Towards Urban General Intelligence Through Urban Foundation Models

Hao Liu, Weijia Zhang, Jindong Han,
Zhao Xu, Hang Ni, Hui Xiong

The Hong Kong University of Science and Technology (Guangzhou) &
The Hong Kong University of Science and Technology

KDD2024
BARCELONA, SPAIN

Tutorial Website

Survey Paper

Github

OUTLINE

Introduction

Challenges of Building UFMs

Overview of UFMs

Prospects of UFMs

Summary

OUTLINE

- **Introduction**
- Challenges of Building UFMs
- Overview of UFMs
- Prospects of UFMs
- Summary

City Evolution – Challenge & Opportunity

- City, as a dynamically evolving entity, is driven by a multitude of factors, such as population growth, environmental changes, and technological advancements.

Challenge

Traffic jam

Energy crisis

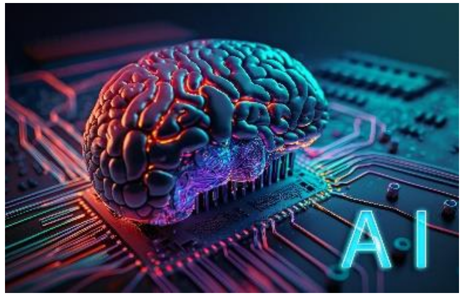
Climate change

…

Opportunity

Urban big data

+



Artificial Intelligence

→

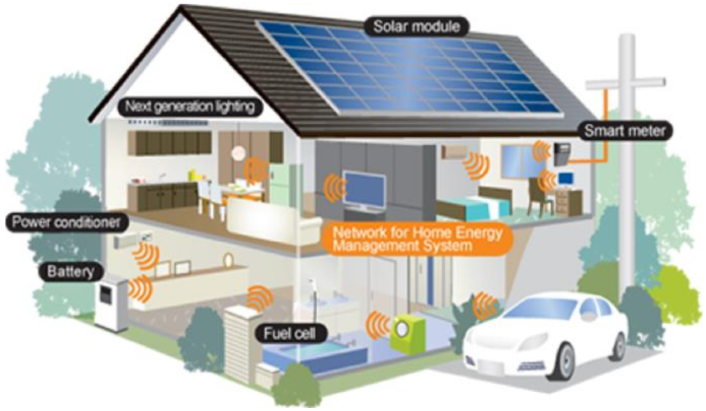
Smarter & sustainable city

United Nation Released the 17 Sustainable Development Goals (SDGs). https://sdgs.un.org/goals

Smart City and Urban Intelligence

- Smart city has deeply penetrated into our daily life.

Intelligent transportation systems

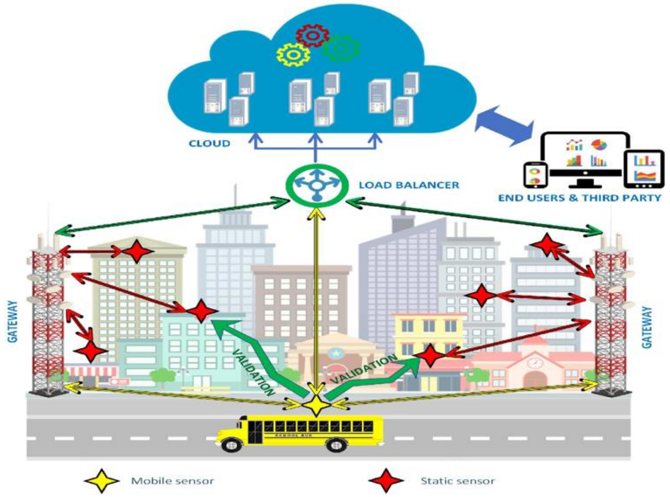


Smart building

E-Government Services

Urban surveillance

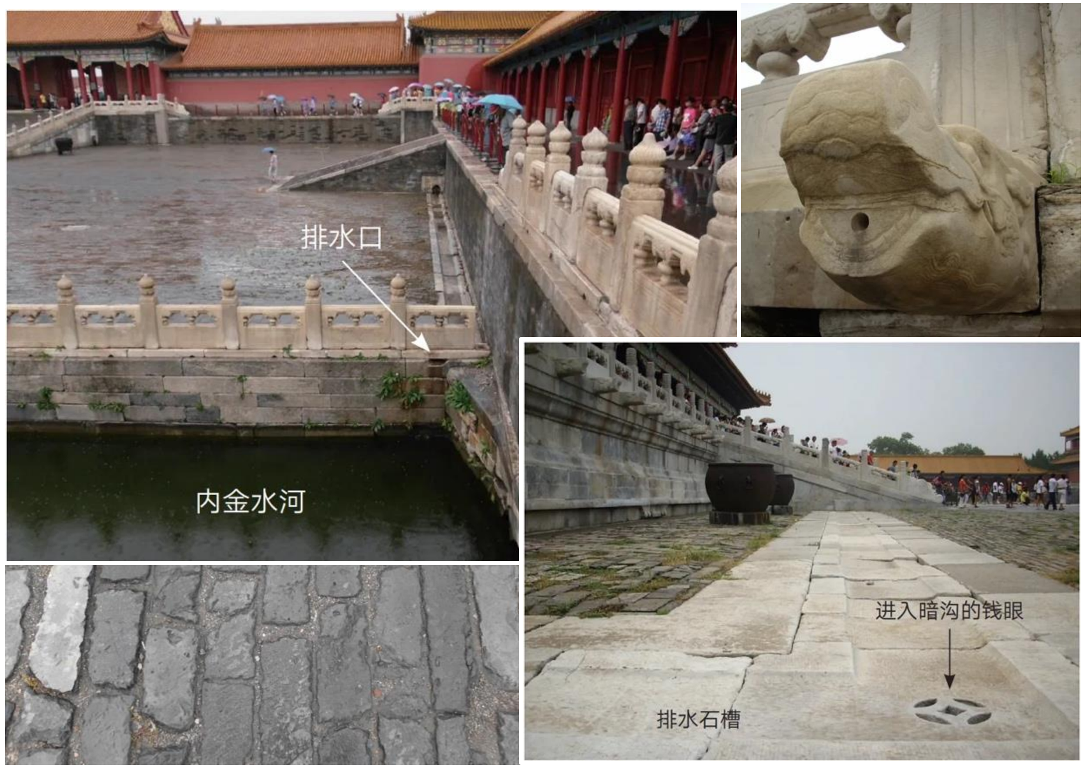
Smart grid



Air pollution analysis

Urban Intelligence: Are We There Yet?

