Towards Urban General Intelligence Through Urban Foundation Models

Hao Liu*
liuh@ust.hk
The Hong Kong University of Science and Technology (Guangzhou) & The Hong Kong University of Science and Technology
Guangzhou, China

Weijia Zhang
wzhang411@connect.hkust-gz.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Jindong Han
jhanao@connect.ust.hk
The Hong Kong University of Science and Technology
Hong Kong, China

Zhao Xu
zxu674@connect.hkust-gz.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Hang Ni
hni017@connect.hkust-gz.edu.cn
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China

Hui Xiong*
xionghui@ust.hk
The Hong Kong University of Science and Technology (Guangzhou) & The Hong Kong University of Science and Technology
Guangzhou, China

ABSTRACT

Machine learning techniques are now integral to the advancement of intelligent urban services, playing a crucial role in elevating the efficiency, sustainability, and livability of urban environments. The recent emergence of foundation models such as ChatGPT marks a revolutionary shift in the fields of machine learning and artificial intelligence. Their unparalleled capabilities in contextual understanding, problem solving, and adaptability across a wide range of tasks suggest that integrating these models into urban domains could have a transformative impact on the development of smart cities. Despite growing interest in Urban Foundation Models (UFMs), this burgeoning field faces challenges such as a lack of clear definitions, systematic reviews, and universalizable solutions.

In this tutorial, we will first introduce the concept of UFM and discuss the unique challenges involved in building them. We then present a data-centric taxonomy that categorizes and clarifies current UFM-related works, based on urban data modalities and types. Furthermore, to foster advancement in this field, we present a promising framework aimed at the prospective realization of UFMs, designed to overcome the identified challenges. Additionally, we explore the application landscape of UFMs, detailing their potential impact in various urban contexts. Relevant papers and open-source resources have been collated and are continuously updated at: https://github.com/usail-hkust/Awesome-Urban-Foundation-Models.

*Corresponding tutors.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Hao Liu, Weijia Zhang, Jindong Han, Zhao Xu, Hang Ni, and Hui Xiong. 2024. Towards Urban General Intelligence Through Urban Foundation Models. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 TARGET AUDIENCE AND PREREQUISITES

This tutorial welcomes conference attendees with interests in how machine learning and foundation models [1, 26, 27, 30] can transform urban environments. It's designed for those with a basic knowledge of data mining and machine learning. Yet, we aim to make the session as clear and straightforward as possible, ensuring it's easily graspable for both academic researchers and industry professionals.

2 TUTORIAL OUTLINE

Introduction. We introduce the basic concepts and definitions related to Urban Foundation Models (UFMs) and how they can pave the way to Urban General Intelligence (UGI).

Challenges of Building UFMs. This section discusses the challenges to building UFMs, including multi-source, multi-granularity, and multi-modal data integration; spatio-temporal reasoning capability; versatility to diverse urban task domains; and privacy and security concerns.

Overview of UFMs. We introduce a data-centric taxonomy for UFMs-related studies to shed light on the progress and efforts made in this field. Based on the urban data modalities, we categorize the existing works on UFMs into six classes:

- Language-based models [2, 7, 11, 16, 24, 42].
- Vision-based models [4, 9, 21, 23, 26, 32, 37].
- Trajectory-based models [8, 12, 17, 18, 20, 34].
- Time series-based models [5, 14, 15, 19, 39].
- Multimodal models [1, 33, 36, 38].
- Others [6, 13].

We will clarify these studies through the lens of their focused pretraining and adaptation techniques.

Prospects of UFMs. To foster advancement in this field, we present a promising framework aimed at the prospective realization of versatile UFMs, designed to overcome the identified challenges.

Applications. This section explores the application landscape of UFMs, detailing their potential impact in various urban domains, such as transportation [28, 35], urban planning [41], energy management [10], environmental monitoring [31], and public safety and security [40].

Summary and Future Directions. We summarize this tutorial and introduce potential future directions.

