
GWF
GEOSPATIAL WORLD FORUM
CLICK TO KNOW MORE

USC University of Southern California
Spatial Sciences Institute

GWF 2024
GEOSPATIAL WORLD FORUM

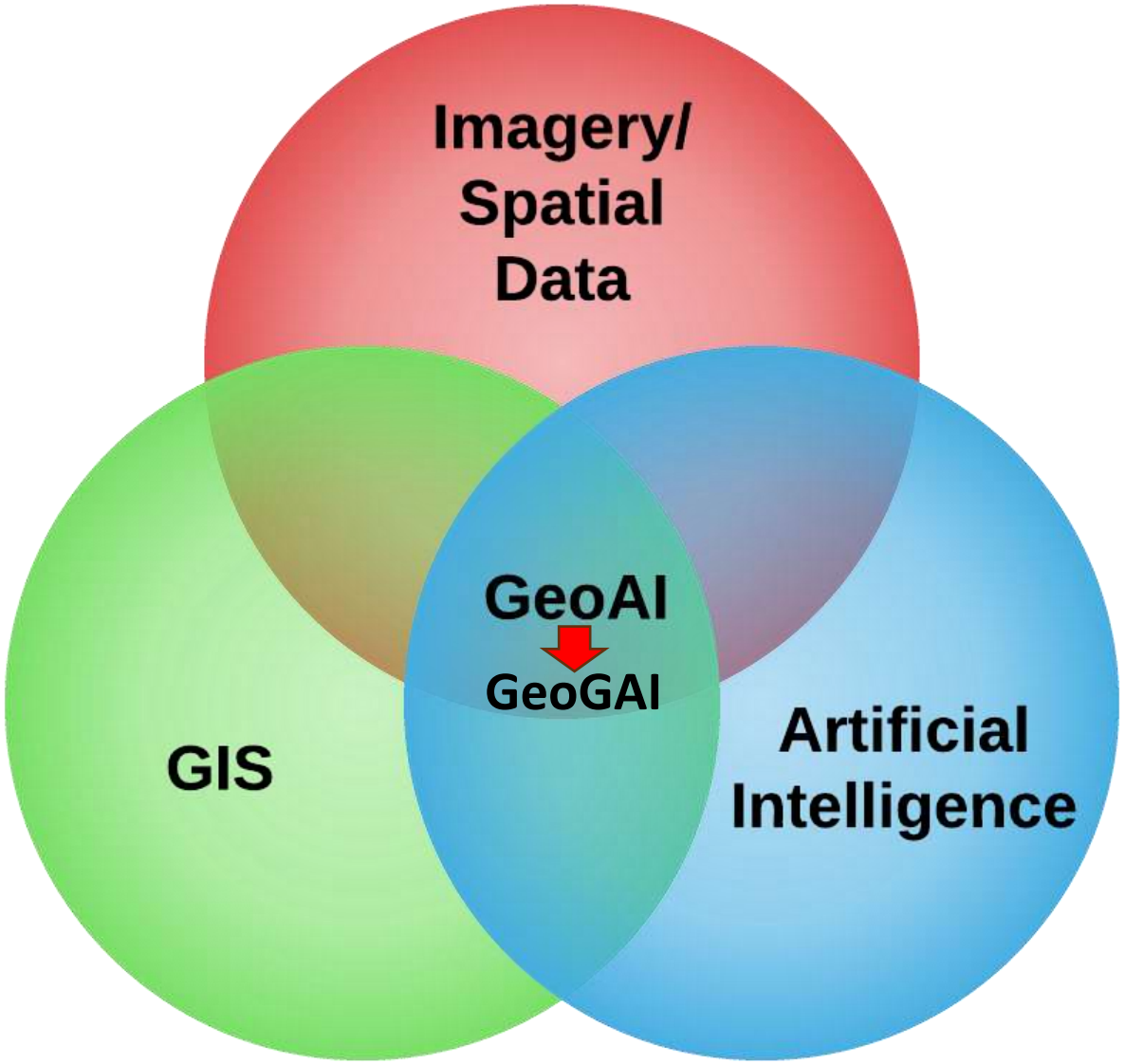
Geospatial World Forum 2024

GeoAI (Geospatial AI) and GeoGAI (Geospatial Generative AI) used in geospatial research: extensive systematic review

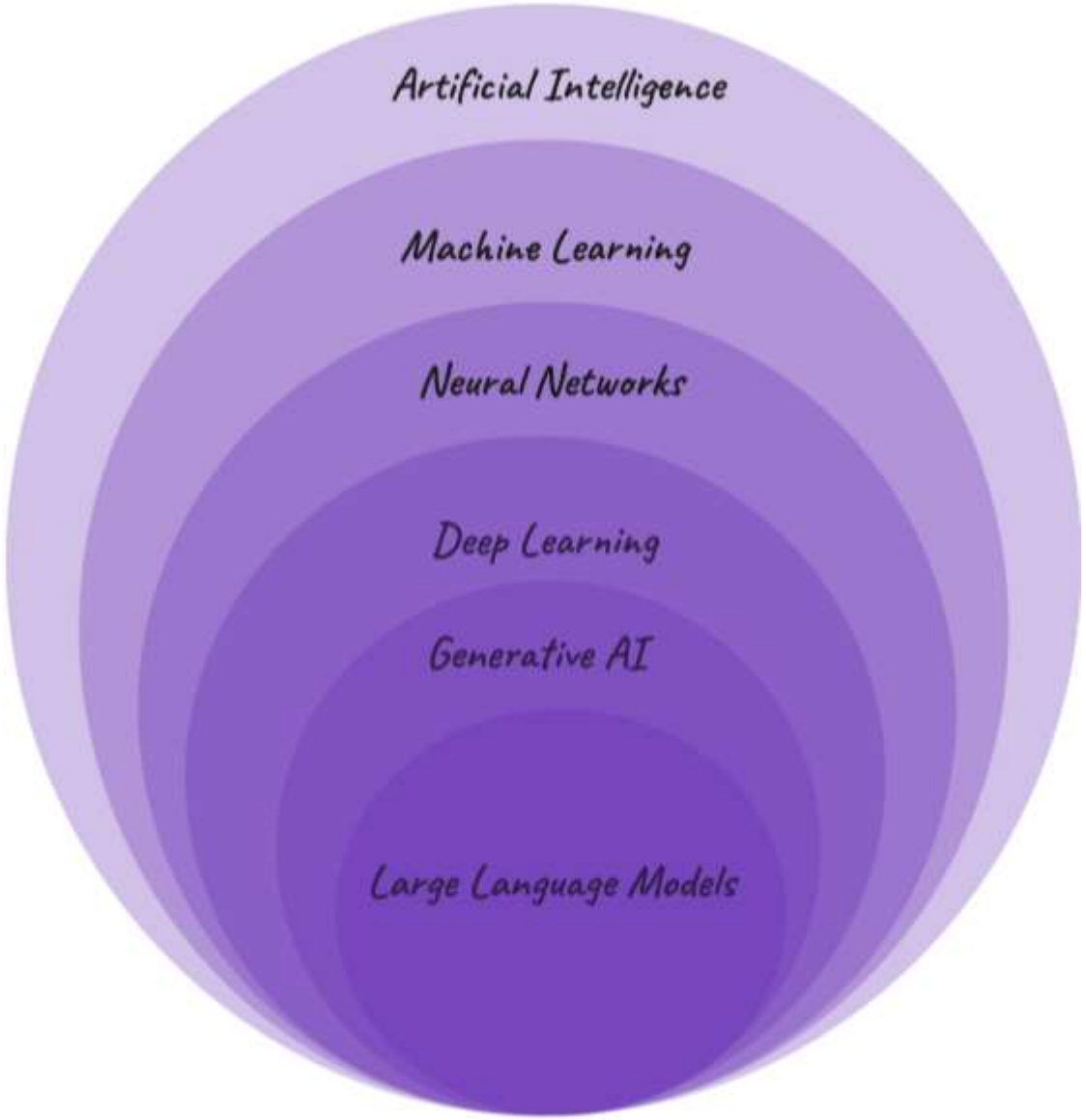
Dr Siqin (Sisi) Wang
Associate Professor
Spatial Sciences Institute
University of Southern California, US




