

INVITED ARTICLE

# GeoAI: Where machine learning and big data converge in GIScience

Wenwen Li

School of Geographical Science and Urban Planning, Arizona State University, Tempe, AZ

*Received: March 2, 2020; accepted: April 16, 2020*

---

**Abstract:** In this paper GeoAI is introduced as an emergent spatial analytical framework for data-intensive GIScience. As the new fuel of geospatial research, GeoAI leverages recent breakthroughs in machine learning and advanced computing to achieve scalable processing and intelligent analysis of geospatial big data. The three-pillar view of GeoAI, its two methodological threads (data-driven and knowledge-driven), as well as their geospatial applications are highlighted. The paper concludes with discussion of remaining challenges and future research directions of GeoAI.

**Keywords:** GeoAI, data-driven, big data, CNN, knowledge graph, convergence research, deep learning

---

## 1 Introduction

In the 2020s, the world is experiencing the greatest challenges ever regarding the degradation of environmental quality, the more frequent occurrence of natural disasters, new and reemerging diseases, and the surge in social unrest, among others. These global issues are largely due to rapid population and economic growth, the excessive consumption of natural resources, and growing levels of social inequality. Global climate change, which yields increasingly extreme heat and droughts, resulted in intensified wildfires across the world. In 2019, the bushfire in Australia caused 33 deaths, more than 2,000 houses were damaged, nearly 15 million acres of land burned, and 480 million animals were killed [6]. In addition, the fire generated significant amounts of hazardous smoke, causing severe air pollution and negative health consequences.

The Arctic, one of the Earth's remaining frontiers, has also undergone dramatic changes due to rising global temperatures. The Arctic permafrost is at high risk of thawing, in which case an immense amount of carbon dioxide and methane will be released, which

will further exaggerate the greenhouse effect and global warming [15]. In January 2020, a novel coronavirus, COVID-19, emerged as a highly infectious disease that quickly spread across the world. As of this writing, in April 2020, it has sickened more than 2 million people and caused at least 130,000 deaths. Action needs to be taken to mitigate the negative effects of these devastating events to ensure sustainable development and the well-being of humankind [1].

It is clear that these global problems are geospatial in nature—they all occur at a certain location on or near the Earth’s surface and have distinct space-time patterns due to different geographical processes or interactive mechanisms between humans and nature. GIScience, which deals with the collection, storage, analysis, and visualization of geospatial data, will no doubt play a central role in tracking and mapping environmental and social phenomena at and across different scales, projecting how these phenomena evolve, crafting theories related to driving factors and processes, and setting policies to prevent and respond to ongoing and emerging issues.

Three recent technological advances have prepared GIScientists to better tackle these problems. First, the rapid development in Earth Observation (EO), wireless sensor networks, Internet and Communication Technology (ICT), and the prevalence of social media platforms have fostered the explosive growth of geospatial big data available at very fine spatial, temporal, and spectral resolutions [8]. These data allow us to study in near-real-time the Earth, its changing environment, and human behavior at an unprecedented scale and level of detail. Second, the recent breakthrough in machine learning, or more generally artificial intelligence (AI) and more specifically deep learning, enables a new research paradigm—data-driven science—relying on which, massive geospatial data that are difficult to handle using traditional spatial analysis methods, can now be analyzed, mined, and visualized. As a result, complex hidden patterns can be revealed, new questions can be asked and answered, and new knowledge can be discovered. Third, the dramatic increase in computational resources, such as Graphics Processing Units (GPUs) offers backbone support for the efficient training of machine learning models with big data. In addition, the availability of cloud computing platforms to the public allows individual researchers to build big data applications. All these revolutionary advances are shaping the future of geospatial research.

## 2 GeoAI: a new power-up for geospatial research

GeoAI, or geospatial artificial intelligence, sits at the junction of AI, geospatial big data, and high performance computing (HPC) to provide a promising solution technology for data- or compute-intensive geospatial problems. Figure 1 illustrates the conceptual, three-pillar view of GeoAI. As an interdisciplinary expansion of AI, the aim of GeoAI is for the machine to gain the intelligence to perform spatial reasoning and analysis like humans. GeoAI develops as AI evolves and it has two major method classes: knowledge-driven, known as the top-down approach, and data-driven, known as the bottom-up approach. Without a doubt, the data-driven approach, led by machine learning, has become the mainstream AI today because of its outstanding ability to learn to make predictions from massive amounts of data without the need to explicitly program the analytical rules. Deep learning, as a recent breakthrough in machine learning, has transformed data analytics paradigm in two ways.



