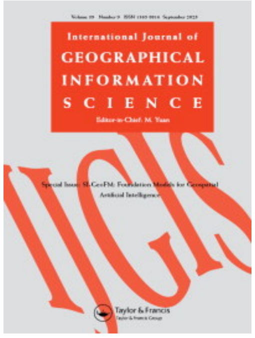

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

GeoFM: how will geo-foundation models reshape spatial data science and GeoAI?

Krzysztof Janowicz, Gengchen Mai, Weiming Huang, Rui Zhu, Ni Lao & Ling Cai

To cite this article: Krzysztof Janowicz, Gengchen Mai, Weiming Huang, Rui Zhu, Ni Lao & Ling Cai (2025) GeoFM: how will geo-foundation models reshape spatial data science and GeoAI?, International Journal of Geographical Information Science, 39:9, 1849-1865, DOI: 10.1080/13658816.2025.2543038

To link to this article: https://doi.org/10.1080/13658816.2025.2543038

Published online: 09 Aug 2025.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2025, VOL. 39, NO. 9, 1849–1865
https://doi.org/10.1080/13658816.2025.2543038

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

GeoFM: how will geo-foundation models reshape spatial data science and GeoAI?

Krzysztof Janowicz[a], Gengchen Mai[b], Weiming Huang[c,d,e], Rui Zhu[f], Ni Lao[g] and Ling Cai[h]

[a]Department of Geography and Regional Research, University of Vienna, Austria; [b]Department of Geography and the Environment, University of Texas at Austin, USA; [c]College of Computing and Data Science, Nanyang Technological University, Singapore; [d]Department of Physical Geography and Ecosystem Science, Lund University, Sweden; [e]School of Geography, University of Leeds, UK; [f]School of Geographical Sciences, University of Bristol, UK; [g]Google, USA; [h]ByteDance Inc., USA

ABSTRACT
The emerging field of geo-foundation models (GeoFM) has the potential to reshape GeoAI and spatial data science research, education, and practice. In this work, we motivate and define the term and put it into its historic context within GeoAI and spatial data science more broadly. Next, we review core datasets, models, and benchmarks. Based on this overview of the state-of-the-art, we introduce key research challenges for future GeoFM research, such as GeoAI scaling laws, geo-alignment of AI, truly multimodal GeoFM, and so on. Finally, we discuss potential risks of GeoFM research and outline the road ahead with a specific focus on the increasing role of international large-scale collaborations and the future of GeoAI and spatial data science education.

ARTICLE HISTORY
Received 30 July 2025
Accepted 30 July 2025

KEYWORDS
GeoAI; foundation models; spatially explicit machine learning; AI alignment

1. Introducing geo-foundation models

In a nutshell, *foundation models* are large and highly versatile AI models pre-trained on massive datasets that can be easily adapted to a wide range of downstream tasks across domain boundaries. Each of these terms is key here, so let us comment on them briefly. **(1)** Interestingly, *model size* is not just another hyperparameter that influences accuracy. Instead, it gives rise to *emerging* properties that only manifest once models grow beyond a certain (relative) size. For instance, according to Wei *et al.* (2022), *chain of thought* reasoning only emerges past approximately 100B parameters.[1] However, model size (e.g., parameter count) is not the only aspect where size matters. AI *scaling laws*, for example, empirically study how model performance improves in relation to parameter count, training data size, and required computing resources (Kaplan *et al.* 2020). **(2)** Until very recently, models were trained with a specific task in

CONTACT Krzysztof Janowicz krzysztof.janowicz@univie.ac.at

mind, e.g., land cover classification. In contrast, the same foundation model may be utilized across multiple tasks. For instance, such a *versatile* model can classify land cover in imagery, segment scenes, describe them in natural language, and suggest data engineering and analytical pipelines to process the now classified data further. **(3)** Finally, a key characteristic of foundation models is their *adaptability*. Today, these models perform on par with many prior single-purpose models with little *fine-tuning* and minimal (task-specific) training data (Brown *et al.* 2020). This adaptability is due to foundation models being pre-trained on massive amounts of highly *heterogeneous* data, even across modalities. Once these models are able to *generalize* and learn more abstract, relational patterns, they can be fine-tuned to a plethora of downstream tasks with little effort or even employed directly for novel tasks via few-shot learning or even zero-shot learning.

Given that foundation models (Bommasani *et al.* 2021) are, to some degree, task-agnostic, why do we need *geo-foundation models* (GeoFM) at all, and what exactly are they or will they be? First, foundation models can only generalize within the scope of their training data. While they have been trained on vast amounts of images, text, tabular data, and other forms of directly Web-accessible data, they have not been exposed to the same degree to geospatial vector data (Mai *et al.* 2024a, 2024b), let alone multi-spectral remote sensing imagery (Cong *et al.* 2022; Fuller *et al.* 2023; Guo *et al.* 2024), time series data, or NetCDF-style array data more generally. Consequently, one would not expect (general-purpose) foundation models such as large language models (LLM) or text-to-image models to predict the next Atlantic Hurricane season. Second, many geospatial tasks are highly specific and require types of reasoning beyond the current abilities of state-of-the-art foundation models. To give a concrete example, current LLM and text-to-image models still struggle greatly with simple topological relations and reasoning (e.g., *Ji et al.* in this issue; *Cohn and Blackwell* 2024). Third, geography is inherently *local/regional* or contextual. It is shaped by the interplay of humans and the environment, as well as cultural, societal, and political factors that determine what is desirable or even *true*. Hence, models need to take space and time into account when determining a proper answer. To give a well-known example, answers about the varied and contested borders of the Kashmir region depend on who asks and when. Interestingly, this difficulty of representing and reasoning in the presence of spatial and temporal scopes extends beyond the models themselves but also affects their data backbones, e.g., geo-knowledge graphs (*Cai et al.* 2021, Zhang *et al.* 2025a in this issue). While *pluralistic* models are under development, current foundation models cannot handle different perspectives, contradicting data, nuanced cultural norms, and varied spatiotemporal scopes well (Janowicz, 2023; Sorensen *et al.* 2024).