What is GeoAI and GeoGAI ?



GeoAI = GIS + Artificial Intelligence



Broad AI Family

How GeoAI and GeoGAI have been used in geospatial research?

Review 1: (by May, 2023)

An extensive systematic review of GeoAI in Human Geography

Review 2: (by Dec, 2023)

A systematic review of GeoGAI (LLMs, ChatGPT, etc) in geospatial studies

Review 1: An extensive systematic review of GeoAI in Human Geography

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

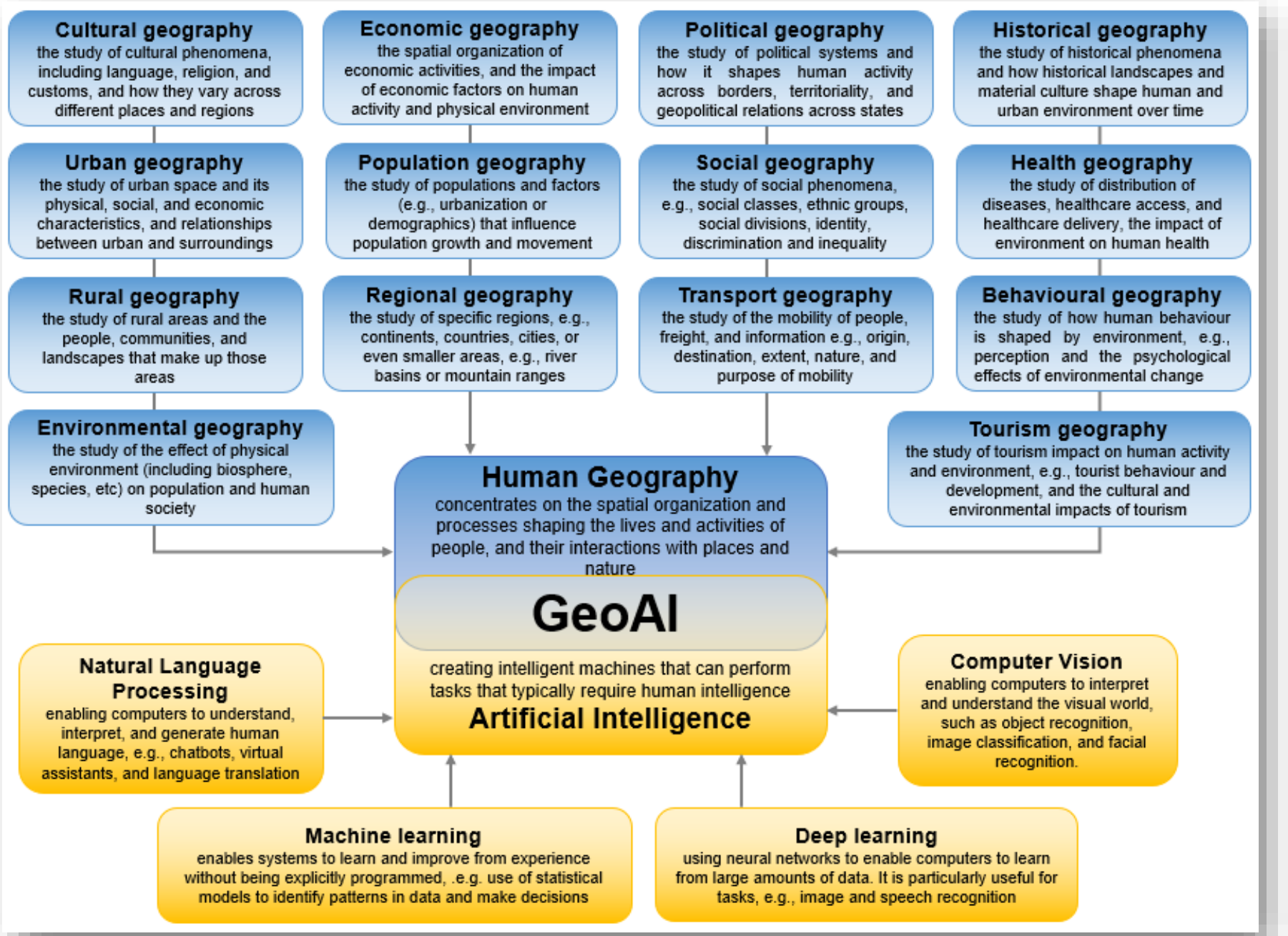
Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review

Siqin Wang [a,b,c,*], Xiao Huang [d,*], Pengyuan Liu [e], Mengxi Zhang [f], Filip Biljecki [g,h], Tao Hu [i], Xiaokang Fu [j,k], Lingbo Liu [j], Xintao Liu [l], Ruomei Wang [a], Yuanyuan Huang [b], Jingjing Yan [v], Jinghan Jiang [m], Michaelmary Chukwu [v], Seyed Reza Naghedi [v], Moein Hemmati [n], Yaxiong Shao [o], Nan Jia [p], Zhiyang Xiao [l], Tian Tian [q], Yaxin Hu [r], Lixiaona Yu [i], Winston Yap [g], Edgardo Macatulad [g], Zhuo Chen [s], Yunhe Cui [t], Koichi Ito [g], Mengbi Ye [g], Zicheng Fan [g], Binyu Lei [g], Shuming Bao [u]