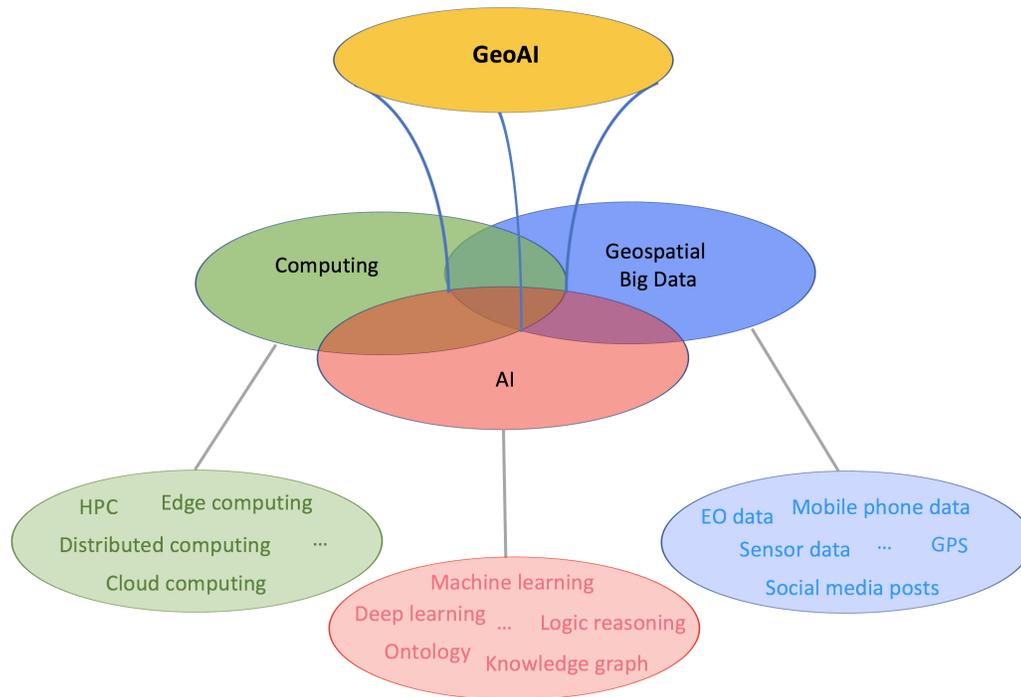


Figure 1: A conceptual, three-pillar view of GeoAI.

First, deep learning models, such as convolutional neural networks (CNNs), have the ability to automatically extract prominent features from the data to help differentiate object classes so that accurate predictions can be made. This is a great advantage over traditional spatial analytical approaches because deep learning allows for a more automatic and intelligent method of feature extraction, a strategy which is especially helpful in solving big data problems in which there is often limited prior knowledge about the underlying patterns and processes within the data. Second, deep learning models introduce a local operation—convolution—into the learning process, such that numerous interdependencies residing in the global computation of traditional neural network models are broken down. This type of model design makes it much easier to be parallelized and trained on a high-performance or distributed computing environment. Even when a model structure goes very deep with thousands to millions of parameters to learn, it is still very likely to converge with strong predictive power.

Machine learning has also powered up the more traditional, top-down, ontological-based GeoAI approaches. These approaches tackle spatial cognition problems, such as semantic similarity measures [10], by leveraging ontology and logical reasoning. Different from data-driven approaches, an ontological approach relies on a knowledge base to provide semantic definitions of real-world entities and their interrelationships in the format of <subject, predicate, object> triples. The knowledge discovery process follows pre-defined reasoning rules and constraints and uses deductive reasoning to ensure that

each newly derived fact can be formally validated with its reasoning path clearly traceable. Although highly interpretable, this approach suffers from two drawbacks: (1) ontology engineering—the process of constructing a knowledge base—heavily or even solely relies on expert knowledge and manual work. Although a very deep structure can be established to describe the complex relationships among entities, the human-centered approach can hardly scale to make the knowledge base comprehensive to ensure its performance; and (2) although ontology tries to capture the complexity in human logic, it needs to be implemented in a way that is machine understandable so some simplification and abstraction are inevitable. This adds another layer of performance challenge in making accurate predictions and decisions.