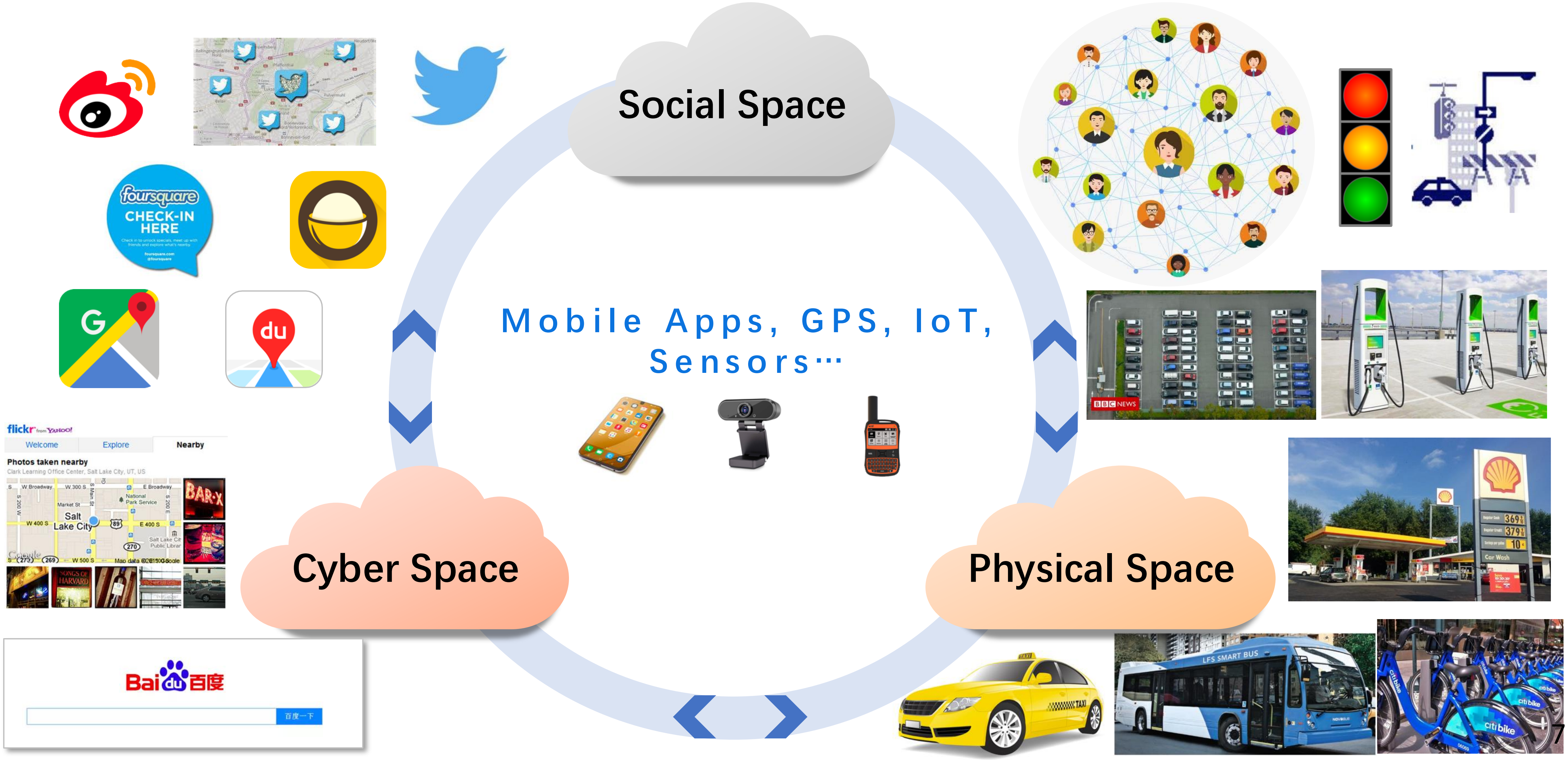
- Urban Intelligence (UI) refers to the strategic integration of various technologies and practices to enhance the management, operation, and development of urban areas.