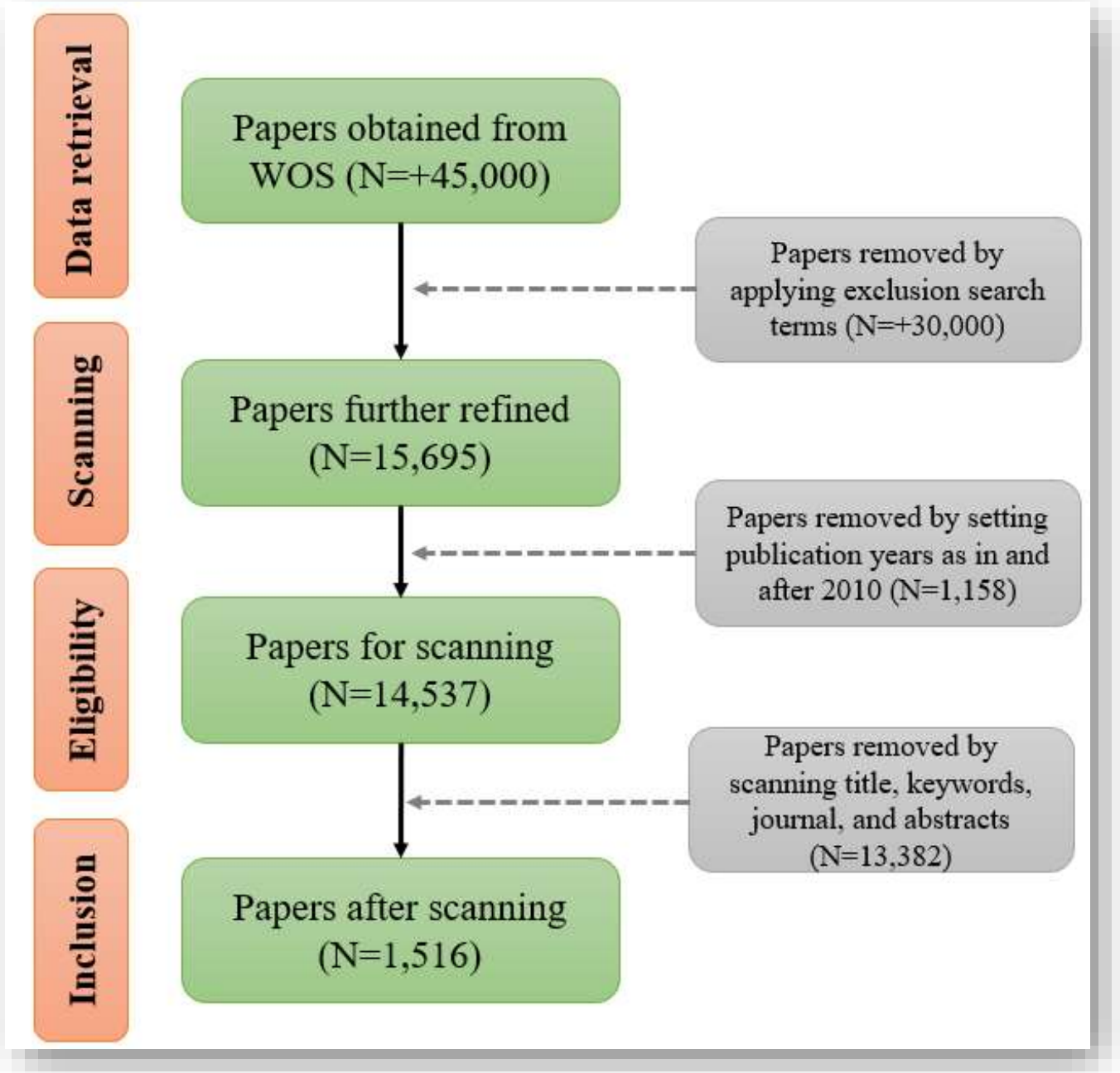
[a] Spatial Sciences Institute, University of Southern California, Los Angeles, United States
[b] University of Queensland, Brisbane, Australia
[c] School of Science, RMIT University, Melbourne, Australia
[d] Department of Environmental Sciences, Emory University, Atlanta, United States
[e] Future Cities Lab Global, Singapore-ETH Centre, Singapore
[f] Carilion School of Medicine, Virginia Tech, Blacksburg, VA, United States
[g] Department of Architecture, National University of Singapore, Singapore
[h] Department of Real Estate, National University of Singapore, Singapore
[i] Department of Geography, Oklahoma State University, Stillwater, OK, United States
[j] Centre for Geographic Analysis, Harvard University, Cambridge, MA, United States
[k] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China
[l] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hong Kong, China
[m] Department of Environmental Science Policy and Management, University of California Berkeley, Californa, United States
[n] Environmental Dynamics Program, Graduate School and International Education, University of Arkansas, Fayetteville, AR, United States
[o] Department of Earth Atmosphere and Environment, Northern Illinois University, Dekalb, IL, United States
[p] Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Michigan State University, Michigan, United States
[q] Department of Urban Planning, School of Urban Design, Wuhan University, Wuhan, China
[r] The Fu Foundation School of Engineering and Applied Science, Columbia University, New York, NY, United States
[s] Harrington Heart and Vascular Institute, University Hospitals, and School of Medicine, Case Western Reserve University, Cleveland, OH, United States
[t] Department of Geography, University of Connecticut, Storrs, CT, United States
[u] China Data Institute, Michigan, United States
[v] Department of Geosciences, University of Arkansas, AR, United States

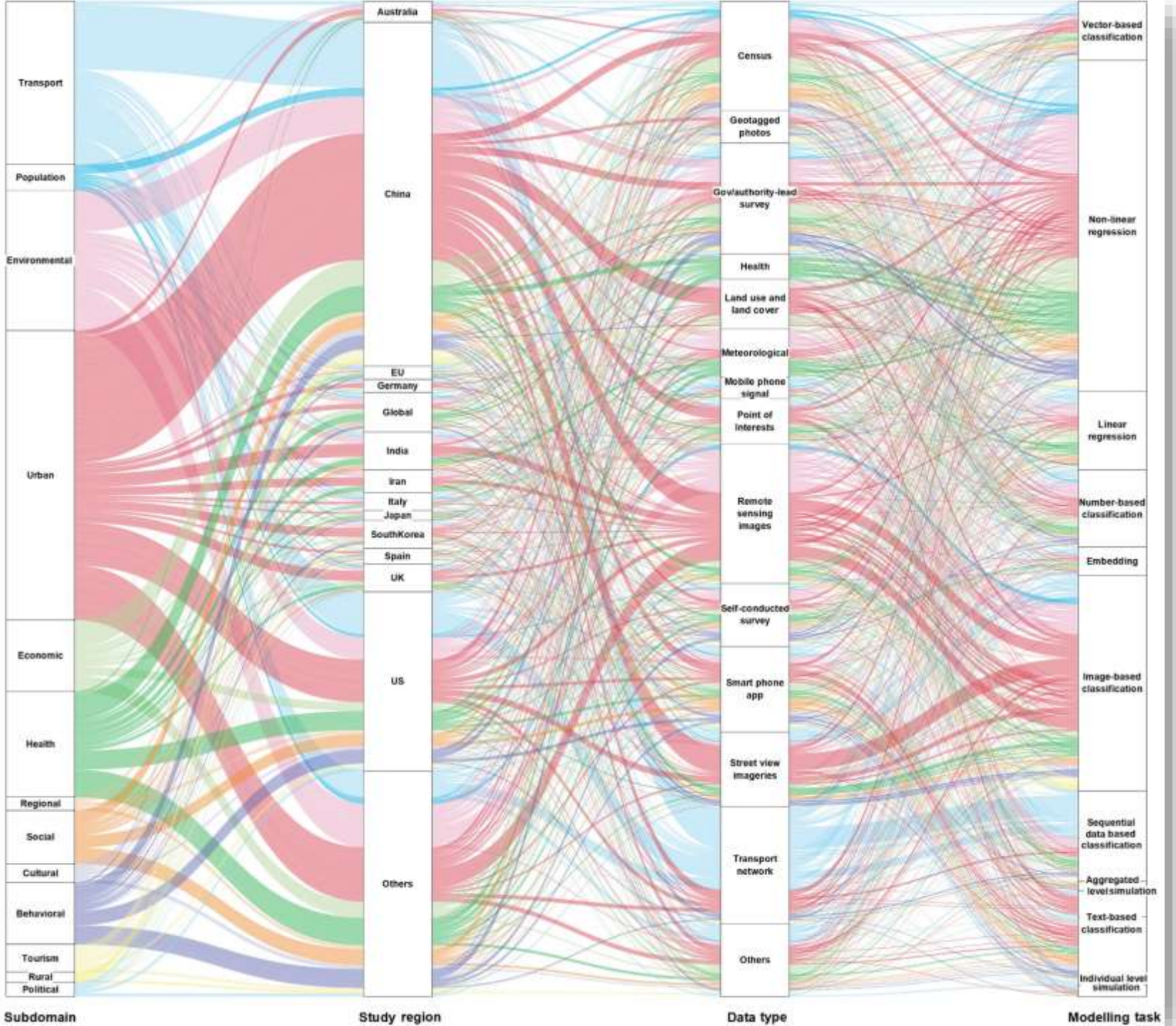
31 co-authors from 22 institutes in 5 countries/regions (US, Australia, China, Singapore, Hong Kong)