The recent notable progress of the knowledge graph and its combined use with machine learning has elevated the ontological approach back to the forefront of GeoAI research. Similar to an ontology, a knowledge graph is based on semantics and designed to infer new knowledge and derive new insights. But the two differ in that an ontology normally emphasizes depth, whereas a knowledge graph orients more with breadth. In this vein, an ontology can serve as the schema which defines the semantic structure of domain knowledge; a knowledge graph will follow this schema to “instantiate” the knowledge base with millions to even trillions of triples for scalable geospatial applications. To achieve a size like this, it is imperative to rely on machine learning for the automatic construction of a knowledge graph. Knowledge inference will also build on recent advances in machine learning, such as graph neural network and embedding techniques to achieve automation and intelligence.

Both methodological threads of GeoAI have widespread applications in geospatial domains. The remote sensing community has extensively used CNN for scene classification (natural and urban), change detection, and other image analysis tasks [9,18,19]. Deep learning has been leveraged to support cartographic tasks such as generalization, smart mapping, and map element inspection [16]. Machine learning has been increasingly used for the semantic and sentiment analysis of social media data and other natural language text documents [3]. In spatial information retrieval, the knowledge graph has become a key component and the backbone technology for intelligent question answering, hidden link prediction, and semantic search, among other things [11]. Multi-dimensional geospatial data, such as lidar and scientific data from numerical simulation models, can also benefit from the processing power such as 3D CNN for 3D object detection and event classification [13]. Time-series data, streamed from Internet of Things (IoT) sensors, can take advantage of Recurrent Neural Networks (RNNs) to achieve real-time predictions and analyses [14]. The diversity in geospatial data and the prevalence of location-based service make GIScience a natural home for these uses and the boom of AI.

### 3 Concluding remarks

Although exciting progress has been made, GeoAI remains in an early stage. Many technical challenges, such as the lack of good quality training data, uncertainty modeling, geographical transferability of a model result, and cross-scale and cross-resolution learning have yet to be addressed [5]. To further advance GeoAI and establish it as a cornerstone of GIScience, progress needs to be made in the following areas:

### 3.1 Integrated GeoAI

As discussed, both data-driven GeoAI and knowledge-driven GeoAI have their strengths and weaknesses. Data-driven models often tackle problems by following the “trial and error” strategy without guidance from theory or prior knowledge. They achieve high predictive power by building complex models while sacrificing model interpretability. Knowledge-driven approaches have strong expressive power but less satisfying predictive performance due to the many constraints applied on a model. An integrated GeoAI approach, which uses domain knowledge to guide the design of data-driven models, will simplify model design, reduce training time, increase model expressiveness, and augment decision power.

### 3.2 Fasten two-way knowledge transfer in GeoAI research

Although a large number of research projects have been conducted for applying AI for geospatial problems, the study of GeoAI should not merely involve a simple import of AI to geography. A two-way knowledge transfer from both “AI” to “Geo” and “Geo” to “AI” has to be enabled in order to establish GeoAI as a research field that creates impact within and beyond the geospatial domain. Some pioneer research along this line (i.e., spatially-explicit GeoAI models) has been proposed [7]. In future research, it is critically important to integrate spatial thinking and spatial principles into the GeoAI model development to make the model smarter at solving geospatial problems and developing artificial geospatial intelligence.

### 3.3 GeoAI-enabled convergent GIScience

The newer science landscape is leaning toward the conduct of convergence research which requires the seamless integration of theories, knowledge, tools, and expertise across traditional discipline boundaries for tackling complex problems in a collaborative manner. Space and time could serve as the nexus of the web of knowledge. For the challenging environmental and public health problems mentioned in the beginning of the paper, GeoAI will and should play a key role in fusing massive datasets, performing intelligent analysis, enabling interactive question answering for both researchers and the general public, as well as providing decision-support capabilities for planners and stakeholders. Hence, besides fundamental research, it is essential to deepen and broaden applications of GeoAI toward addressing big and bigger problems, for the benefit of society in realizing convergent GIScience.