Drainage System in Forbidden City

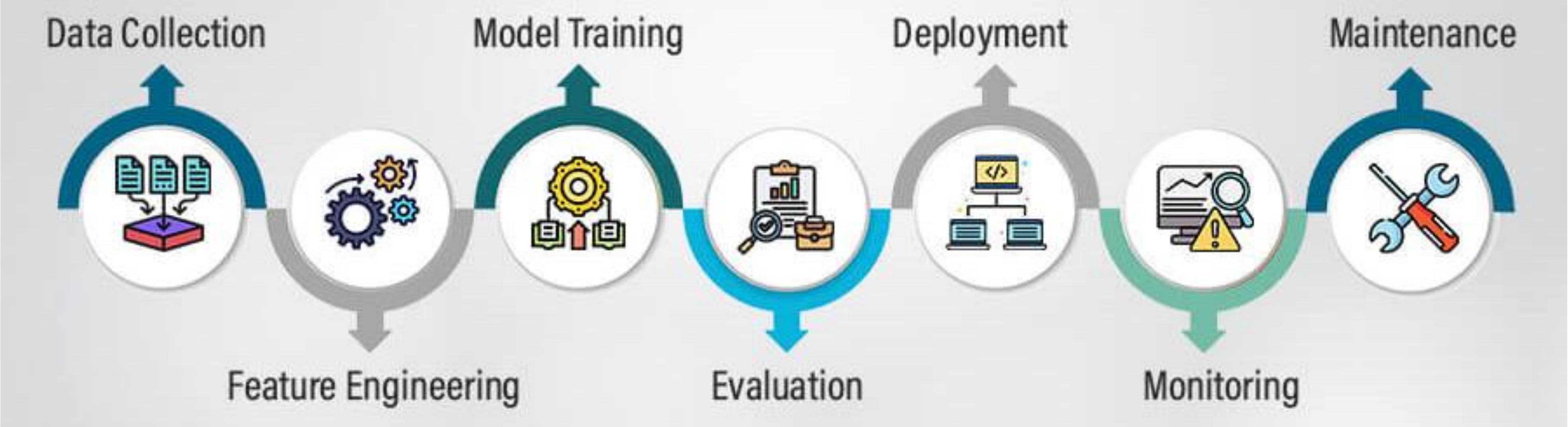
Beijing flood, 2023

Limitation 1: Interconnected Urban Big Data



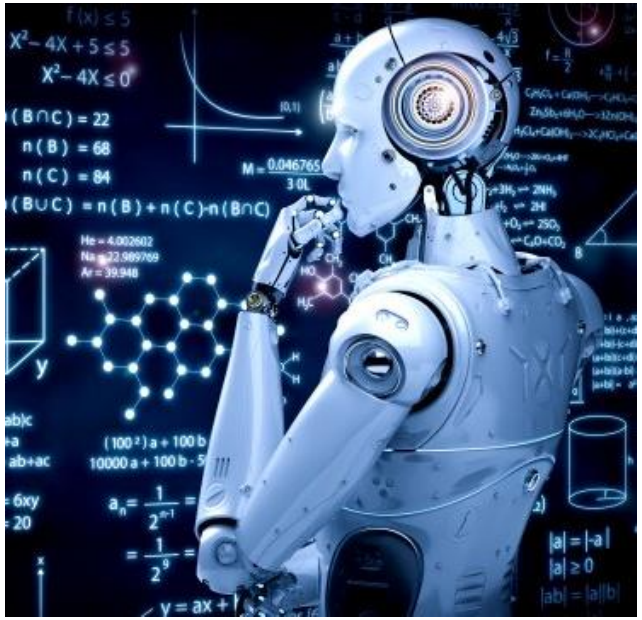
Limitation 2: Task Specific Machine Learning Modeling

- Most data-driven Urban Intelligence solutions follow the machine learning development pipeline
  - Requires abundant labeled data for training -> limited budget/ authority
  - Possess specific skill for designated tasks -> limited generalizability
  - The limit of the model approx. to the quality of the data -> limited knowledge



Urban General Intelligence (UGI)

- Analogous to Artificial General Intelligence (AGI), UGI is envisioned to autonomously perform any intellectual task related to urban contexts, rivaling or even surpassing human capabilities, thereby transforming cities into more livable, resilient, and adaptive spaces.



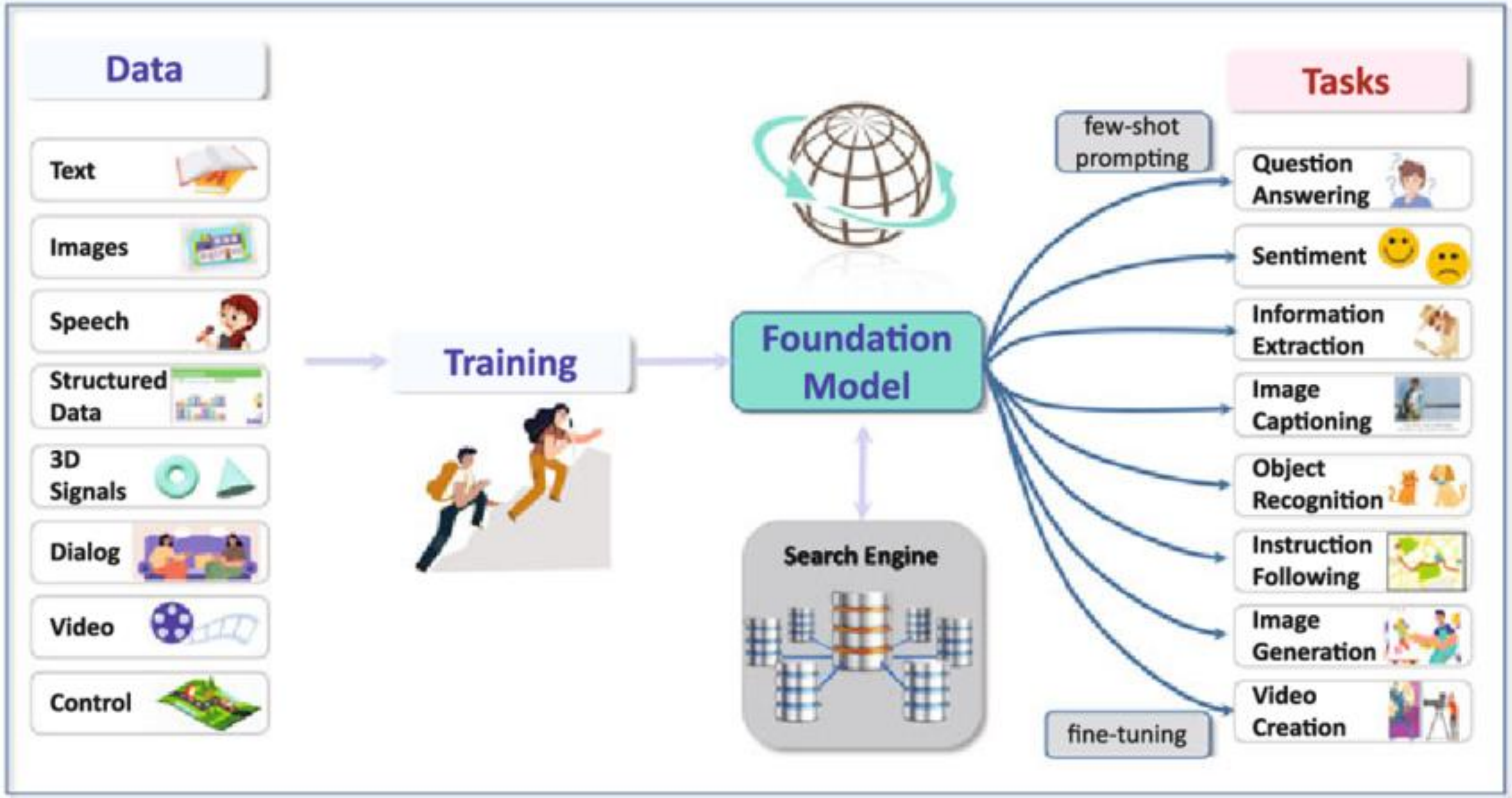
AGI

VS.

UGI

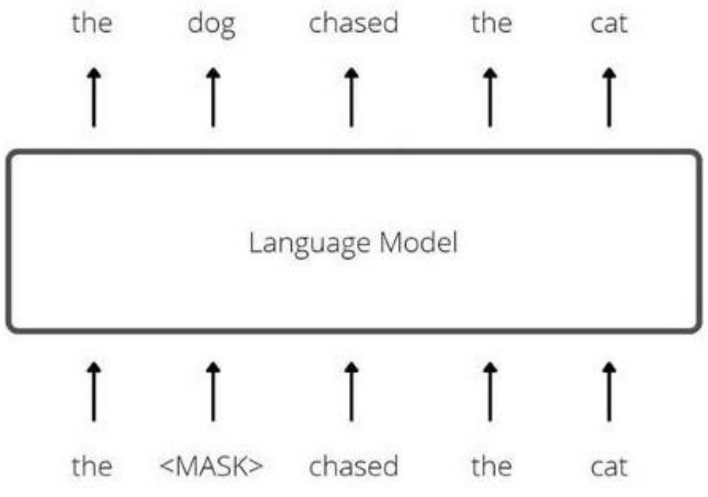
Towards UGI: Are Foundation Models a Potential Solution?