Keyword scoping



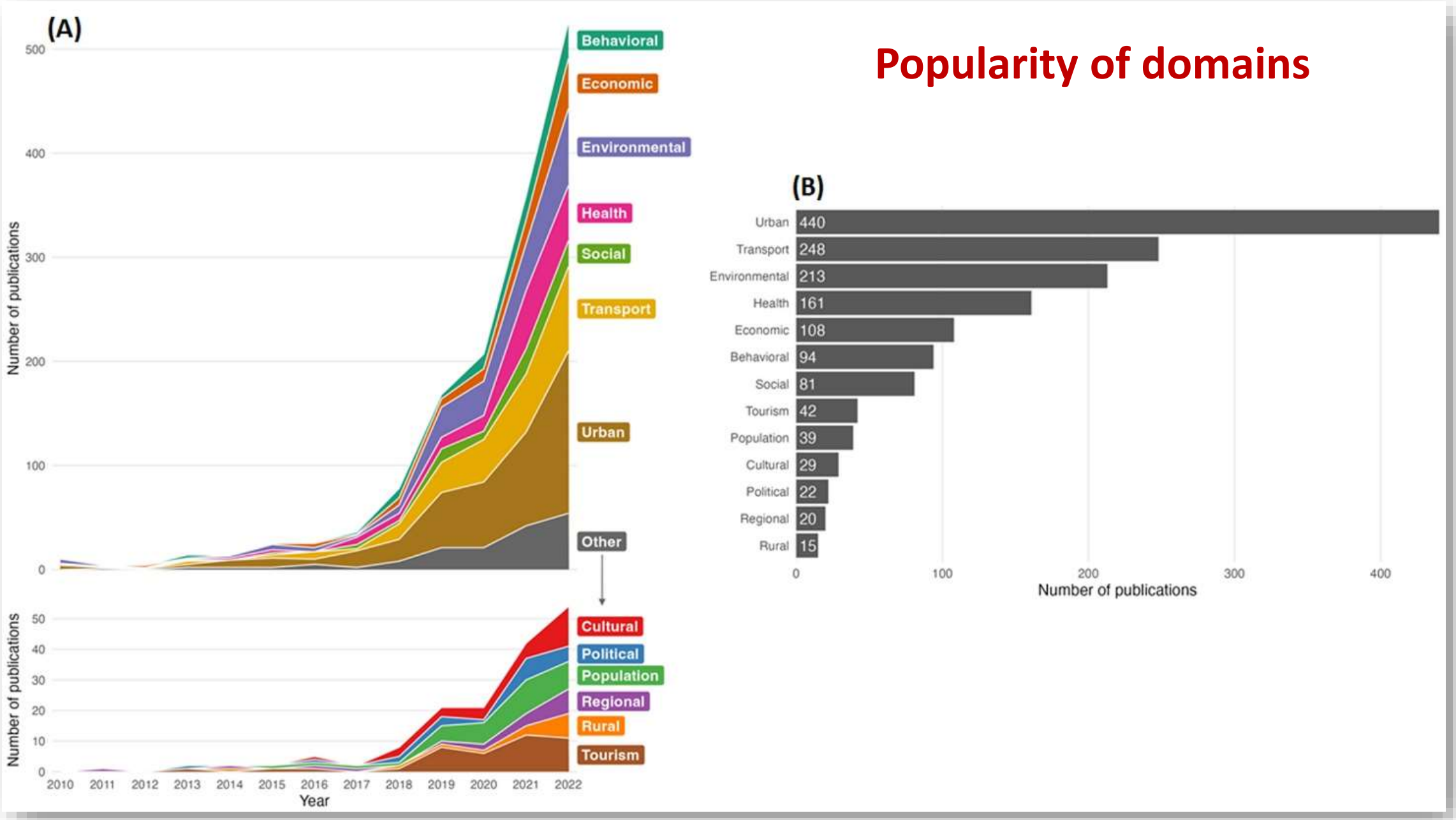
Scan 1,526 papers out of 15,695





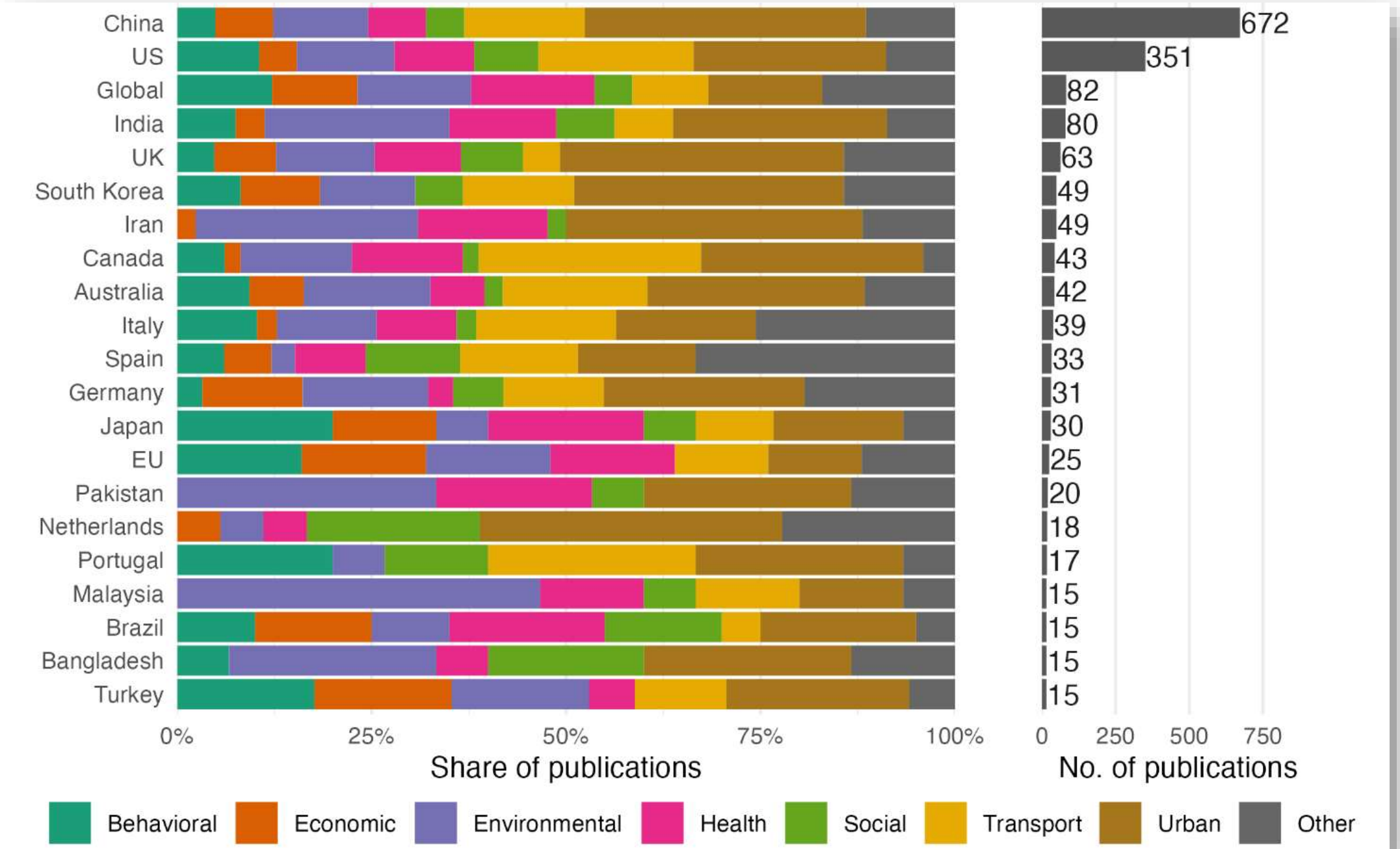
Sankey diagram showing the distribution of **study areas, data types and modeling tasks** GeoAI achieved in each subdomain of human geography

Popularity of domains

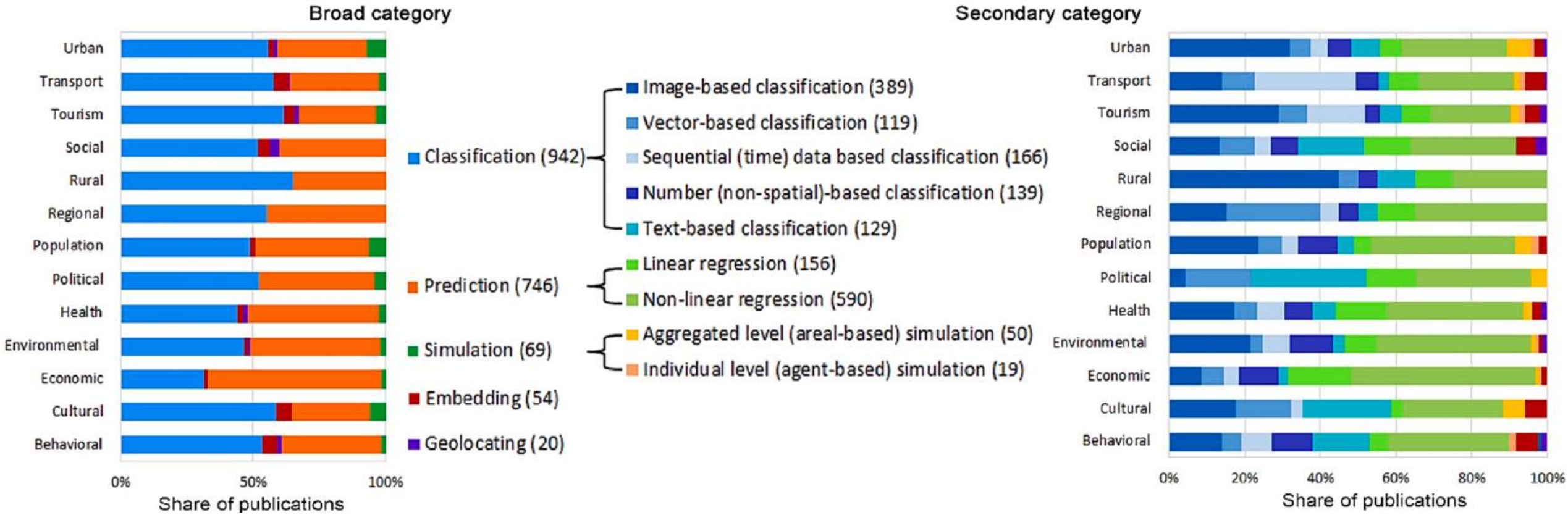


(A) Number of papers published over the years and in each subdomain by accounting for the primary subdomain (B) total number of papers published in each subdomain from 2010 to 2023.

Empirical studies in different countries



GeoAI Modeling Tasks



Broad category	Secondary category	Exemplary models and algorithms
GeoAI models Classification	Image-based	• Convolutional Neural Networks (CNNs) (LeNet, AlexNet, ResNet, VGG, Inception, EfficientNet, DenseNet, MobileNet) • Artificial Neural Network • Support vector machine (SVM) and one-class SVM • Gradient Boosting algorithms (GBM) (e.g., XGBoost, LightGBM, Catboost) • K-Nearest Neighbour • Naïve Bayes Algorithm • Deep Belief Networks (DBNs) • Autoencoder (AE) • Siamese networks • Isolation forest • Local outlier factor • Angle-based outlier detector • Histogram-based outlier detection • Autoencoders (variational