, A., ET AL. Charting new territories: Exploring the geographic and geospatial capabilities of multimodal LLMs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), 554–563. doi:10.1109/CVPR55264.2024.00061.

[53] SIEBER, R., BRANDUSESCU, A., SANGIAMBUT, S., CORBETT, J., AND BUI, T. A.-T. What is civic participation in artificial intelligence? *Environment And Planning B: Urban Analytics and City Science* (2024), 23998083241296200. doi:10.1177/23998083241296200.

[54] SOLDAINI, L., KINNEY, R., BHAGIA, A., WANG, R., LIANG, C., LAM, D., BHATIA, V., LEMASTERS, M., BHARDWAJ, A., DODGE, J., ET AL. Dolma: An open corpus of three trillion tokens for language model pretraining research. *arXiv preprint arXiv:2402.00159* (2024).

[55] STRUBELL, E., GANESH, A., AND MCCALLUM, A. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243* (2019).

[56] TOKAYEV, K. J. Ethical implications of large language models: A multidimensional exploration of societal, economic, and technical concerns. *International Journal of Social Analytics 8*, 9 (2023), 17–33.

[57] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł., AND POLOSUKHIN, I. Attention is all you need. In *Advances in Neural Information Processing Systems* (2017), vol. 30.

[58] VELIČKOVIĆ, P., CUCURULL, G., CASANOVA, A., ROMERO, A., LIO, P., AND BENGIO, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[59] WANG, R., SUN, K., AND KUHN, J. A pipeline of neural-symbolic integration to enhance spatial reasoning in large language models. *arXiv preprint arXiv:2411.18564* (2024).

[60] WU, J., GAN, W., CHAO, H.-C., LI, Q., AND CHEN, Y. Geospatial big data: Survey and challenges. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2024). doi:10.1109/JSTARS.2024.3353025.

[61] XING, J., AND SIEBER, R. The challenges of integrating explainable artificial intelligence into geoai. *Transactions in GIS 27*, 3 (2023), 626–645. doi:10.1111/tgis.13042.

[62] XUE, H., PRAKASH VOUTHAROJA, B., AND SALIM, F. D. Leveraging language foundation models for human mobility forecasting. In *Conference on Advances in Geographic Information Systems* (2022), pp. 1–9. doi:10.1145/3557915.3561057.

[63] YAN, L., SHA, L., ZHAO, L., WANG, C., WANG, J., CHEN, B., LIU, Y., CHEN, S. S., PAN, J., AND ZHANG, J. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology 55*, 1 (2024), 90–112. doi:10.1111/bjet.13384.

[64] YILDIRIM, I., AND PAUL, L. From task structures to world models: what do LLMs know? *Trends in Cognitive Sciences 28*, 5 (2024), 404–415. doi:10.1016/j.tics.2024.02.008.

[65] ZHANG, Y., WEI, C., WU, S., CHU, W., CHEN, Y., FU, X., SHAO, Y., AND WANG, F. GeoGPT: Understanding and processing geospatial tasks through an autonomous GPT. *arXiv preprint arXiv:2307.07930* (2023).

[66] ZHANG, Z., HAN, X., LIU, Z., JIANG, X., SUN, M., AND LIU, Q. ERNIE: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129* (2019).

[67] ZHOU, Y., LIU, H., SRIVASTAVA, T., ZHOU, K., MOUTINHO RIBEIRO, J. P., BROWN, Z., DUAN, X., AND JAVED, S. Hypothesis generation with large language models. *arXiv preprint arXiv:2404.04326* (2024).

[68] ZHU, J., DANG, P., CAO, Y., LI, C., LIAO, L., LI, S.-H., YANG, Z., LU, Y., TANG, B.-F., XU, Y., ET AL. A flood knowledge-constrained large language model interactable with GIS: Enhancing public risk perception of floods. *International Journal of Geographical Information Science 38*, 4 (2024), 603–625. doi:10.1080/13658816.2023.2280763.

[69] ZHU, X., GUO, X., CAO, S., LI, S., AND GONG, J. StructuGraphRAG: Structured document-informed knowledge graphs for retrieval-augmented generation. In *Proceedings of the AAAI Symposium Series* (2024), vol. 4, pp. 242–251. doi:10.1609/aaaiss.v4i1.31798.