- Recent advances on Large Language Model (e.g., ChatGPT) and Vision Transformer (e.g., SAM) demonstrate remarkable generalization and emergent abilities on a wide spectrum of tasks.

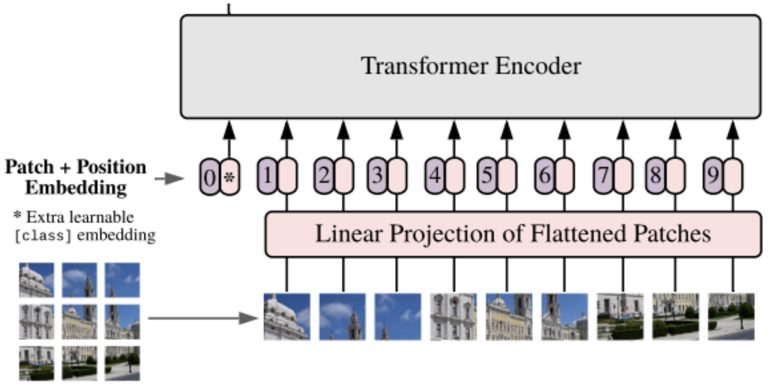


Basics of Foundation Model

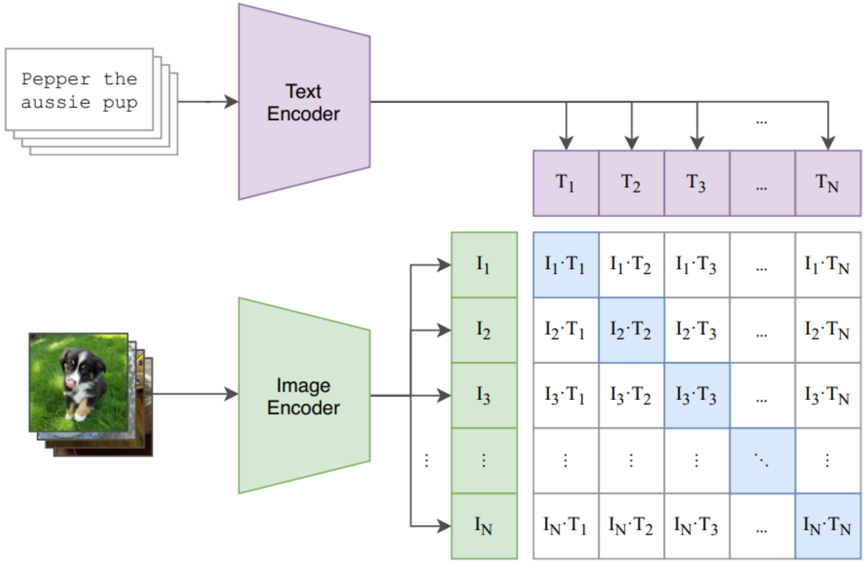
- Pretraining of language, vision, and multi-modal foundation models.



Masked and next token prediction


Transformer Encoder
Patch + Position Embedding
* Extra learnable [class] embedding
0
*
1
2
3
4
5
6
7
8
9
Linear Projection of Flattened Patches