Given these three motivating factors, **GeoFM can be defined as follows**: *Geo-foundational models are foundation models specifically trained on heterogeneous spatiotemporal data, capable of reliably performing advanced spatiotemporal reasoning, and designed to incorporate spatial, temporal, and other contextual factors into their output to support a wide range of (geo)spatial downstream tasks in geography and neighboring disciplines that benefit from a spatial or geographic perspective.*

2. From GeoAI to GeoFM

Over the past years, geographical artificial intelligence (GeoAI) has established itself as a rapidly growing subfield of spatial data science and geographic information science. Hence, before we explore the role geo-foundation models will play in the future and how they will advance GeoAI research, it is worth taking a step back to revisit the value proposition of GeoAI (Janowicz *et al.* 2020).

Simply put, GeoAI advances along two major dimensions: **(1)** it applies novel methods and technologies from the broader AI and machine learning community to geographic and geospatial research questions and **(2)** it feeds its own, novel theoretical and methodological contributions back to the broader AI community, e.g., by developing methods for spatial representation learning (Mai *et al.* 2022; Cepeda *et al.* 2023; Yu *et al.* 2024; Mai *et al.* 2024b; Chen *et al.* 2025c) applicable to a wide range of use cases in domains such as biodiversity studies (Cole *et al.* 2023; Mai *et al.* 2023b), health (Zhang *et al.* 2025b), transportation (Cai *et al.* 2020; Zhang *et al.* 2022; Rao *et al.* 2023), disaster mitigation (Sui *et al.* 2024; Li *et al.* 2025a; Chen *et al.* 2025b), sustainability (Wu *et al.* 2024), urban sensing (Huang *et al.* 2023), and so forth.

A common example of GeoAI research would be the work on detecting building footprints from remotely sensed imagery or predicting traffic flow. In the past, researchers would customize (neural) models for these singular tasks and train them on geography-specific datasets. Such workflows also made it relatively easy to clarify the research contributions of such work, as many of the developed models would be spatially explicit, i.e., they encode location explicitly instead of treating them as just yet another attribute. To give a simple example, location embeddings can be trained separately and concatenated with the embeddings representing learned building footprints, land classes, and so on (Mac Aodha *et al.* 2019; Yan *et al.* 2019; Mai *et al.* 2020).

With the advent of AlexNet (Krizhevsky *et al.* 2012) and ResNet (He *et al.* 2016) and related deep learning architectures, these GeoAI workflows have shifted rapidly. Instead of custom-tailored models trained *from scratch*, one would rather utilize a pre-trained model as a *feature extractor* by removing its output layer and integrating the remaining model into the task-specific neural architecture by feeding it into new layers. Such *transfer learning* dramatically speeds up training and improves model accuracy as the new model has already learned many rich and more abstract patterns, which only need to be adapted to the task at hand. Although modern foundation models were not yet on the horizon in the early 2010s, it was already clear that the era of custom, single-purpose models was slowly giving way to workflows developed around reuse and transferability. This shift raises a key question for GeoAI research: *how can we distinguish progress driven by GeoAI-specific innovation from improvements mostly gained through the application of transfer learning (and related methods) from general-purpose models?*

With the advent of foundation models in 2017,[2] this question is becoming more urgent than ever. Although not initially developed for this purpose, foundation models, such as LLM, are now widely used for a multitude of tasks and use cases within GeoAI research and applications. The successful combination of few-shot, prompt engineering, and transfer-learning methods on top of powerful general-purpose models raises the old question again: *is spatial really special?* Geo-foundation models may offer one potential answer, especially if they follow the second part of the GeoAI definition introduced above, namely, if they contribute back to the broader AI and machine learning community.

> *Just as LLMs encode the syntax, semantics, and pragmatics of human language, GeoFM could encode the language of space, i.e., the place-agnostic properties that define geography – spatial dependence and heterogeneity (Anselin,1988) and its related concepts such as scale, adjacency, spatial and temporal scopes, and so on.*

Of course, this is a challenging endeavor, for instance, because it raises serious concerns about the global representation of local geography (Liu *et al.* 2025a). Still, early models, including SatMAE (Cong *et al.* 2022), Prithvi (Jakubik *et al.* 2023), or AllenAI's Satlas[3] point towards a potential path.

3. Current models, benchmarks, and datasets

Here we provide a brief overview of existing GeoFM and GeoAI models as well as core datasets and benchmarks for evaluating these models.

3.1. Geo-Foundation models

Geo-foundation models are still a new and rapidly evolving research field. Based on the role FMs or GeoFMs play in each study, we can roughly classify the existing GeoFM-related research into the following categories: 1) adapting existing FMs on geospatial tasks via prompt engineering and task-specific fine-tuning; 2) developing advanced LLM agent frameworks for geospatial tasks; and 3) developing novel geo-foundation models via geo-aware model training and fine-tuning.

Most of the existing GeoFM-related research, including most papers in this issue (Gong *et al.* 2024; Hsu *et al.* 2024; Ji *et al.* 2025; Chen *et al.* 2025a; Zhang *et al.* 2025c), falls into the first category – FM adaptation and evaluation on geospatial tasks via prompt engineering and task-specific fine-tuning. As a critical component of GeoFM research, these studies can be seen as the first step to explore and investigate the advantages and disadvantages of existing FMs when applied to various geospatial tasks. So far, we have seen many successes across multiple tasks, including sustainability index prediction (Manvi *et al.* 2024), place name and location description recognition (Hu *et al.* 2023; Mai *et al.* 2024a), image geolocalization (Haas *et al.* 2024; Zhou *et al.* 2024b), vessel trajectory prediction (*Chen et al.*; in this issue), map reading and question answering (*Zhang et al.*; in this issue), geometry-based spatial reasoning (*Ji et al.*; in this issue), building function identification (*Gong et al.*; in this issue), remote sensing image object detection and instance segmentation (Osco *et al.* 2023; Zhang *et al.* 2024a) and (*Hsu et al.*; in this issue), and so on.

Despite the above successes, many studies acknowledge the limitations of current FMs in handling a range of geospatial tasks—particularly those involving novel data modalities that are not yet supported, such as geospatial vector data and network data (Manvi *et al.* 2024; Mai *et al.* 2024a; Ji *et al.* 2025). One solution to bypass these limitations is to utilize LLMs as agents that can synthesize geospatial processing workflows using existing geospatial toolsets, which corresponds to the second category of

GeoFM research. Until now, many agenetic frameworks have been developed to leverage multiple general-purpose FMs or GeoFMs for various types of geospatial tasks, including various remote sensing tasks (e.g., RSAgent (Xu *et al.* 2024)), spatial-reasoning question answering (e.g., Spatial-RAG (Yu *et al.* 2025)), spatial analysis (e.g., Autonomous GIS (Li and Ning, 2023), GeoGPT (Zhang *et al.* 2024e)), historical map understanding (e.g., PEACE (Huang *et al.* 2025)), map symbol editing (e.g., MapGPT (Zhang *et al.* 2024c)), map style transfer (e.g., CartoAgent (Wang *et al.* 2025; in this issue)), spatial optimization problem (e.g., RegionDefiner (Feng and Cao, 2024; in this issue)), spatial cognition and routing (e.g., Hybrid Mind (*Yang et al.*; in this issue)), geospatial image superresolution (e.g., 4KAgent (Zuo *et al.* 2025)), among others.