types) • Hidden Markov models • Fuzzy logic-based outlier detection • Deep-learning based methods (Conditional neural network, RNN) • YoLo model family • R-CNN model family (R-CNN, Fast R-CNN, Mask R-CNN, R-FCN, Cascade R-CNN) • CenterNet model family (Single Shot Detector (SSD), DSSD, RON, CornerNet) • Histogram of Oriented Gradients (HOG) • Region-Based Segmentation • Edge Segmentation • K-Means • Convolutional Encoder-Decoder Architecture (e.g. SegNet, U-Net, Fully Convolutional Networks (FCN) • Multi-Scale and Pyramid Network Based Models (FPN) • Pyramid Scene Parsing Network (PSPNet), Mask R-CNN, Fast R-CNN) • Dilated Convolutional Models and DeepLab Family
	Vector-based	• Spatially constrained multivariate clustering • Multivariate clustering • Density-based clustering • Image segmentation • Hot spot analysis • Cluster and outlier analysis • Space-time pattern mining • Hierarchical clustering analysis (HCA) • Density-based spatial clustering of applications with noise (DBSCAN) • Spectral clustering • Affinity propagation (AP) • Gaussian mixture model (GMM)
	Sequential (time) data-based	• Hidden Markov Models • Long Short-Term Memory networks (LSTM) • Recurrent Neural Networks (RNN) • Conditional Random Fields