3 TUTORS BIOGRAPHY

3.1 Presenters

- **Hao Liu** is currently an assistant professor at the Artificial Intelligence Thrust, Hong Kong University of Science and Technology (Guangzhou). Prior to that, he was a senior research scientist at Baidu Research and a postdoctoral fellow at HKUST. He received the Ph.D. degree from the Hong Kong University of Science and Technology (HKUST), in 2017. His general research interests are in data mining, machine learning, and big data management, with a special focus on mobile analytics and urban computing. He has published prolifically in refereed journals and conference proceedings, such as TKDE, VLDBJ, KDD, NeurIPS, VLDB, SIGIR, WWW, AAAI, and IJCAI.
- **Weijia Zhang** is currently a Ph.D. student at the Artificial Intelligence Thrust, Hong Kong University of Science and Technology (Guangzhou). His research interests include spatio-temporal data mining, urban general intelligence, and sequence modeling. His works on urban intelligence have been published in several prestigious conferences and journals, such as KDD, WWW, VLDB, AAAI, ICDM, and TKDE.
- **Jindong Han** is currently a Ph.D. student at the Hong Kong University of Science and Technology. His research interests include spatiotemporal data mining, urban computing, and large language models. He has published several research papers in prestigious conferences and journals, such as KDD, VLDB, AAAI, TKDE, and VLDBJ. He received the first prize award in the Fresh Air competition of KDD Cup 2018.
- **Hui Xiong** is a Chair Professor, Associate Vice President (Knowledge Transfer), and Head of the Artificial Intelligence Thrust at Hong Kong University of Science and Technology (Guangzhou). His research interests span artificial intelligence, data mining, and mobile computing. He obtained his PhD in Computer Science from the University of Minnesota, USA. Dr. Xiong has served on numerous organization and program committees for conferences, including as Program Co-Chair for the Industrial and Government Track for the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Program Co-Chair for the IEEE 2013 International Conference on Data Mining (ICDM), General Co-Chair for the 2015 IEEE International Conference on Data Mining (ICDM), and Program Co-Chair of the Research Track for the 2018 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. He received several awards, such as the 2021 AAAI Best Paper Award and the 2011 IEEE ICDM Best Research Paper award. For his significant contributions to data mining and mobile computing, he was elected as a Fellow of both AAAS and IEEE in 2020.

3.2 Additional contributors

- **Zhao Xu** received his B.E. degree from Tsinghua University. He is currently an MPhil student at the Hong Kong University of Science and Technology (Guangzhou). His works have been published by influential conferences and journals such as IEEE TMC and ACM SIGSPATIAL, aiming at developing reliable, trustworthy large models for real-world applications.
- **Hang Ni** received his B.E. degree from Northwestern Polytechnical University. He is currently a Ph.D. candidate at Artificial Intelligence Thrust, Hong Kong University of Science and Technology (Guangzhou). His research interests lie in areas of graph learning and urban computing. Some of his works have been accepted by top conferences and journals such as CIKM and KBS.

4 AUDIENCE PARTICIPATION

To foster a dynamic and interactive tutorial environment, we plan to engage our audience through several key strategies. Firstly, we will contextualize the content by seamlessly integrating it with everyone's urban life experiences, ensuring that the material is both relatable and captivating for the audience. Secondly, we'll allocate dedicated Q&A sessions after each major topic, inviting questions and fostering a two-way dialogue to ensure clarity and engagement. Encouraging active participation, we aim to create an atmosphere where attendees feel comfortable to inquire, discuss, and explore ideas. Additionally, by making all presentation materials publicly available before the tutorial begins, we enable participants to prepare questions in advance, enhancing the overall interactive experience.

5 POTENTIAL SOCIETAL IMPACTS

This tutorial aims to bridge the gap between AI and urban science. Our goal is to equip AI scientists with insights into urban science, enabling them to refine AI methods for urban challenges [2, 22, 29]. Concurrently, urban scientists will discover the potential of advanced AI methods, like foundation models [1, 26, 27, 30], to tackle urban issues, fostering the development of impactful applications [3, 25, 26]. This cross-disciplinary exchange is designed to spur collaboration between the two fields, enhancing the pursuit of smart city solutions and urban general intelligence. Ultimately, we anticipate inspiring more researchers to explore UFMs, accelerating progress toward smarter and more sustainable urban environments.

REFERENCES

[1] Pasquale Balsebre, Weiming Huang, Gao Cong, and Yi Li. 2023. City Foundation Models for Learning General Purpose Representations from OpenStreetMap. *arXiv e-prints* (2023), arXiv–2310.

[2] Prabin Bhandari, Antonios Anastasopoulos, and Dieter Pfoser. 2023. Are Large Language Models Geospatially Knowledgeable?. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*. 1–4.