Other than these two research directions, another promising direction is to develop new task-agnostic geo-foundation models (GeoFMs) by using existing or novel pretraining objectives. Since most GeoFMs are developed based on existing neural architectures such as language models, vision transformers, segmentation decoders, etc., we further classify the current GeoFMs in four categories based on the data modalities they support and their application scenarios: geospatial language foundation models, geospatial vision foundation models, geospatial graph foundation models, and geospatial multimodal foundation models.

Geospatial language foundation models are developed by fine-tuning general-purpose LLMs on geo-referenced text corpora to support various purely language tasks such as geographic entity recognition, spatial relation extraction, geographic question answering, etc. Examples include K2 (Deng *et al.* 2023) and BB-GeoGPT (Zhang *et al.* 2024d).

Compared with geospatial language foundation models, there are relatively more papers on geospatial vision foundation models (vision GeoFMs), especially for remote sensing foundation models such as SatMAE (Cong *et al.* 2022), SatMAE++ (Noman *et al.* 2024), S2MAE (Li *et al.* 2024c), SpectralGPT (Hong *et al.* 2024), CROMA (Fuller *et al.* 2023), SkySense (Guo *et al.* 2024), Prithvi (Jakubik *et al.* 2023), among others. Other than these vision GeoFMs that are pretrained based on the masked autoencoder (MOE) objective, there are other vision generative GeoFMs which are pretrained on diffusion-based objectives such as DiffusionSat (Khanna *et al.* 2024) and CRS-Diff (Tang *et al.* 2024). Note that although many of these vision GeoFMs claim to be able to handle multi-modal data (e.g., SpectralGPT, CROMA, and SkySense), they usually refer to the ability of handling various types of remote sensing (RS) images, including optical multispectral RS imagery, optical hyperspectral RS imagery, Synthetic Aperture Radar (SAR), thermal infrared (TIR) imagery, LiDAR (Light Detection and Ranging) imagery, panchromatic imagery, etc, which can be considered as subcategories of vision data. Thus, we still classify them as vision GeoFMs instead of multimodal GeoFMs.

Geospatial graph foundation models are among the new types of GeoFMs recently developed to handle geospatial relational data such as large-scale geographic knowledge graphs (Zhu, 2024), spatial-social networks, and so forth. One promising example is Google's Population Dynamic Foundation Model (PDFM)(Agarwal *et al.* 2024), which is a large-scale graph neural network-based GeoFM that is able to handle places (e.g., zipcodes and counties), different place characteristics (e.g., Google search trend, business data, weather and climate data), and their geospatial relations (e.g., topological relations). PDFM can be used on multiple geospatial tasks that can not be done by other GeoFMs, such as population health outcome forecasting, vector-based superresolution and imputation, etc. Another example in this vein is Garner (Zhou *et al.* 2024a), a road network representation learning model based on a multi-view graph learning process. They used street view images to enrich the information of road networks, and the learned road representations are proven to be useful in tasks such as road function classification and average speed estimation.

Multimodal GeoFMs can handle multiple geospatial data modalities such as geotagged text, geospatial vector data, geospatial imagery, geospatial network data, etc. The most popular types are vision-language GeoFMs. Examples include EarthGPT (Zhang *et al.* 2024b), RemoteCLIP (Liu *et al.* 2024), GeoChat (Kuckreja *et al.* 2024), SkyEyeGPT (Zhan *et al.* 2025), GRAFT (Mall *et al.* 2024), MapReader (*Zhang et al.*; in this issue), etc. Other than vision-language GeoFMs, we also see other types of multimodal GeoFMs. For instance, to jointly consider both geospatial imagery and their location metadata, we have multiple vision-location GeoFMs, including CSP (Mai *et al.* 2023a), SatCLIP (Klemmer *et al.* 2023), GeoCLIP (Cepeda *et al.* 2023), RANGE (Dhakal *et al.* 2025), and GAIR (Liu *et al.* 2025b). In addition, Balsebre *et al.* (2024) used OpenStreetMap to pre-train an early version of CityFM which learns representations for multiple types of geographic objects that can be applied to different analyses, such as average speed estimation on road segments, functions of individual buildings, and population density in different regions.

3.2. Benchmarks and datasets

Multiple benchmarks have been developed to evaluate these GeoFMs. These benchmarks, most of which are designed for remote sensing vision and vision-language foundation models, include GEO-Bench (Lacoste *et al.* 2023), PANGAEA (Marsocci *et al.* 2024), and VRSBench (Li *et al.* 2024b). In addition to these remote sensing-centric benchmarks, TorchSpatial (Wu *et al.* 2024) has been designed as a benchmark for vision-location GeoFMs and geographic bias quantification. MapEval (Dihan *et al.* 2024), MapQA (Chang *et al.* 2022), and POI-QA (Han *et al.* 2025) are map and POI question answering datasets designed for vision-language GeoFMs. Yang *et al.* (2025) have developed a benchmark to evaluate the spatial cognition abilities of LLMs. Finally, GeoGrid-Bench (Jiang *et al.* 2025) is a multimodal grid-based geospatial benchmark designed for climate vision-language FMs.

4. Key research challenges in geo-foundation models

Despite this sizeable early literature on GeoFM, many challenges still remain to be solved. This section outlines several key challenges. However, it is worth noting that this list is not meant to be complete or even representative; instead, we encourage the community to keep discussing along these and other directions.

4.1. Forms of geo-foundation models

GeoFMs are emerging in two main forms: generative models like LLMs and representation models that produce embeddings for downstream tasks. On the utilization side, the prevailing trend favors generative LLMs, largely due to their intuitive natural language interface and versatility in answering almost any question that can be phrased as natural language. In contrast, representation models provide superior performance on quantitative prediction tasks, such as assessing population health risks with PDFM, which can be challenging for LLMs (Mai *et al.* 2024a). In a way, the design is task-dependent, e.g., generative solutions seem to be more plausible if we pursue an agent knowledgeable about geospatial literature, while representation models could be more suitable if we pursue better numerical outcomes for e.g., predictions of population dynamics. In this regard, questions remain: should we pursue one of these paths for different types of tasks, or should they be combined to gain the advantages of both?