Without a doubt, the advances in GeoAI will help address societal challenges such as disease outbreaks, natural disasters, and climate change. With situations related to COVID-19, for example, GeoAI is being applied to support the rapid collection of data from multiple sources regarding the global environment, social economy, transportation, human mobility, and confirmed cases. The advanced analytics and machine learning capabilities that GeoAI offers facilitate the timely identification of vulnerable populations in this pandemic [4]. It is also being leveraged to evaluate social distancing effects [17] and to forecast impacts on hospital resources [12]. Together, these will improve the understanding of disease transmission patterns and provide scientific support for governments to enact plans to protect its citizens and save lives.

This technology can also be used to detect patterns and discover explanatory spatial models from big data [2]. In this way, the mechanics of spatial processes, such as the spread of wildfire and the formation and development of tropical storms can be better understood. All these disastrous events, whether naturally occurring or caused by humans, are related to long-term changes throughout the world. GeoAI, through its combined use with big data and knowledge-driven approaches, will contribute to the discovery and validation of new theories and science, thereby enriching and helping sustain our global environment and society.

## Acknowledgments

This research was in part supported by National Science Foundation, grant number 1853864, 1455349, 1937908, and 1936677.

## References

- [1] ELLIOTT, J. *An introduction to sustainable development*. Routledge, 2012.
- [2] GAHEGAN, M. Fourth paradigm giscience? prospects for automated discovery and explanation from data. *International Journal of Geographical Information Science* 34, 1 (2020), 1–21.
- [3] GAO, H., AND LIU, H. Data analysis on location-based social networks. In *Mobile social networking*. Springer, 2014, pp. 165–194.
- [4] HEROSIAN, L. Clinical AI Leader Jvion Launches COVID Community Vulnerability Map. <https://jvion.com/news/press/clinical-ai-leader-jvion-launches-covid-community-vulnerability-map>, 2020. Last Accessed April 15, 2020.
- [5] HU, Y., GAO, S., LUNGA, D., LI, W., NEWSAM, S., AND BHADURI, B. Geoai at acm sigspatial: progress, challenges, and future directions. *SIGSPATIAL Special* 11, 2 (2019), 5–15.
- [6] HUTT, R. Australia bushfires: 5 things to know about the crisis. <https://www.weforum.org/agenda/2020/01/australia-bushfires-size-impact-wildlife-emissions/>, 2020. Last Accessed Feb 20, 2020.
- [7] JANOWICZ, K., GAO, S., MCKENZIE, G., HU, Y., AND BHADURI, B. Geoai: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond.
- [8] LI, W., BATTY, M., AND GOODCHILD, M. F. Real-time gis for smart cities. *International Journal of Geographic Information Science* 34, 2 (2020), 311–324.
- [9] LI, W., AND HSU, C.-Y. Automated terrain feature identification from remote sensing imagery: a deep learning approach. *International Journal of Geographical Information Science* 34, 4 (2020), 637–660.



- [10] LI, W., RASKIN, R., AND GOODCHILD, M. F. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *International Journal of Geographical Information Science* 26, 8 (2012), 1415–1435.
- [11] MAI, G., JANOWICZ, K., YAN, B., AND SCHEIDER, S. Deeply integrating linked data with geographic information systems. *Transactions in GIS* 23, 3 (2019), 579–600.
- [12] MURRAY, C. Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator-days and deaths by us state in the next 4 months. *medRxiv* (2020).
- [13] QI, C. R., SU, H., MO, K., AND GUIBAS, L. J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 652–660.
- [14] REICHSTEIN, M., CAMPS-VALLS, G., STEVENS, B., JUNG, M., DENZLER, J., CARVALHAIS, N., ET AL. Deep learning and process understanding for data-driven earth system science. *Nature* 566, 7743 (2019), 195–204.
- [15] SMITH, J. Climate change: Scientific evidence and the industry of denial. *The Missouri Review* 40, 3 (2017), 187–201.
- [16] TOUYA, G., ZHANG, X., AND LOKHAT, I. Is deep learning the new agent for map generalization? *International Journal of Cartography* 5, 2-3 (2019), 142–157.
- [17] XU, P., DREDZE, M., AND BRONIATOWSKI, D. A. The twitter social mobility index: Measuring social distancing practices from geolocated tweets. *arXiv preprint arXiv:2004.02397* (2020).
- [18] ZHANG, L., ZHANG, L., AND DU, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4, 2 (2016), 22–40.
- [19] ZHU, X. X., TUIA, D., MOU, L., XIA, G.-S., ZHANG, L., XU, F., AND FRAUNDORFER, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 5, 4 (2017), 8–36.