[3] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619, 7970 (2023), 533–538.
[4] Keyan Chen, Chenyang Liu, Hao Chen, Haotian Zhang, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. 2023. RSPrompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model. *arXiv preprint arXiv:2306.16269* (2023).
[5] Yakun Chen, Xianzhi Wang, and Guandong Xu. 2023. GATGPT: A Pre-trained Large Language Model with Graph Attention Network for Spatiotemporal Imputation. *arXiv preprint arXiv:2311.14332* (2023).
[6] Longchao Da, Minchiuan Gao, Hao Mei, and Hua Wei. 2023. Llm powered sim-to-real transfer for traffic signal control. *arXiv preprint arXiv:2308.14284* (2023).
[7] Ruixue Ding, Boli Chen, Pengjun Xie, Fei Huang, Xin Li, Qiang Zhang, and Yao Xu. 2023. MGeo: Multi-Modal Geographic Language Model Pre-Training. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 185–194.
[8] Vinayak Gupta and Srikanta Bedathur. 2022. Doing more with less: overcoming data scarcity for poi recommendation via cross-region transfer. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 3 (2022), 1–24.
[9] Lukas Haas, Silas Alberti, and Michal Skreta. 2023. Learning Generalized Zero-Shot Learners for Open-Domain Image Geolocalization. *arXiv preprint arXiv:2302.00275* (2023).
[10] Chenghao Huang, Siyang Li, Ruohong Liu, Hao Wang, and Yize Chen. 2023. Large Foundation Models for Power Systems. *arXiv preprint arXiv:2312.07044* (2023).
[11] Jizhou Huang, Haifeng Wang, Yibo Sun, Yunsheng Shi, Zhengjie Huang, An Zhuo, and Shikun Feng. 2022. ERNIE-GeoL: A Geography-and-Language Pre-trained Model and its Applications in Baidu Maps. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3029–3039.
[12] Jiawei Jiang, Dayan Pan, Houxing Ren, Xiaohan Jiang, Chao Li, and Jingyuan Wang. 2023. Self-supervised Trajectory Representation Learning with Temporal Regularities and Travel Semantics. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*. IEEE, 843–855. https://doi.org/10.1109/ICDE55515.2023.00070
[13] Siqi Lai, Zhao Xu, Weijia Zhang, Hao Liu, and Hui Xiong. 2023. Large Language Models as Traffic Signal Control Agents: Capacity and Opportunity. *arXiv preprint arXiv:2312.16044* (2023).
[14] Zhonghang Li, Lianghao Xia, Jiabin Tang, Yong Xu, Lei Shi, Long Xia, Dawei Yin, and Chao Huang. 2024. UrbanGPT: Spatio-Temporal Large Language Models. *arXiv preprint arXiv:2403.00813* (2024).
[15] Zhonghang Li, Lianghao Xia, Yong Xu, and Chao Huang. 2023. GPT-ST: Generative Pre-Training of Spatio-Temporal Graph Neural Networks. *arXiv preprint arXiv:2311.04245* (2023).
[16] Zekun Li, Wenxuan Zhou, Yao-Yi Chiang, and Muhao Chen. 2023. GeoLM: Empowering Language Models for Geospatially Grounded Language Understanding. *arXiv preprint arXiv:2310.14478* (2023).
[17] Yuebing Liang, Yichao Liu, Xiaohan Wang, and Zhan Zhao. 2023. Exploring Large Language Models for Human Mobility Prediction under Public Events. *CoRR* abs/2311.17351 (2023). https://doi.org/10.48550/ARXIV.2311.17351 arXiv:2311.17351
[18] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. 2021. Pre-training Context and Time Aware Location Embeddings from Spatial-Temporal Trajectories for User Next Location Prediction. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*. AAAI Press, 4241–4248. https://doi.org/10.1609/AAAI.V35I5.16548
[19] Chenxi Liu, Sun Yang, Qianxiong Xu, Zhishuai Li, Cheng Long, Ziyue Li, and Rui Zhao. 2024. Spatial-temporal large language model for traffic prediction. *arXiv preprint arXiv:2401.10134* (2024).
[20] Hao Liu, Jindong Han, Yanjie Fu, Yanyan Li, Kai Chen, and Hui Xiong. 2023. Unified route representation learning for multi-modal transportation recommendation with spatiotemporal pre-training. *VLDB J.* 32, 2 (2023), 325–342. https://doi.org/10.1007/S00778-022-00748-Y
[21] Yu Liu, Xin Zhang, Jingtao Ding, Yanxin Xi, and Yong Li. 2023. Knowledge-infused contrastive learning for urban imagery-based socioeconomic prediction. In *Proceedings of the ACM Web Conference 2023*. 4150–4160.
[22] Gengchen Mai, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, et al. 2023. On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv preprint arXiv:2304.06798* (2023).