From another angle, we also observe that there are three major ways of realizing GeoFM or using generalist FM to tackle geospatial tasks. First, a large portion of research explored effective means of prompting general-purpose FMs to perform geospatial tasks, e.g., Huang *et al*. (2024); Hu *et al*. (2023), (Wang *et al.* 2025; in this issue, *Ji et al.*; in this issue, and *Chen et al.*; in this issue). In such processes, the injection of geospatial domain knowledge is usually a key to make general-purpose FMs spatially aware. Second, a growing number of studies finetuned general-purpose FMs as GeoFMs, e.g., Manvi *et al*. (2024), (*Zhang et al.*; in this issue). Third, a few studies also developed GeoFM from scratch, mainly in remote sensing, e.g., Cong *et al*. (2022); Hong *et al*. (2024), with large-scale geospatial datasets. For now, it is unclear whether one of the paths is preferred to approach the vision of generally capable GeoFM so that the research community could consolidate our efforts, or if this is task-dependent, and, hence, varying paths should be taken for different types of tasks.

4.2. Truly multimodal GeoFM

It is widely acknowledged that multimodal data fusion is crucial for developing GeoFMs. This is due to the complementary nature of geospatial data, e.g., remote sensing captures physical surface properties, while points of interest reflect socioeconomic functions afforded by urban spaces. However, building effective multimodal GeoFMs requires aligning data not only semantically (e.g., matching vector shapes of rivers with their image representations) but also through spatiotemporal relationships. This involves reasoning about proximity, containment, connectivity, and temporal sequences across diverse modalities. Designing architectures that can jointly process such heterogeneous data, scale to large datasets, and accomplish effective cross-modality alignment remains a major open challenge. This is especially true for mode changes between model inputs and model outputs, which is very common in GIS data processing and analysis. It is worth noting that we are interested in models that are truly able to perform these operations (e.g., from rasters to vectors) latently, not simply by scripting.

4.3. The human dimension of GeoFM

A further critical challenge lies in reconciling the (seemingly) objective, quantitative nature of GeoFM with the subjective and complex human experience of geographic and urban spaces. Although FM show promise for analyzing physical and (collective) socioeconomic systems, the human dimension is not often included in pre-training. A fundamental question is whether those subjective and complex human experiences should become part of GeoFM. To this end, it is increasingly recognized that considering the human dimension is pivotal to developing data-driven solutions for tackling geospatial and urban challenges that, ultimately, are for and about humans (Janowicz, 2023; Liu *et al.* 2023; Birkin *et al.* 2025; Yue *et al.* 2025). The representation and reasoning of *social space*, which encompasses how different people experience their environment, form community bonds, and develop a unique attachment to a location, therefore becomes an unresolved hurdle for GeoFMs. We often portray current times as data-rich and us as drowning in data, but this is a very biased view. While there is plenty of (near-real-time) data about some geographic areas and data layers, we are starving for data from less represented areas or attributes that are less often recorded. As will be discussed below, this raises concerns about GeoFM misrepresenting geography, be it by introducing bias or by learning representations that do not align with those of groups or societies. Finally, so far we have argued about the role of human conceptualizations and experiences from a *static* perspective; however, representations of geographic space shift across space, time, and culture (Shi *et al.* 2025), and we need GeoFM to account for these changes.

4.4. Incorporating spatial priors

To enhance spatial awareness and better adapt to the nature of geospatial data, it is widely recognized that spatial priors should ideally be incorporated into the pre-training of GeoFM (Mai *et al.* 2024a). Among different types of spatial priors, the modeling of spatial proximity is perhaps the most common approach. This can be implemented in various ways, such as using location encoding (Klemmer *et al.* 2023; Mai *et al.* 2023a) or graph learning processes that smooth over neighborhoods (Huang *et al.* 2023; Agarwal *et al.* 2024). The similarity of geographic environments (configuration) has also been used to capture higher-order semantic similarity beyond spatial proximity (Zhou *et al.* 2024a). However, it remains unclear whether such spatial priors are adequate and robust across diverse geographic contexts and tasks – it is even possible that the modeling of spatial priors could compromise effectiveness, e.g., through over-smoothing. It is also uncertain whether other useful forms of spatial prior are underexplored. Moreover, those priors change across scale, resolution, modality, and so forth, and it is presently not clear how to best handle those. For instance, should they be explicitly engineered or implicitly learned? From a technical viewpoint, each type of spatial prior can be incorporated into pre-training in multiple ways, and the development of robust and effective technical solutions is actively being pursued. Finally, going one step further, we expect to see growing interest in the integration of spatiotemporal priors. Similar arguments can be made about evaluation and loss functions (Wiedemann *et al.* 2025).

4.5. Future datasets, benchmarks, and scaling laws

Another continual quest for advancing GeoFMs is to fill the gap in pre-training datasets and evaluation benchmarks. Although dedicated pre-training datasets are emerging, such as the language-focused BB-GeoSFT (Zhang *et al.* 2024d) and the multimodal ChatEarthNet (Yuan *et al.* 2025), they remain modest in scale compared to general-purpose corpora like LAION-5B (Schuhmann *et al.* 2022). This scale difference may be acceptable for models developed by fine-tuning generalist FMs, but it presents a significant bottleneck for training powerful, spatially-native (open) models from scratch.

A critical frontier is the creation of large-scale, multimodal geospatial datasets that fuse diverse sources like remote sensing imagery, street view imagery, points of interest, and human trajectories. The primary challenge extends beyond sheer scale – it lies in ensuring robust semantic, spatial, and temporal (Zhao *et al.* 2025) alignment across these heterogeneous modalities. The development of such large-scale and multimodal datasets is intrinsically linked to the quest for more comprehensive benchmarks. We need evaluation frameworks that move beyond simple GeoQA or classification to rigorously assess a model's capacity for complex, real-world tasks, such as dynamic urban analysis, cross-modal geographic retrieval, and human-environment interaction modeling. Without co-evolving our data and benchmarks, the true potential of GeoFMs will remain constrained.