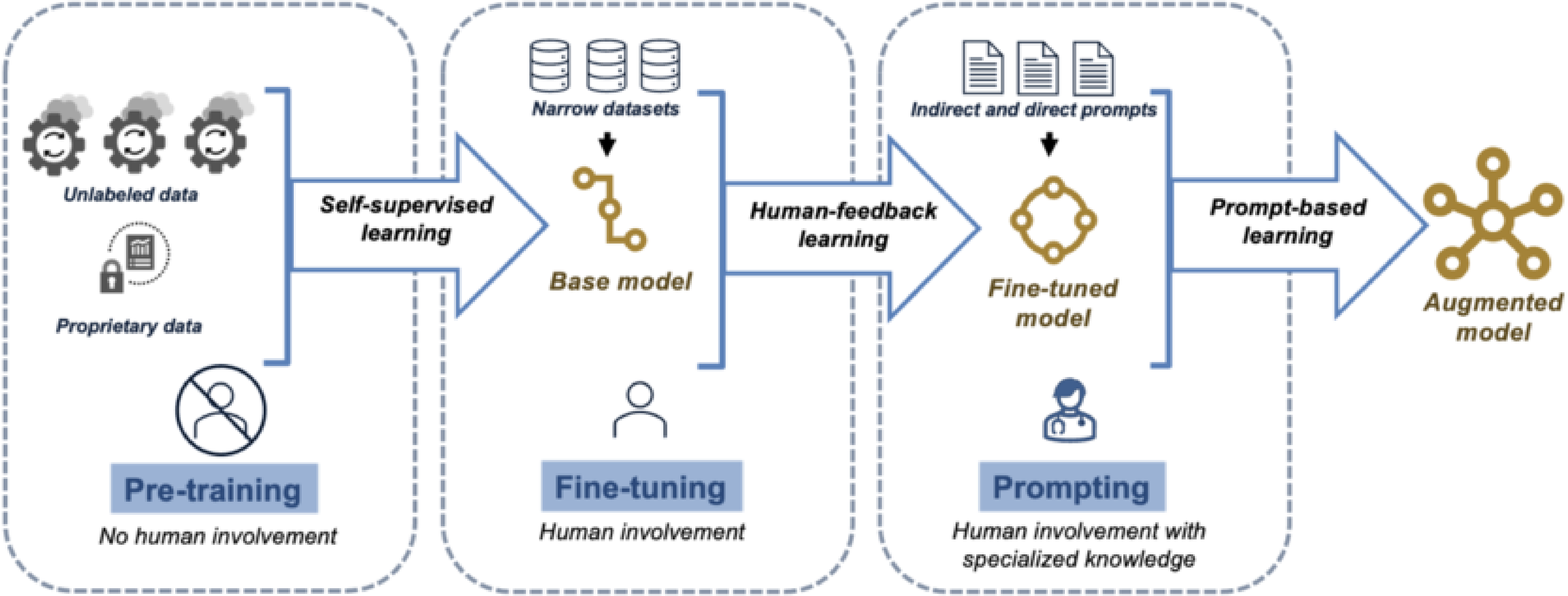
Masked patch prediction



Cross-modal alignment

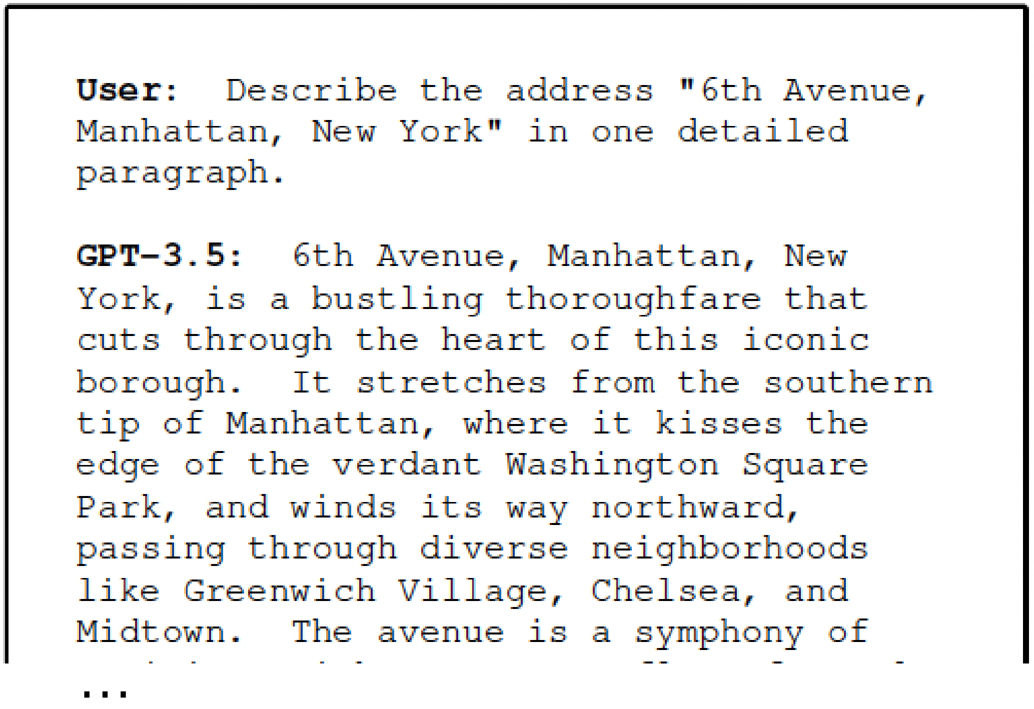
Basics of Foundation Model

- Typical foundation model training and serving stages.



Can LLMs Memorize Spatial Knowledge?

- LLM has compressed world knowledge in training corpus obtained from the Internet
  - Does LLM memorized geospatial knowledge? If yes, how to extract them?

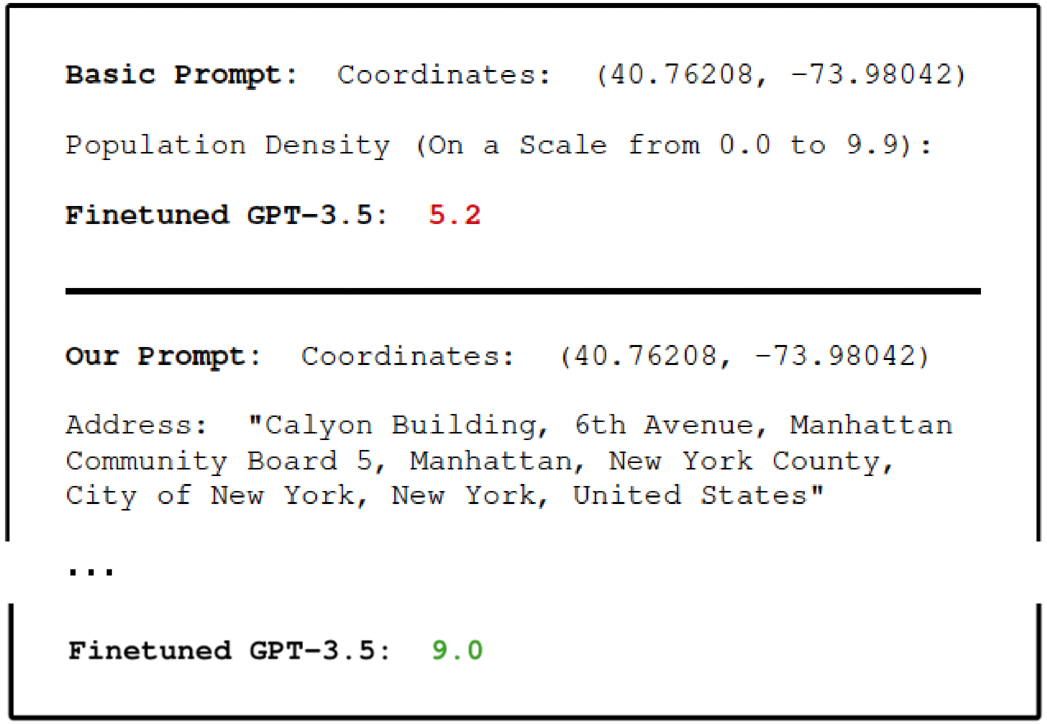


User: Describe the address "6th Avenue, Manhattan, New York" in one detailed paragraph.

GPT-3.5: 6th Avenue, Manhattan, New York, is a bustling thoroughfare that cuts through the heart of this iconic borough. It stretches from the southern tip of Manhattan, where it kisses the edge of the verdant Washington Square Park, and winds its way northward, passing through diverse neighborhoods like Greenwich Village, Chelsea, and Midtown. The avenue is a symphony of

...

GPT-3.5 can describe addresses



Basic Prompt: Coordinates: (40.76208, -73.98042)

Population Density (On a Scale from 0.0 to 9.9):

Finetuned GPT-3.5: 5.2

Our Prompt: Coordinates: (40.76208, -73.98042)

Address: "Calyon Building, 6th Avenue, Manhattan Community Board 5, Manhattan, New York County, City of New York, New York, United States"

...