[23] Xin Man, Chenghong Zhang, Changyu Li, and Jie Shao. 2023. W-MAE: Pre-trained weather model with masked autoencoder for multi-variable weather forecasting. *arXiv preprint arXiv:2304.08754* (2023).
[24] Rohin Manvi, Samar Khanna, Gengchen Mai, Marshall Burke, David Lobell, and Stefano Ermon. 2023. Geollm: Extracting geospatial knowledge from large language models. *arXiv preprint arXiv:2310.06213* (2023).
[25] Grey Nearing, Deborah Cohen, Vusumuzi Dube, Martin Gauch, Oren Gilon, Shaun Harrigan, Avinatan Hassidim, Daniel Klotz, Frederik Kratzert, Asher Metzger, Sella Nevo, Florian Pappenberger, Christel Prudhomme, Guy Shalev, Shlomo Shenzis, Tadele Yednkachw Tekalign, Dana Weitzner, and Yossi Matias. 2024. Global prediction of extreme floods in ungauged watersheds. *Nature* 627, 8004 (2024), 559–563.
[26] Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. 2023. ClimaX: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343* (2023).
[27] OpenAI et al. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
[28] Wang Peng. 2023. DUOMO/TransGPT. https://github.com/DUOMO/TransGPT
[29] Yifeng Shi, Feng Lv, Xinliang Wang, Chunlong Xia, Shaojie Li, Shujie Yang, Teng Xi, and Gang Zhang. 2023. Open-TransMind: A New Baseline and Benchmark for 1st Foundation Model Challenge of Intelligent Transportation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6327–6334.
[30] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).
[31] Saeid Ashraf Vaghefi, Dominik Stammbach, Veruska Muccione, Julia Bingler, Jingwei Ni, Mathias Kraus, Simon Allen, Chiara Colesanti-Senni, Tobias Wekhof, Tobias Schimanski, et al. 2023. Chatclimate: Grounding conversational AI in climate science. *Communications Earth & Environment* 4, 1 (2023), 480.
[32] Di Wang, Qiming Zhang, Yufei Xu, Jing Zhang, Bo Du, Dacheng Tao, and Liangpei Zhang. 2022. Advancing plain vision transformer toward remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing* 61 (2022), 1–15.
[33] Peng Wang, Xiang Wei, Fangxu Hu, and Wenjuan Han. 2024. TransGPT: Multi-modal Generative Pre-trained Transformer for Transportation. *arXiv preprint arXiv:2402.07233* (2024).
[34] Xinglei Wang, Meng Fang, Zichao Zeng, and Tao Cheng. 2023. Where Would I Go Next? Large Language Models as Human Mobility Predictors. *CoRR* abs/2308.15197 (2023). https://doi.org/10.48550/ARXIV.2308.15197 arXiv:2308.15197
[35] Xuhong Wang, Ding Wang, Liang Chen, and Yilun Lin. 2023. Building Transportation Foundation Model via Generative Graph Transformer. *arXiv preprint arXiv:2305.14826* (2023).
[36] Yibo Yan, Haomin Wen, Siru Zhong, Wei Chen, Haodong Chen, Qingsong Wen, Roger Zimmermann, and Yuxuan Liang. 2023. When Urban Region Profiling Meets Large Language Models. *arXiv preprint arXiv:2310.18340* (2023).
[37] Fanglong Yao, Wanxuan Lu, Heming Yang, Liangyu Xu, Chenglong Liu, Leiyi Hu, Hongfeng Yu, Nayu Liu, Chubo Deng, Deke Tang, et al. 2023. RingMo-sense: Remote sensing foundation model for spatiotemporal prediction via spatiotemporal evolution disentangling. *IEEE Transactions on Geoscience and Remote Sensing* (2023).
[38] Siyao Zhang, Daocheng Fu, Zhao Zhang, Bin Yu, and Pinlong Cai. 2023. Trafficgpt: Viewing, processing and interacting with traffic foundation models. *arXiv preprint arXiv:2309.06719* (2023).
[39] Zijian Zhang, Xiangyu Zhao, Qidong Liu, Chunxu Zhang, Qian Ma, Wanyu Wang, Hongwei Zhao, Yiqi Wang, and Zitao Liu. 2023. PromptST: Prompt-Enhanced Spatio-Temporal Multi-Attribute Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 3195–3205.
[40] Ou Zheng, Mohamed Abdel-Aty, Dongdong Wang, Chenzhu Wang, and Shengxuan Ding. 2023. TrafficSafetyGPT: Tuning a Pre-trained Large Language Model to a Domain-Specific Expert in Transportation Safety. *arXiv preprint arXiv:2307.15311* (2023).
[41] Zhilun Zhou, Yuming Lin, Depeng Jin, and Yong Li. 2024. Large Language Model for Participatory Urban Planning. *arXiv preprint arXiv:2402.17161* (2024).
[42] He Zhu, Wenjia Zhang, Nuoxian Huang, Boyang Li, Luyao Niu, Zipei Fan, Tianle Lun, Yicheng Tao, Junyou Su, Zhaoya Gong, Chenyu Fang, and Xing Liu. 2024. PlanGPT: Enhancing Urban Planning with Tailored Language Model and Efficient Retrieval. *arXiv preprint arXiv:2402.19273* (2024).