Unfortunately, however, recent findings in AI scaling laws suggest a potential bottleneck for future GeoFM. Hoffmann *et al*. (2022) argue that many present (general-purpose) foundation models have scaled in parameter size without also scaling proportionally in training token size. This imbalance may be even more troublesome for GeoFM. Geographic data is not independent and identically distributed; instead, it may exhibit strong spatial autocorrelation. As a result, more data, e.g., denser samples, do not translate to more information (content). While empirical *GeoAI scaling laws* have yet to be established, the proper ratio between model size and required geodata may be even more challenging compared to general-purpose FM. Clearly, a lot of work remains to be done.

4.6. Geography according to foundation models

We have already hinted at several challenges at the intersection of technology and society, and more concretely at issues of representation. While we see great potential for GeoFMs, these models also introduce significant ethical concerns relating to bias, fairness, trust, and so on (Janowicz, 2023; Li *et al.* 2024a). One very interesting and underexplored aspect revolves around the question of how (geo)-foundation models represent geographic space, to what degree these representations differ from human cognition, and whether this may cause misalignment (Russell, 2019) in AI systems. Hence, for the years ahead, it is crucial to ask what *Geography looks like according to ChatGPT*.

As GeoAI systems become more autonomous, these *agentic AI* will act on or inform decisions about the physical world. Hence, it is essential to align their actions with societal goals, values, and norms. For many reasons, such goals are often not reflected in training data. Simply put, data is, by definition, from the past. For instance, while it is (still) true that a majority of industry leaders are male, we do not want GenAI models to exclusively or disproportionately represent male leadership when asked to depict an office scene or give career advice. In a geographic context, similar issues arise: just because a region has historically been underrepresented or economically disadvantaged, should not marginalize it in future AI output. This is also true for direct (text-to-image) depiction of the space around us. Prior work has shown that current foundation models may form strong geographic *defaults* (Liu *et al.* 2025a), which may cement certain geographic perspectives at the cost of others. Consequently, without understanding what future GeoFM models will know, using them for decision support or data-driven policy may lead to unexpected or unintended outcomes. Finally, most present work on AI alignment does not account for regional, e.g., cultural, differences. However, as geographers, we know that the aforementioned societal goals, values, and norms vary greatly across geographic space and time – without any being inherently superior to others. This calls for pluralistic alignment (Sorensen *et al.* 2024) approaches and, more specifically, for novel *geo-alignment* research.

5. The road ahead

Now that we have outlined key research challenges and their potential ethical implications, it is worth closing with a broader look ahead.

5.1. Competition, collaboration, convergence

Many noteworthy contributions to the recent GeoAI and Spatial Data Science literature have already been authored by international and interdisciplinary teams. Given the rapid increases in model size, required training data size, necessary compute, and potential for harm, the next major breakthroughs may be too big for single teams or even universities alone. In other domains, e.g., astrophysics, it is common to (try to) reach a community-wide consensus and author joint decadal surveys (National Academies of Sciences, Engineering, and Medicine *et al.* 2021) to establish overarching objectives, pathways, and roadmaps to showcase strategic unity to decision makers and funding agencies. In our field, we do not yet have such a tradition despite some noteworthy counterexamples with long-lasting impact, such as the National Center for Geographic Information and Analysis (NCGIA), established in 1988, or some of the *specialist meetings* organized over the past decade.[4] As researchers who are often involved in stiff competition to get our results out months (or even just weeks) before others do, we will have to learn to better blend competition and collaboration. In the near future – e.g., in the area of autonomous, agentic GeoMachina systems (Janowicz *et al.* 2020; Li *et al.* 2025b) – we *will* be forced to decide whether the potential cost of progress along some dimensions is worth the risk (Jonas, 1984). Going one step further, we may require *convergence* approaches similar to those promoted by the National Science Foundation's (NSF) Convergence Accelerator to rally government agencies, industry, non-governmental organizations (NGOs), and universities behind common goals and rapidly prototype potential solutions.

5.2. Education and training

With (geo-)foundation models advancing rapidly and agentic GIS agents already on the horizon, we will have to rethink how we train future talent and which skills will matter. In 2019, we argued that such agentic GeoMachina systems will be able to replace junior GIS analysts by 2030 (Janowicz *et al.* 2020) – it seems we may get there before. Today, LLM-based chatbots and early autonomous GIS systems (Li *et al.* 2025b) can retrieve data, suggest the correct data analysis steps, and geovisualize the results for many basic GIS needs. Many tasks that have taken considerable human expertise and work will become fully automated within the next 2-5 years. This, of course, has advantages such as opening up GIS analysis to the masses, revolutionizing education, and cost savings, but also many disadvantages that are not yet well understood. From politicians and futurologists to core AI researchers, it is fair to say that we all misjudged the effects of AI on the workforce.

Until very recently, society bet that AI (often confused with robotics) would replace jobs such as truck drivers, warehouse workers, cashiers, and even (super) models long before it would replace creative jobs (Frey and Osborne, 2017). Occupations such as photographers, computer programmers, database administrators, film and video editors, graphic designers, and scientists would be safe for the decades ahead. In fact, (Frey and Osborne, 2017, p. 48) close by arguing that *'[…] computerisation [will be] principally confined to low-skill and low-wage occupations. Our findings thus imply that as technology races ahead, low-skill workers will reallocate to tasks that are non-susceptible to computerisation – i.e., tasks requiring creative and social intelligence.'* For years, this was also a very convenient stance from a marketing perspective: AI will free us from monotonic work so we can focus on highly creative and intellectually demanding work. It seems highly unlikely somebody would make such claims in 2025. In fact, it's those (product) photographers, graphic designers, programmers, writers, and copy editors who are among the most affected by the rise of generative AI. To what extent current AI systems are truly *creative* is up for debate,[5] but their effect on creative work is already widely felt. Hence, as a community, we need to understand how the GIS and spatial data science job market will change and what skills our students will require in the future. For example, there are many social intelligence skills involved in spatial data science, but we typically do not highlight them in our classes or textbooks. More generally, AI's long-term impact on education and training is still underexplored (Kasneci *et al.* 2023; Latif *et al.* 2023).

Intuitively, one could expect a decline in skills among (GIScience) students as well as a drop in confidence. Just as the widespread use of digital navigation systems has impacted our spatial memory (Dahmani and Bohbot, 2020), habitual use of GenAI/FM may reduce spatial thinking more broadly. While most of us spend less than 30 minutes handling navigation systems per day, we may interact more frequently with various AI analysts and tools in the near future. Now, giving up some of our skills is not a first; however, in this case, it is happening at a speed not seen before and by systems not truly designed for autonomous tasks. Thus, skills that help us better interact with such agents, critically think about their outputs, align AI with societal goals, and so on, will increase in importance.