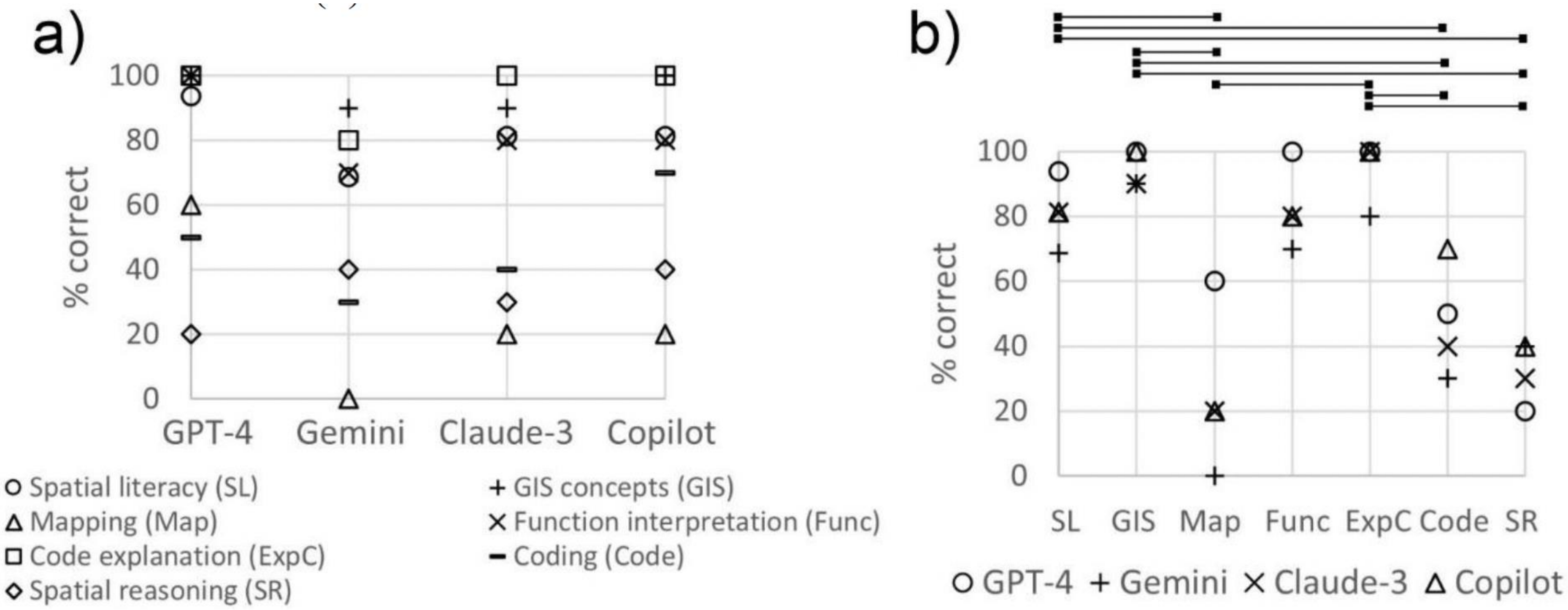
Finetuned GPT-3.5: 9.0

Prompting with additional map info

Rohin Manvi, et al. GeoLLM: Extracting Geospatial Knowledge from Large Language Models. ICLR 2024.

Quantitative Study of LLM's Spatial Capability?

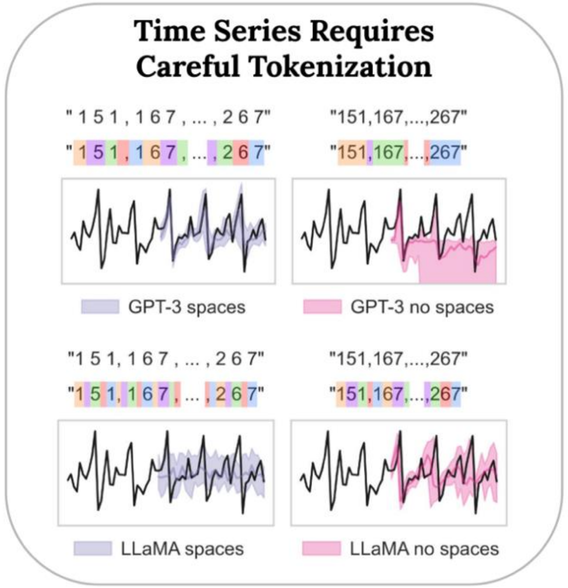
- LLM generally performed well on tasks related to spatial literacy, GIS theory, but revealed weaknesses in mapping and spatial reasoning

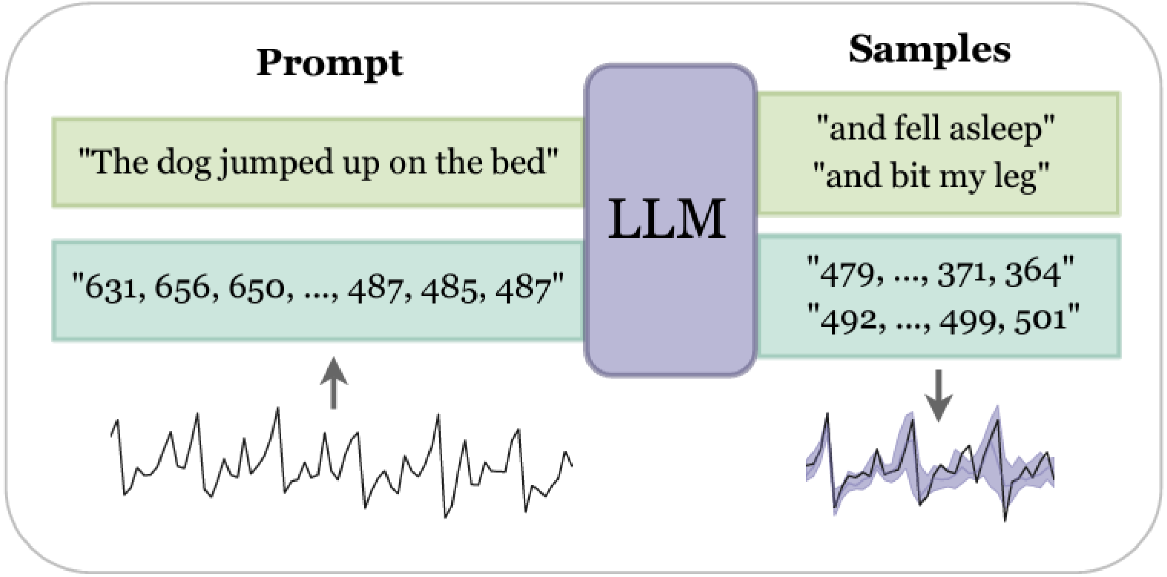


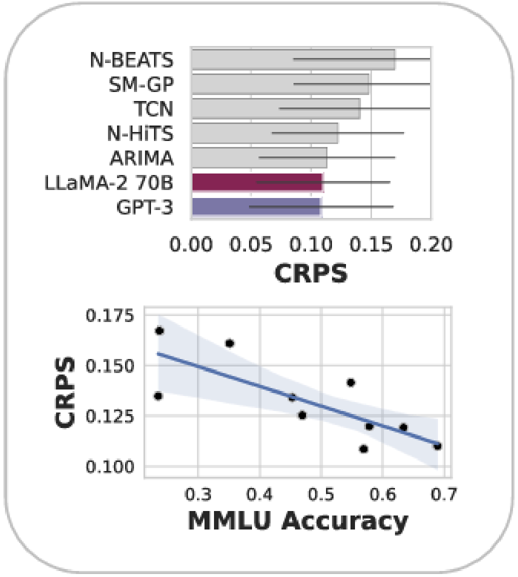
Performance of 4 mainstream LLMs on 7 different spatial tasks.

