香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

数据科学与分析学域
DATA SCIENCE AND ANALYTICS THRUST
信息枢纽
INFORMATION HUB

DSA Research Experiences for Undergraduates

Research Project

Section1: Faculty Information

Full Name	Guoming Tang	Tel	13618480331
Thrust/Hub	DSA/INFO	Office	W3-306
Email	guomingtang@hkust-gz.edu.cn		

Section2: Research Project Proposal

Project Title	Introducing LLMs to Building Energy Modeling
Project Description (max 800 words)	As global urbanization accelerates, the high energy consumption of official buildings underscores the urgent need for advanced modeling methods. Traditional building energy modeling (BEM) approaches face major challenges: **forward modeling**, based on physical principles, requires extensive input and expertise, while **inverse modeling**, which uses historical data, struggles with limited datasets and lacks interpretability. Integrating large language models (LLMs) into BEM presents opportunities to overcome these limitations. LLMs provide distinct advantages over traditional methods, including **explainability** through natural language insights, **generalization** by leveraging domain knowledge, and a seamless transition **from prediction to actionable recommendations**. These features make LLMs ideal for advancing energy efficiency and sustainability in buildings. **This project will explore the following questions:** 1. How can time-series sensor data (e.g., temperature, humidity, HVAC operations) be effectively preprocessed and represented for LLM-based modeling? [1] 2. Can LLMs outperform traditional neural networks in energy consumption prediction by leveraging reasoning capabilities and domain knowledge (using fine-tuning methods like LoRA)? [2, 3] 3. How can LLMs provide interpretable explanations and actionable optimization suggestions alongside numerical predictions? 4. How can LLMs' expert knowledge lower the barrier to using complex forward modeling methods (e.g., developing LLM agents for BEM)? [4] **You will gain the following opportunities:** 1. **Real-world Data**: Access real data from hundreds of sensors in large buildings across multiple locations, process and apply them for BEM. 2. **LLM Agent Development**: Gain hands-on experience in developing LLM agents, exploring techniques like RAG, prompt engineering, and fine-tuning, and experiment with their applications.

	3. **Research Skills & Papers**: Identify real-world problems, transform them into research topics, and produce high-quality academic papers (if possible). 4. **Internship Opportunities**: Students participating in this project may have the opportunity to intern at relevant companies. [1] Li Z, Deldari S, Chen L, et al. Sensorllm: Aligning large language models with motion sensors for human activity recognition[J]. arXiv preprint arXiv:2410.10624, 2024. [2] Gruver N, Finzi M, Qiu S, et al. Large language models are zero-shot time series forecasters[J]. Advances in Neural Information Processing Systems, 2023, 36: 19622-19635. [3] Jin M, Wang S, Ma L, et al. Time-LLM: Time series forecasting by reprogramming large language models[C]. International Conference on Learning Representations (ICLR), 2024. [4] Zhang L, Chen Z, Ford V. Advancing building energy modeling with large language models: Exploration and case studies[J]. Energy and Buildings, 2024, 323: 114788.
Proposed Research Duration	Start Date: Now End Date: Aug. 2025
Student/Researcher Duties	Students should read papers in related fields, actively engage in team discussions, and contribute ideas to the project. Additionally, they will be responsible for part of coding tasks. Exploring different cutting-edge directions to solve problems is encouraged.
Technical Skills Required	☑Python ☐ Machine Learning ☐ Big Data ☐ R ☑ Deep Learning ☐ SQL ☐ C/C++ ☐ Other: _____
Preferred Student/Researcher Background	Programming, Academic Reading & Writing Skills, Teamwork Ability
Maximum Number of Students/Researchers	☐ 1 ☑ 2

Section3: Pre-Application Research Exposure Meeting

Faculty members are encouraged to schedule a Research Exposure Meeting to introduce students to their projects.

Preferred Date	Mar. 14-16, 2025
Preferred Time	Morning or afternoon time, 1 hour
Meeting Mode	☑ In-Person ☐ Online
Venue (if in-person)	W3-306
Meeting Link (if online)	