6. Summary and conclusions

In this work, we presented perspectives on the emerging study of geo-foundation models. We motivated and defined the term, put it into its historical context within

GeoAI and spatial data science more broadly, and reviewed core datasets, models, and benchmarks. Next, we proposed key research challenges for future GeoFM research, such as GeoAI scaling laws, geo-alignment of AI, truly multimodal GeoFM, and so on, and motivated them based on current research trends within the GeoAI community. We identified potential risks and outlined the road ahead by discussing potential impacts on education and the increasing need for broad international collaboration. We believe that current and next-generation geo-foundation models will be defining steps for the future of GeoAI and spatial data science, and hope that our community's contributions will fuel progress beyond our own research field, all while balancing innovation and responsibility.

Notes

1. The exact count depends on the model architecture and does not matter here. *Emergence* is defined from the moment *chain of thought* begins to consistently outperform standard prompting.
2. If we take the seminal paper by Vaswani et al. (2017) as the (somewhat arbitrary) earliest starting point.
3. https://satlas.allen.ai/.
4. See Goodchild et al. (2025) for a recent example describing the outcomes of such a meeting.
5. The same argument can be made about *intelligence* in general and Searle's Chinese Room counterargument.

Acknowledgments

We thank Prof. Yang Yue at The Hong Kong University of Science and Technology (Guangzhou) for in-depth discussions on the human dimension of GeoFM and the members of the STKO lab for their feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Krzysztof Janowicz is a professor of Geoinformatics and chair of the Department of Geography and Regional Research at the University of Vienna, Austria. His research interests include cross-domain knowledge representation, geo-semantics, geospatial artificial intelligence, and AI ethics.

Gengchen Mai is an Assistant Professor at the Department of Geography and the Environment, the University of Texas at Austin, USA. He is also a visiting faculty researcher at Google Research. His research interests include geo-foundation models, geographic question answering, and spatially explicit artificial intelligence models.

Weiming Huang is a Lecturer at the School of Geography, University of Leeds, UK. He obtained his Ph.D. in Geographical Information Science from Lund University, Sweden, and was a Wallenberg Postdoctoral Fellow at Nanyang Technological University and Lund University. His research interests include spatial data mining and geospatial foundation models.

Rui Zhu is a Senior Lecturer in the School of Geographical Sciences at the University of Bristol, UK. His research focuses on geospatial semantics, geospatial artificial intelligence (GeoAI), and spatially explicit machine learning.

Ni Lao is a Research Scientist at Google, USA. He is an expert in machine learning, knowledge graphs, and natural language understanding.

Ling Cai is a Senior Machine Learning Engineer at Bytedance. Her interest lies in Generative AI for E-Commerce and Symbolic Reasoning.

ORCID

Krzysztof Janowicz http://orcid.org/0009-0003-1968-887X
Gengchen Mai http://orcid.org/0000-0002-7818-7309
Weiming Huang http://orcid.org/0000-0002-3208-4208

Data availability statement

No data or codes were generated or used in this review paper.

References

Agarwal, M., *et al.*, 2024. General geospatial inference with a population dynamics foundation model. *arXiv preprint arXiv:2411.07207*.
Anselin, L., 1988. *Spatial econometrics: methods and models*, volume 4. Dordrecht: Springer Science & Business Media.
Balsebre, P., *et al.*, 2024. City foundation models for learning general purpose representations from openstreetmap. *Proceedings of the 33rd ACM international conference on information and knowledge management*, 87–97.
Birkin, M., *et al.*, 2025. Digital twins and ai for healthy and sustainable cities. *Computers, Environment and Urban Systems*, 120, 102305.
Bommasani, R., *et al.*, 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
Brown, T., *et al.*, 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
Cai, L., *et al.*, 2020. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Transactions in GIS*, 24 (3), 736–755.
Cai, L., *et al.*, 2021. Time in a box: advancing knowledge graph completion with temporal scopes. *Proceedings of the 11th knowledge capture conference*, 121–128.
Cepeda, V.V., Nayak, G.K., and Shah, M., 2023. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Thirty-seventh conference on neural information processing systems*.
Chang, S., *et al.*, 2022. MapQA: A dataset for question answering on choropleth maps. *arXiv preprint arXiv:2211.08545*.
Chen, N., *et al.*, 2025a. SEMINT: an LLM-empowered long-term vessel trajectory prediction framework. *International Journal of Geographical Information Science*, 1–35. https://doi.org/10.1080/13658816.2025.2487990
Chen, X., *et al.*, 2025b. Harnessing geospatial artificial intelligence and deep learning for landslide inventory mapping: Advances, challenges, and emerging directions. *Remote Sensing*, 17 (11), 1856.
Chen, Y., *et al.*, 2025c. Self-supervised representation learning for geospatial objects: A survey. *Information Fusion*, 123, 103265.

Cohn, A.G., and Blackwell, R.E., 2024. Evaluating the ability of large language models to reason about cardinal directions–dataset.

Cole, E., *et al.*, 2023. Spatial implicit neural representations for global-scale species mapping. In *International conference on machine learning*. PMLR.

Cong, Y., *et al.*, 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35, 197–211.

Dahmani, L., and Bohbot, V.D., 2020. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10 (1), 6310.

Deng, C., *et al.*, 2023. Learning a foundation language model for geoscience knowledge understanding and utilization. *arXiv preprint arXiv:2306.05064*.

Dhakal, A., *et al.*, 2025. Range: Retrieval augmented neural fields for multi-resolution geo-embeddings. In *Proceedings of the computer vision and pattern recognition conference*, 24680–24689.

Dihan, M.L., *et al.*, 2024. MapEval: A map-based evaluation of geo-spatial reasoning in foundation models. *arXiv preprint arXiv:2501.00316*.

Feng, X., and Cao, Y., 2024. Enhancing the accessibility of regionalization techniques through large language models: a case study in conversational agent guidance. *International Journal of Geographical Information Science*, 1–24. In this issue.

Frey, C.B., and Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.

Fuller, A., Millard, K., and Green, J.R., 2023. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. In *Thirty-seventh conference on neural information processing systems*.