Can We Stimulate LLMs for Temporal Analysis?

- LLM can be adopted for solving time series forecasting by translating timeseries to nature language
  - GPT-3 and LLaMA-2 can zero-shot extrapolate time series comparable to or exceeding the performance of task-specific time series models.



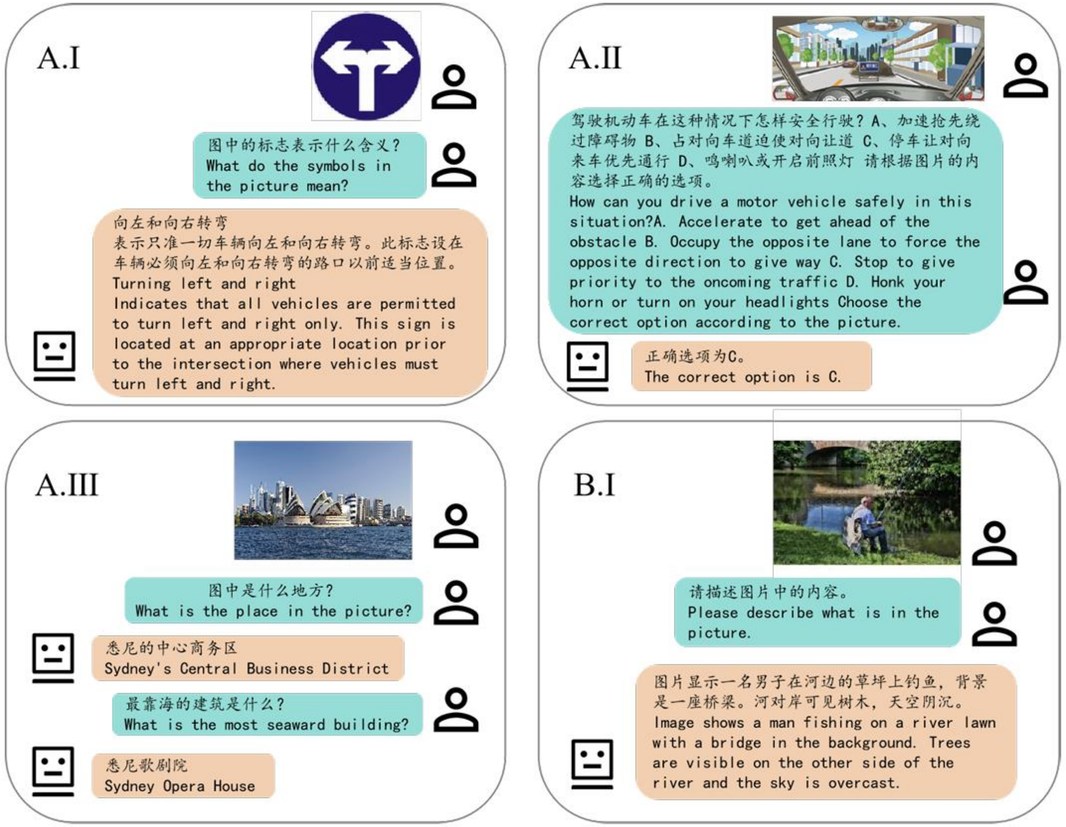




Nate Gruver, et al. Large Language Models Are Zero-Shot Time Series Forecasters. NeurIPS 2023.

Can We Teach LLMs for Solving Urban Tasks?

- LLM can be finetuned to accomplish transportation tasks
  - Knowledge injected: traffic engineering documents, research papers, examination documents, etc.
  - Supported Tasks: synthetic traffic scenarios, explaining traffic phenomena, answering traffic-related questions, providing traffic recommendations, and generating traffic reports.



A.I

图中的标志表示什么含义？
What do the symbols in the picture mean?

向左和向右转弯
表示只准一切车辆向左和向右转弯。此标志设在车辆必须向左和向右转弯的路口以前适当位置。
Turning left and right
Indicates that all vehicles are permitted to turn left and right only. This sign is located at an appropriate location prior to the intersection where vehicles must turn left and right.

A.II

驾驶机动车在这种情况下怎样安全行驶？A、加速抢先绕过障碍物 B、占对向车道迫使对向让道 C、停车让对向来车优先通行 D、鸣喇叭或开启前照灯 请根据图片的内容选择正确的选项。
How can you drive a motor vehicle safely in this situation?A. Accelerate to get ahead of the obstacle B. Occupy the opposite lane to force the opposite direction to give way C. Stop to give priority to the oncoming traffic D. Honk your horn or turn on your headlights Choose the correct option according to the picture.

正确选项为C。
The correct option is C.

A.III

图中是什么地方？
What is the place in the picture?

悉尼的中心商务区
Sydney's Central Business District

最靠海的建筑是什么？
What is the most seaward building?

悉尼歌剧院
Sydney Opera House

B.I

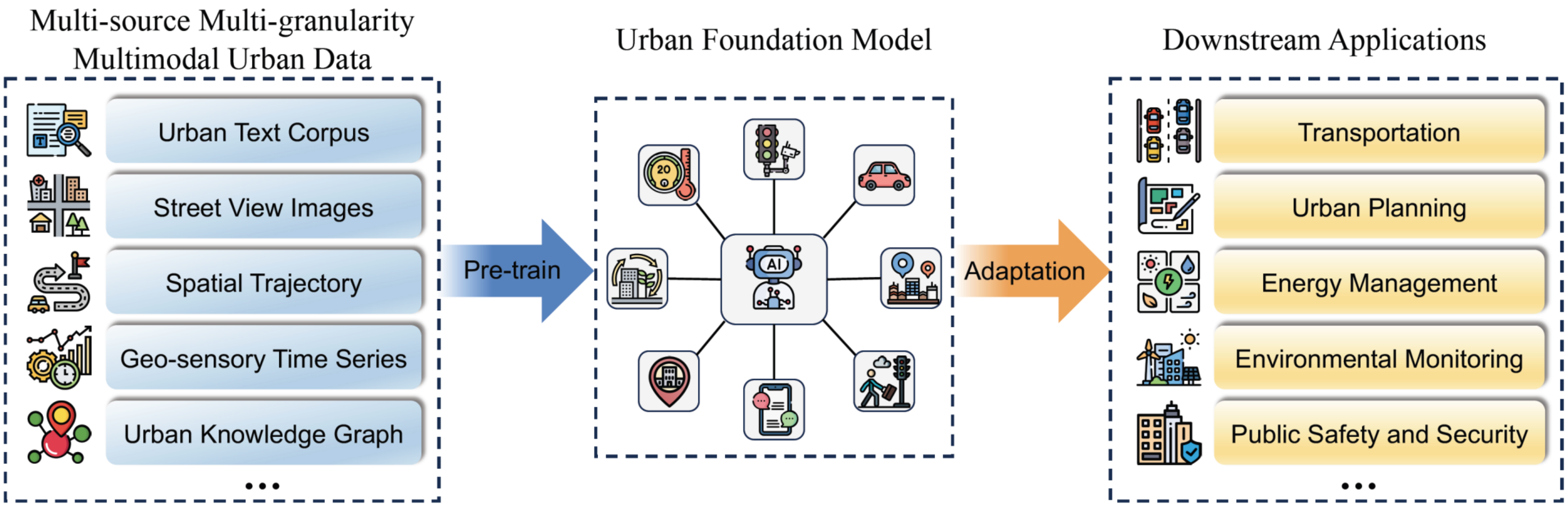
请描述图片中的内容。
Please describe what is in the picture.