	Number (non-spatial)-based	• Support vector machine (SVM) • Gradient Boosting algorithms (GBM) (e.g., XGBoost, lightGBM, CatBoost) • Decision tree / Random Forest • Means algorithm • Fuzzy logic-based algorithms • DBSCAN • Spectral clustering • Hierarchical clustering • Affinity Propagation
	Text-based	• Latent Dirichlet allocation (LDA) / RNN • Word2Vec • Doc2Vec • Bag-of-words model • n-gram model • Transformers-based methods (BERT, XLM, GPT, RoBERTa, XLNet, DistilBERT etc) • ELMo • RNN • LSTM • Word2Vec • Doc2Vec • Bag-of-words model • n-gram model • Transformers-based methods (BERT, XLM, GPT, RoBERTa, XLNet, DistilBERT etc) • ELMo
Prediction	Linear	• Generalized linear model (GLM), including Lasso regression, Ridge regression, Polynomial Regression, Bayesian linear regression; Logistic regression, Gamma regression, Poisson regression, Bernoulli regression, Binomial regression, Multinomial regression, Exponential regression, (Inverse) Gaussian regression
	Non-linear	• Support vector machine • Artificially Neural Network (ANN) • Gradient Boosting algorithms (GBM) (e.g., XGBoost, lightGBM, CatBoost) • Empirical Bayesian Kriging regression prediction • Forest based prediction (random forest, decision tree) • Graph Convolutional Neural Network • Generalised additive model (GAM) and GeoGAM • Bayesian hierarchical model (BHM) • Second-dimension spatial association • Geographically optimal similarity model
Simulation	Aggregated level (areal-based)	• Cellular Automata • Deep neural network • Deep enforcement learning • Tabular Q-learning
	Individual level (agent-based)	• Agent based modelling
Embedding	—	• Principal component analysis (PCA) • Independent Component Analysis (ICA) • Linear Discriminant Analysis (LDA) • Locally Linear Embedding (LLE) • t-distributed Stochastic Neighbour Embedding (t-SNE) • Auto-encoder model family

Review 2: A systematic review of GeoGAI (LLMs, ChatGPT, etc) in geospatial studies

The promise of using GPT and generative artificial intelligence (GAI) models in geospatial science: a systematic review

Siqin Wang[1, 2, 3+], Tao Hu*[4+], Huang Xiao[5+], Yun Li[6+], Ce Zhang*[7+], Huan Ning[8], Rui Zhu[7], Zhenlong Li[8], Xinyue Ye[9]

[1] Spatial Sciences Institute, University of Southern California, Los Angeles, US. siqinwan@usc.edu

[2] School of Earth and Environmental Sciences, The University of Queensland, Brisbane, Queensland, Australia. s.wang6@uq.edu.au

[3] School of Science, Royal Melbourne Institute of Technology (RMIT), Melbourne, Australia.

[4] Department of Geography, Oklahoma State University, Oklahoma, US, tao.hu@okstate.edu

[5] Department of Environmental Sciences at Emory University, Atlanta, US, xiao.huang2@emory.edu

[6] Department of Computer Science, Emory University, Atlanta, US. yli230@emory.edu

[7] School of Geographical Sciences, University of Bristol, Bristol, UK.

C.Z: ce.zhang@bristol.ac.uk; R.Z: rui.zhu@bristol.ac.uk

[8] Geoinformation and Big Data Research Lab, Department of Geography, The Pennsylvania State University, University Park, US.

H.N.: hning@psu.edu; Z.L.: zhenlong@psu.edu

[9] Department of Landscape Architecture and Urban Planning & Departemnt of Computer Science & Engineering, Texas A&M University, College Station, TX, US, xinyue.ye@tamu.edu

Accepted by *International Journal of Digital Earth* – to be online soon

Keyword scoping

Keywords:

1) paper topics (including article titles, abstracts, and keywords) relevant to large pretrained models, including "GPT", "generative artificial intelligence", "generative AI", "GAI", "artificial general intelligence", "AGI", "large language model" and "foundation model";

2) paper topics relevant to geospatial science, including "geospatial", "geograph*“, "spatial" OR "spatialtemporal" and "spatiotemporal";

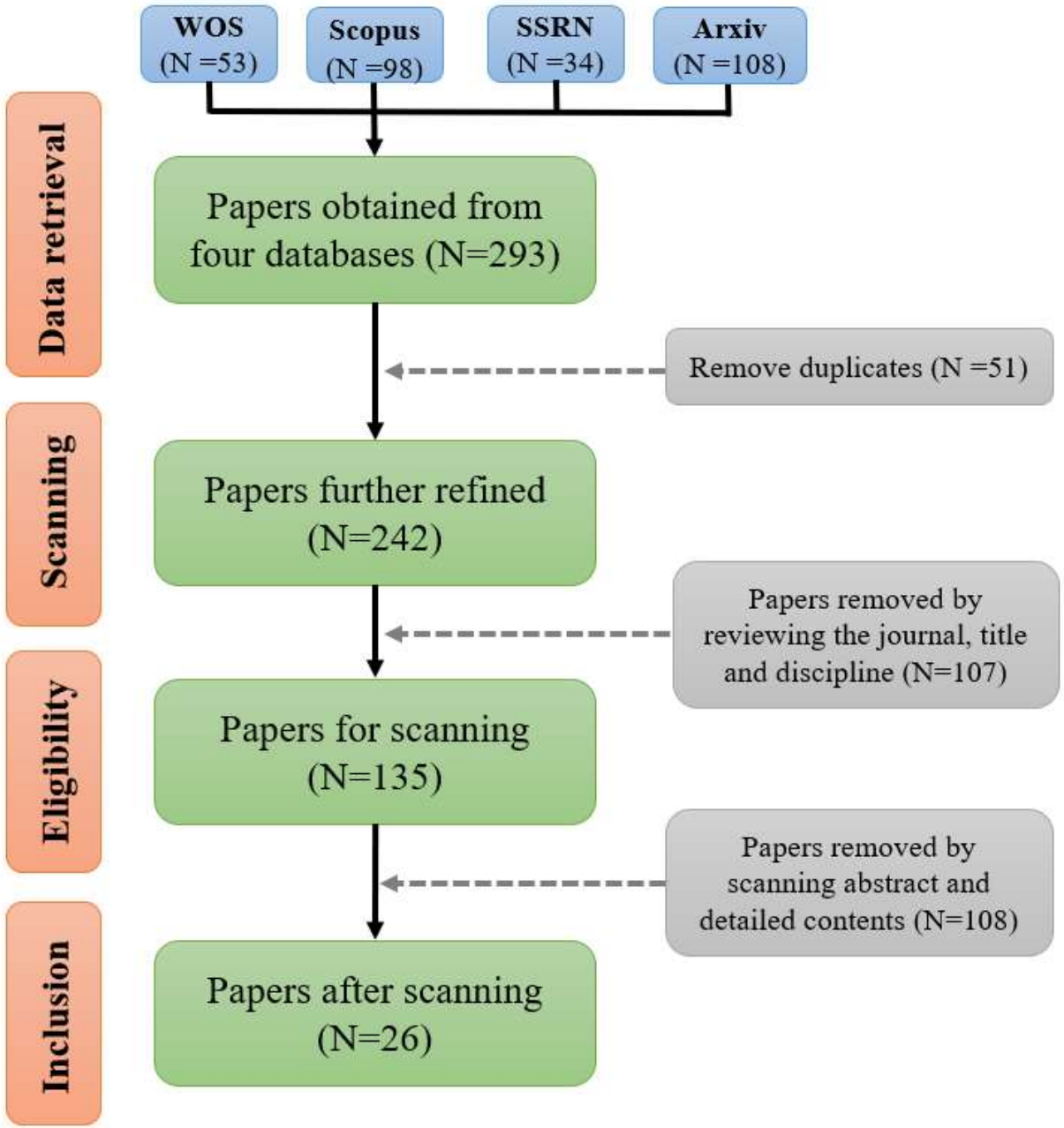
Database:

1) WOS and Scope, article types defined as “peer-reviewed journal articles” and “conference preceding papers”;

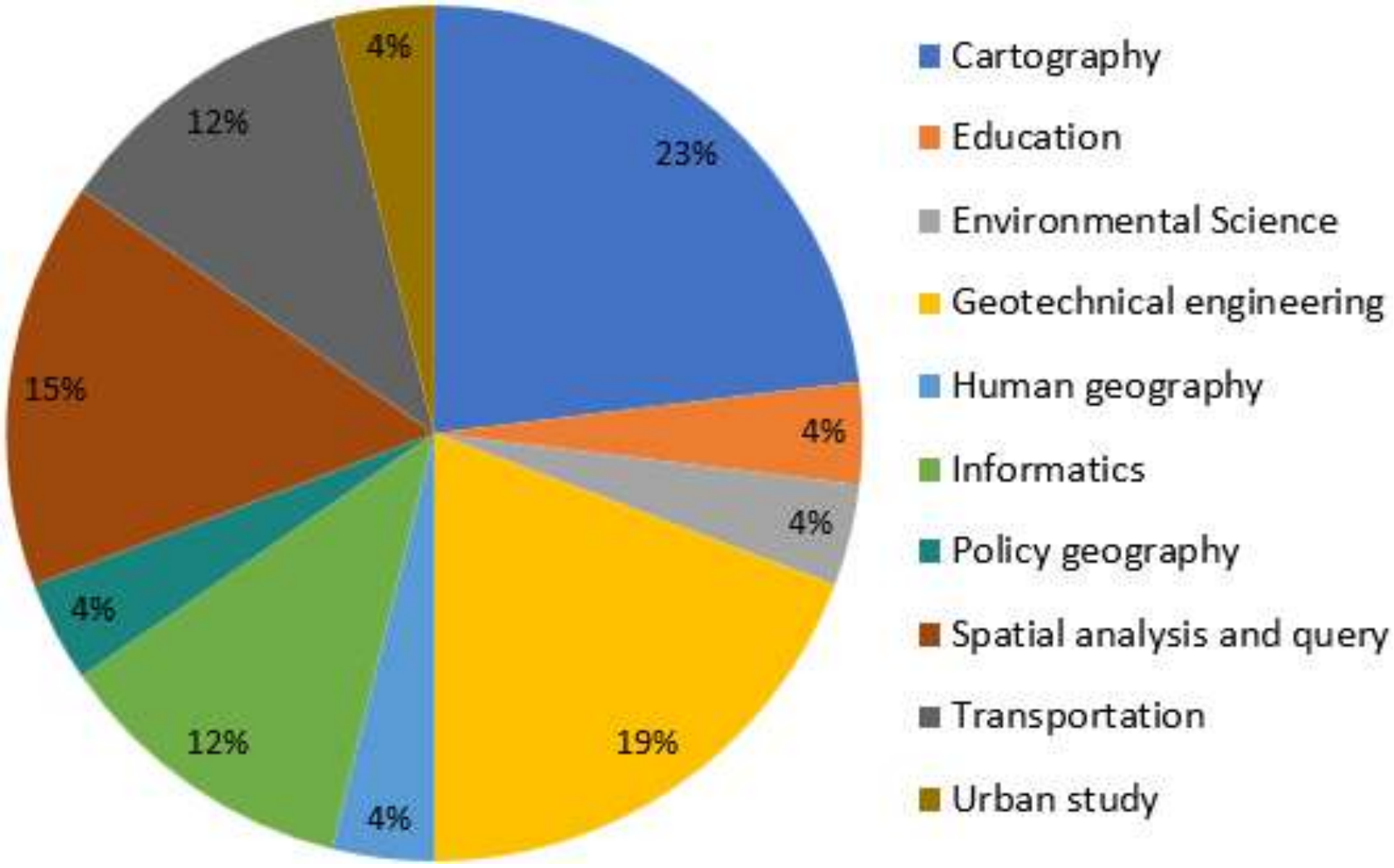
2) SSRN and arXiv, article types include “preprint”, “conference papers” and “ongoing papers”;