Gong, Z., *et al.*, 2024. BF-SAM: enhancing SAM through multi-modal fusion for fine-grained building function identification. *International Journal of Geographical Information Science*, 1–27. https://doi.org/10.1080/13658816.2024.2399142

Goodchild, M., *et al.*, 2025. Locational data and the public interest. *IEEE Transactions on Technology and Society*, 6 (2), 129–154.

Guo, X., *et al.*, 2024. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27672–27683.

Haas, L., *et al.*, 2024. Pigeon: Predicting image geolocations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12893–12902.

Han, X., *et al.*, 2025. A dataset for spatiotemporal-sensitive POI question answering. *arXiv preprint arXiv:2505.10928*.

He, K., *et al.*, 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hoffmann, J., *et al.*, 2022. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.

Hong, D., *et al.*, 2024. Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46 (8), 5227–5244.

Hsu, C.-Y., Li, W., and Wang, S., 2024. Geospatial foundation models for image analysis: evaluating and enhancing NASA-IBM Prithvi's domain adaptability. *International Journal of Geographical Information Science*, 1–30. https://doi.org/10.1080/13658816.2024.2397441

Hu, Y., *et al.*, 2023. Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*, 37 (11), 2289–2318.

Huang, W., Wang, J., and Cong, G., 2024. Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. *International Journal of Geographical Information Science*, 38 (7), 1414–1442.

Huang, W., *et al.*, 2023. Learning urban region representations with pois and hierarchical graph infomax. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 134–145.

Huang, Y., *et al.*, 2025. PEACE: Empowering geologic map holistic understanding with MLLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3899–3908.

Jakubik, J., *et al.*, 2023. Foundation models for generalist geospatial artificial intelligence. *arXiv preprint arXiv:2310.18660*.

Janowicz, K., 2023. Philosophical foundations of geoai: Exploring sustainability, diversity, and bias in geoai and spatial data science. *In*: S. Gao, Y. Hu, and W. Li, eds. *Handbook of geospatial artificial intelligence*, 26–42. Boca Raton: CRC Press.

Janowicz, K., *et al.*, 2020. Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond.

Ji, Y., *et al.*, 2025. Foundation models for geospatial reasoning: assessing the capabilities of large language models in understanding geometries and topological spatial relations. *International Journal of Geographical Information Science*, 1–38. In this issue.

Jiang, B., *et al.*, 2025. GeoGrid-Bench: Can foundation models understand multimodal gridded geo-spatial data? *arXiv preprint arXiv:2505.10714*.

Jonas, H., 1984. *The imperative of responsibility: in search of an ethics for the technological age*. Chicago: The University of Chicago Press.

Kaplan, J., *et al.*, 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Kasneci, E., *et al.*, 2023. Chatgpt for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Khanna, S., *et al.*, 2024. Diffusionsat: A generative foundation model for satellite imagery. In *The Twelfth international conference on learning representations*.

Klemmer, K., *et al.*, 2023. Towards global, general-purpose pretrained geographic location encoders. In *NeurIPS 2023 workshop on distribution shifts: new frontiers with foundation models*.

Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

Kuckreja, K., *et al.*, 2024. Geochat: grounded large vision-language model for remote sensing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 27831–27840.

Lacoste, A., *et al.*, 2023. Geo-bench: toward foundation models for earth monitoring. In *Advances in neural information processing systems*, 36, 51080–51093.

Latif, E., *et al.*, 2023. Artificial general intelligence (agi) for education. *arXiv preprint arXiv: 2304.12479*.

Li, H., *et al.*, 2025a. Cross-view geolocalization and disaster mapping with street-view and VHR satellite imagery: a case study of hurricane IAN. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 841–854.

Li, W., *et al.*, 2024a. GeoAI for science and the science of GeoAI. *Journal of Spatial Information Science*, (29), 1–17. https://doi.org/10.5311/JOSIS.2024.29.349

Li, X., Ding, J., and Elhoseiny, M., 2024b. VRSBench: a versatile vision-language benchmark dataset for remote sensing image understanding. *In Advances in Neural Information Processing Systems*, 37, 3229–3242.

Li, X., Hong, D., and Chanussot, J., 2024c. S2mae: a spatial-spectral pretraining foundation model for spectral remote sensing data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 24088–24097.

Li, Z., and Ning, H., 2023. Autonomous Gis: the next-generation AI-powered GIS. *International Journal of Digital Earth*, 16 (2), 4668–4686.

Li, Z., *et al.*, 2025b. Giscience in the era of artificial intelligence: A research agenda towards autonomous gis. *arXiv preprint arXiv:2503.23633*.

Liu, F., *et al.*, 2024. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–16.

Liu, P., *et al.*, 2023. Towards human-centric digital twins: leveraging computer vision and graph models to predict outdoor comfort. *Sustainable Cities and Society*, 93, 104480.

Liu, Z., *et al.*, 2025a. Operationalizing geographic diversity for the evaluation of AI-generated content. *Transactions in GIS*, 29 (3), e70057.

Liu, Z., *et al.*, 2025b. Gair: Improving multimodal geo-foundation model with geo-aligned implicit representations. *arXiv preprint arXiv:2503.16683*.

Mac Aodha, O., Cole, E., and Perona, P., 2019. Presence-only geographical priors for fine-grained image classification. In *Proceedings of the IEEE international conference on computer vision*, 9596–9606.

Mai, G., *et al.*, 2024a. On the opportunities and challenges of foundation models for geoai (vision paper). *ACM Transactions on Spatial Algorithms and Systems*, 10 (2), 1–46.

Mai, G., *et al.*, 2020. Se-kge: a location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting. *Transactions in GIS*, 24 (3), 623–655.

Mai, G., *et al.*, 2022. A review of location encoding for geoai: methods and applications. *International Journal of Geographical Information Science*, 36 (4), 639–673.

Mai, G., *et al.*, 2023a. Csp: self-supervised contrastive spatial pre-training for geospatial-visual representations. In *the Fortieth international conference on machine learning (ICML 2023)*.

Mai, G., *et al.*, 2023b. Sphere2vec: a general-purpose location representation learning over a spherical surface for large-scale geospatial predictions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202, 439–462.

Mai, G., *et al.*, 2024b. Srl: Towards a general-purpose framework for spatial representation learning. In *Proceedings of the 32nd ACM international conference on advances in geographic information systems*, 465–468.

Mall, U., *et al.*, 2024. Remote sensing vision-language foundation models without annotations via ground remote alignment. In *The Twelfth international conference on learning representations*.