图片显示一名男子在河边的草坪上钓鱼，背景是一座桥梁。河对岸可见树木，天空阴沉。
Image shows a man fishing on a river lawn with a bridge in the background. Trees are visible on the other side of the river and the sky is overcast.

LLM on solving traffic sign recognition, driving test, landmark recognition, captioning. LLM can be armed with multi-modal capability.

Peng Wang, et al. TransGPT: Multi-modal Generative Pre-trained Transformer for Transportation. Arxiv 2024.

Urban Foundation Models (UFMs): A Way to UGI



Multi-source Multi-granularity Multimodal Urban Data

- Urban Text Corpus
- Street View Images
- Spatial Trajectory
- Geo-sensory Time Series
- Urban Knowledge Graph
- ...

Pre-train

Urban Foundation Model

Adaptation

Downstream Applications

- Transportation
- Urban Planning
- Energy Management
- Environmental Monitoring
- Public Safety and Security
- ...

Urban Foundation Models (UFMs) are a family of large-scale models pre-trained on vast amounts of multi-source, multi-granularity, and multimodal urban data. They acquire notable general-purpose capabilities in the pre-training phase, exhibiting remarkable emergent abilities and adaptability dedicated to a range of urban application domains.

OUTLINE

- Introduction
- **Challenges of Building UFMs**
- Overview of UFMs
- Prospects of UFMs
- Summary

Challenges of Existing Foundation Models for UGI

Data Misalignment

- Multi-domain heterogeneous urban data distributed in different parties/stakeholders.
- A bunch of well-developed urban tools with rich urban knowledge encoded.

Skill Misalignment

- Tailored skill requirements to accomplish urban tasks, e.g., spatial & temporal calculation, multi-modal signal consolidation capabilities.
- Do way really need chat ability in all urban tasks?

Task Misalignment

- Misaligned objectives between pretrain/fine-tune of general purpose foundation model and urban tasks.
- Alignment of human preferences may limit the urban problem-solving capability.

Characteristics of Urban Big Data

Urban big data

Multi-source

Sensor networks,
Mobile devices,
Satellites,
…

Multi-granularity

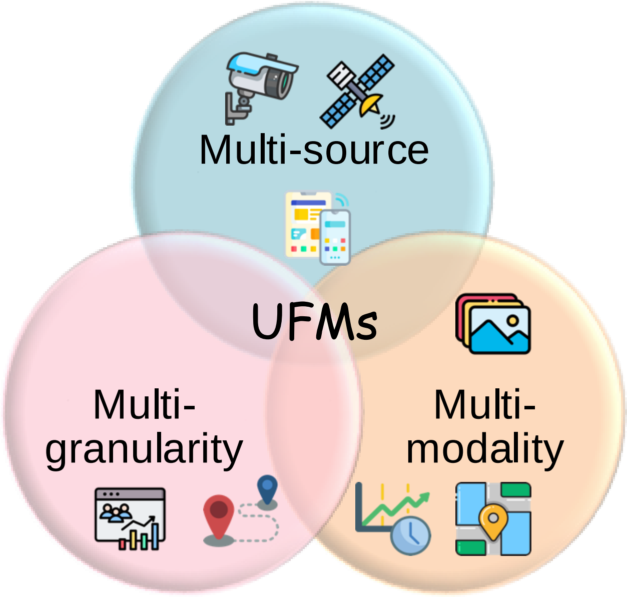
Macro-level:
Regional demand,
Population movement,
Traffic flow
Micro-level:
Trajectory,
Social media

Multi-modal

Text,
Image,
Trajectory,
Time series,
POI,
…

Challenges of Data Misalignment

- **Multi-source, multi-granularity, and multi-modal data integration**
  - Integrating diverse data sources at various granularities, from city-wide patterns to specific local details, poses a significant challenge due to the varied data modalities including text, images, and sensor readings.



Data heterogeneity

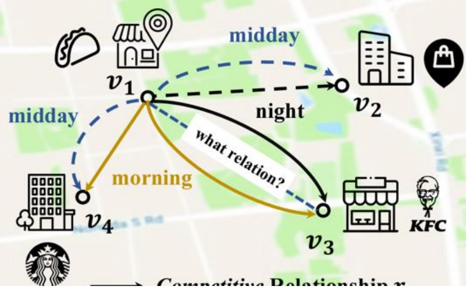
- **Privacy and security concerns**
  - Privacy and security are paramount, necessitating adherence to laws like GDPR, strong data anonymization, and defenses against data breaches and adversarial attacks to build trust and ensure ethical use.

Privacy protection

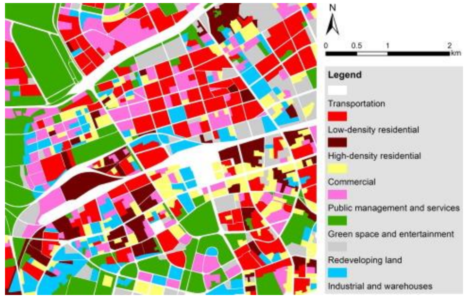
Challenges of Skill & Task Misalignment

Spatio-temporal reasoning skills

- Spatio-temporal reasoning aims to understand dynamic urban phenomena, which evolve over space and time through sophisticated modeling of temporal sequences and spatial distributions.



Spatial understanding



Spatial planning

Dynamic forecasting

Decision making

Versatility to diverse urban task domains

- UFMs need to adapt to diverse urban task domains like transportation, energy, and environmental monitoring, ensuring versatility and effective generalization across different urban scenarios.

Traffic

Environment

Energy

Climate