Timeframe: after Nov, 2022

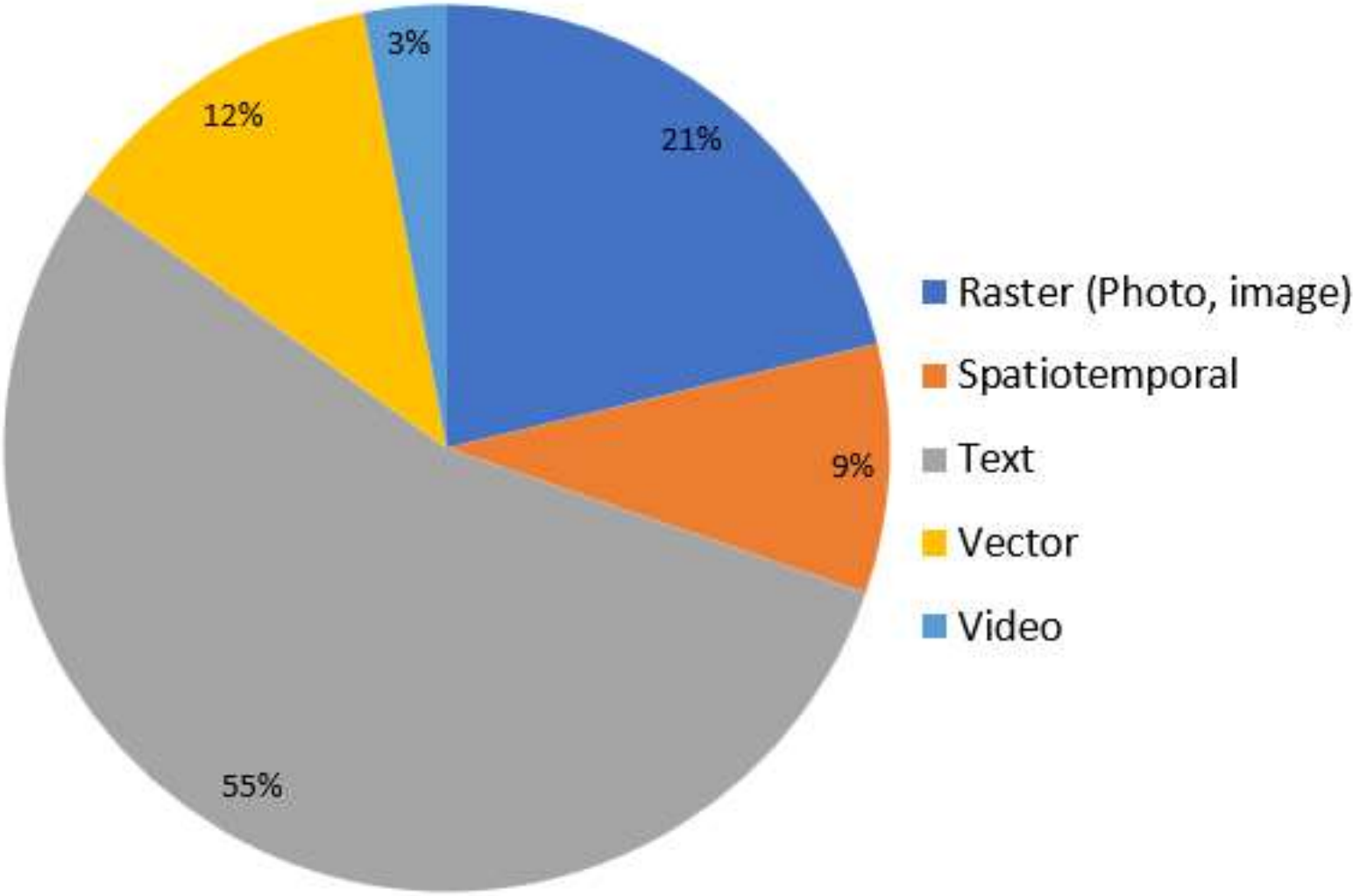
Scan 26 papers out of 293



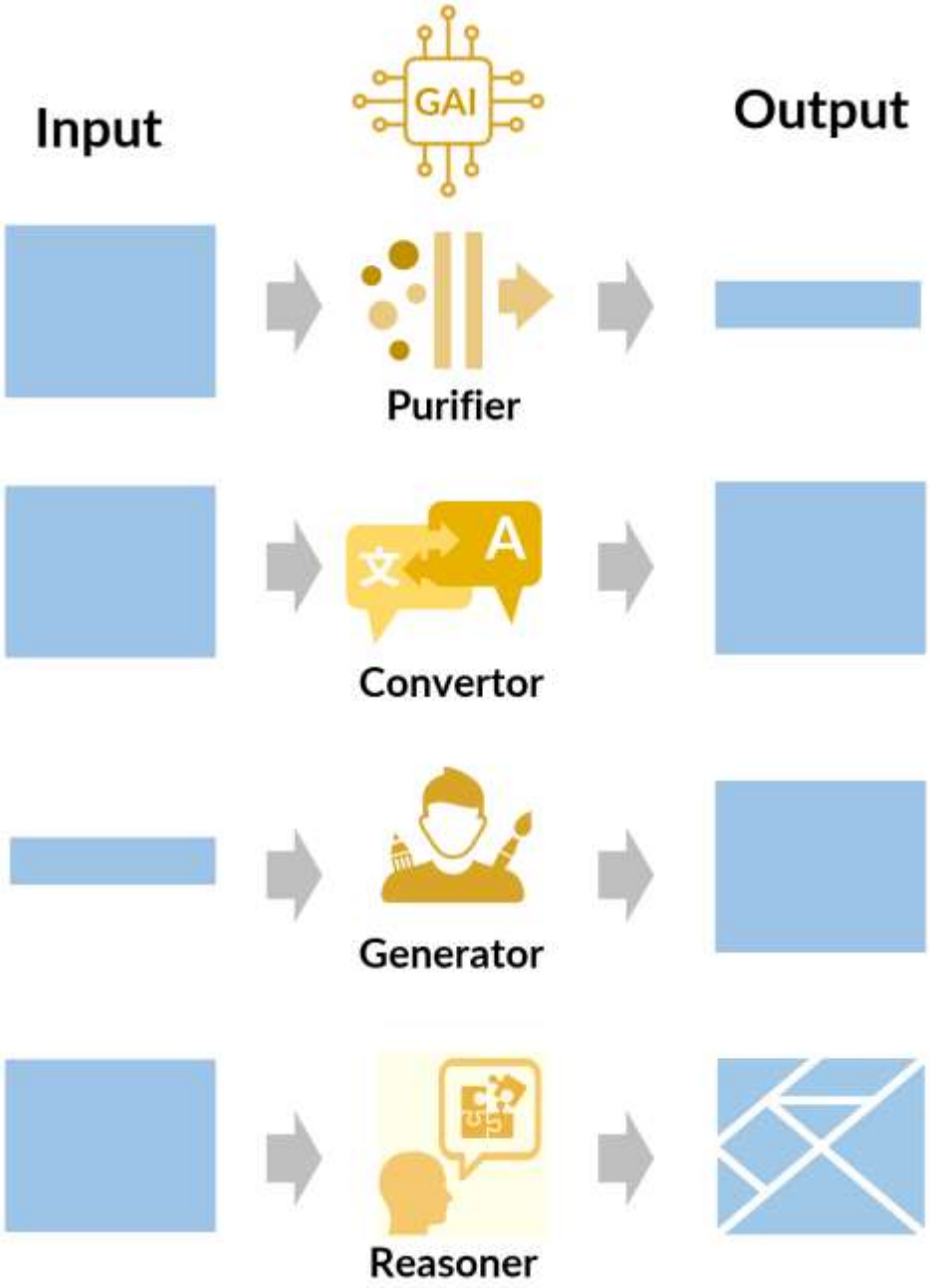
Domains of GeoGAI



Data types used in GeoGAI



Four roles that GeoGAI play



Role	Information source in the GAI output	Implication
Purifier	Input data	Feature engineering may be helpful since it can amplify the interested characteristics of the input data to facilitate the GAI extraction.
Converter	Input data	Adopting appropriate embeddings or representations for input and output data.
Generator	Training data	the variety and sufficiency of the training data and the model capability
Reasoner	Input data, training data (especially text and code)	Providing necessary and sufficient information in the input for GAI's further reasoning.

Future directions for GeoGAI

1. Future GAI functions to be integrated with geospatial science

- *Enhancing geospatial data query and recommendation*
- *Building a robust LLM-based assessment framework*

2. Multi-modal foundation models to develop highly effective GeoAI

- *Feasibility of cross-scale geospatial science research*
- *Next-generation autonomous geospatial science*

3. From GeoGAI to GeoAGI (artificial general intelligence): a promising way to go

- *Bridging human cognition and geospatial intelligence*
- *Pioneering intelligent disaster management*

4. Ethnical and privacy concerns: old problems in a new disguise

Q&A

Welcome any comments/questions

siqinwan@usc.edu