Manvi, R., *et al.*, 2024. Geollm: Extracting geospatial knowledge from large language models. In *Proceedings of the Twelfth international conference on learning representations (ICLR 2024)*.

Marsocci, V., *et al.*, 2024. Pangaea: A global and inclusive benchmark for geospatial foundation models. *arXiv preprint arXiv:2412.04204*.

National Academies of Sciences, Engineering, and Medicine, *et al.*, 2021. *Pathways to discovery in astronomy and astrophysics for the 2020s*. Washington, DC: The National Academies Press.

Noman, M., *et al.*, 2024. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 27811–27819.

Osco, L.P., *et al.*, 2023. The segment anything model (sam) for remote sensing applications: from zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, 124, 103540.

Rao, J., Gao, S., and Zhu, S., 2023. Cats: Conditional adversarial trajectory synthesis for privacy-preserving trajectory data publication using deep learning approaches. *International Journal of Geographical Information Science*, 37 (12), 2538–2574.

Russell, S., 2019. *Human compatible: AI and the problem of control*. Penguin UK.

Schuhmann, C., *et al.*, 2022. Laion-5b: an open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35, 25278–25294.

Shi, M., *et al.*, 2025. Defining concept drift and its variants in research data management: A scientometric case study on geographic information science. *Transactions in GIS*, 29 (3), e70058.

Sorensen, T., *et al.*, 2024. A roadmap to pluralistic alignment. *arXiv preprint arXiv:2402.05070*.

Sui, T., *et al.*, 2024. Biau-net: wildfire burnt area mapping using bi-temporal sentinel-2 imagery and u-net with attention mechanism. *International Journal of Applied Earth Observation and Geoinformation*, 132, 104034.

Tang, D., *et al.*, 2024. Crs-diff: controllable generative remote sensing foundation model. *CoRR*.

Vaswani, A., *et al.*, 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, C., *et al.*, 2025. Cartoagent: a multimodal large language model-powered multi-agent cartographic framework for map style transfer and evaluation. *International Journal of Geographical Information Science*, 1–34. In this issue.

Wei, J., *et al.*, 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Wiedemann, N., Uscidda, T., and Raubal, M., 2025. GeOT: a spatially explicit framework for evaluating spatio-temporal predictions. *International Journal of Geographical Information Science*, 1–31. https://doi.org/10.1080/13658816.2025.2481101

Wu, N., *et al.*, 2024. Torchspatial: A location encoding framework and benchmark for spatial representation learning. *Advances in Neural Information Processing Systems*, 37, 81437–81460.
Xu, W., *et al.*, 2024. RS-Agent: automating remote sensing tasks through intelligent agent. *arXiv preprint arXiv:2406.07089*.
Yan, B., *et al.*, 2019. A spatially explicit reinforcement learning model for geographic knowledge graph summarization. *Transactions in GIS*, 23 (3), 620–640.
Yang, A., *et al.*, 2025. Evaluating and enhancing spatial cognition abilities of large language models. *International Journal of Geographical Information Science*, 1–36.
Yu, D., *et al.*, 2025. Spatial-rag: spatial retrieval augmented generation for real-world spatial reasoning questions. *arXiv preprint arXiv:2502.18470*.
Yu, D., *et al.*, 2024. Polygongnn: representation learning for polygonal geometries with heterogeneous visibility graph. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 4012–4022.
Yuan, Z., *et al.*, 2025. Chatearthnet: a global-scale image–text dataset empowering vision–language geo-foundation models. *Earth System Science Data*, 17 (3), 1245–1263.
Yue, Y., *et al.*, 2025. Shaping future sustainable cities with ai-powered urban informatics: Toward human-ai symbiosis. *Computational Urban Science*, 5 (1), 31.
Zhan, Y., Xiong, Z., and Yuan, Y., 2025. Skyeyegpt: unifying remote sensing vision-language tasks via instruction tuning with large language model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 221, 64–77.
Zhang, C., *et al.*, 2025a. Enhancing semantic accuracy in geographic knowledge graph embeddings through temporal encoding. *International Journal of Geographical Information Science* (In press).
Zhang, J., *et al.*, 2025b. Space: a spatial counterfactual explainable deep learning model for predicting out-of-hospital cardiac arrest survival outcome. *International Journal of Geographical Information Science*, 1–32.
Zhang, J., *et al.*, 2024a. Text2seg: zero-shot remote sensing image semantic segmentation via text-guided visual foundation models. In *Proceedings of the 7th ACM SIGSPATIAL International workshop on AI for geographic knowledge discovery*, 63–66.
Zhang, W., *et al.*, 2024b. 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.
Zhang, X., *et al.*, 2022. Forecasting the crowd: an effective and efficient neural network for citywide crowd information prediction at a fine spatio-temporal scale. *Transportation Research Part C: Emerging Technologies*, 143, 103854.
Zhang, Y., *et al.*, 2024c. MapGPT: an autonomous framework for mapping by integrating large language model and cartographic tools. *Cartography and Geographic Information Science*, 51 (6), 717–743.
Zhang, Y., *et al.*, 2024d. Bb-geogpt: a framework for learning a large language model for geographic information science. *Information Processing & Management*, 61 (5), 103808.
Zhang, Y., *et al.*, 2024e. Geogpt: an assistant for understanding and processing geospatial tasks. *International Journal of Applied Earth Observation and Geoinformation*, 131, 103976.
Zhang, Y., *et al.*, 2025c. Mapreader: a framework for learning a visual language model for map analysis. *International Journal of Geographical Information Science*, 1–36.
Zhao, T., *et al.*, 2025. Quantifying seasonal bias in street view imagery for urban form assessment: a global analysis of 40 cities. *Computers, Environment and Urban Systems*, 120, 102302.
Zhou, H., *et al.*, 2024a. Road network representation learning with the third law of geography. *Advances in Neural Information Processing Systems*, 37, 11789–11813.
Zhou, Z., *et al.*, 2024b. Img2loc: Revisiting image geolocalization using multi-modality foundation models and image-based retrieval-augmented generation. In *Proceedings of the 47th international acm sigir conference on research and development in information retrieval*, 2749–2754.
Zhu, R., 2024. *Geospatial knowledge graphs*. New Jersey, US: John Wiley & Sons, Ltd, 1–7.
Zuo, Y., *et al.*, 2025. 4KAgent: agentic any image to 4K super-resolution. *arXiv preprint arXiv: 2507.07105*.