A Practical Introduction to LLMs in Biomedical Data Science Research

Robert Tang, Qiao Jin, Hufeng Zhou, Shubo Tian, Zhiyong Lu, Mark Gerstein

https://llm4biomed.github.io/

1

Organizing team

Robert Xiangru Tang
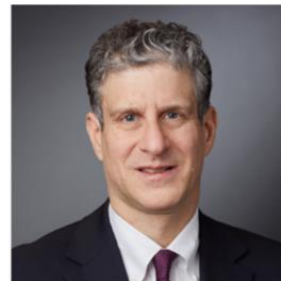Yale University

Qiao Jin
NCBI/NLM/NIH

Hufeng Zhou
Harvard University

Shubo Tian
NCBI/NLM/NIH

Zhiyong Lu
NCBI/NLM/NIH

Mark Gerstein
Yale University

Learning Objectives

- Familiarizing with the key aspects of large-scale biomedical data.
- Leveraging LLMs to handle and interpret vast amounts of biomedical data.
- Learning cutting-edge research topics from two invited talks.
- Utilizing OpenAI APIs for GPTs and open-source LLMs in Python.
- Integrating LLMs to enhance their coding efficiency in bioinformatics.
- Deploying LLMs for biomedical question-answering and academic literature exploration.

Tutorial Agenda: Part 1 - Monday, July 8, 2024

Time	Section	Presenter
Part 1 (Monday, July 8, 2024)		
14:00 - 14:10	Overview and Welcome	Robert Tang
14:10 - 14:40	Introduction to LLMs with a Focus on Biomedical Data Science	Shubo Tian
14:40 - 15:10	How to Use GPT-3.5 and GPT-4 with Python	Qiao Jin
15:10 - 15:30	How to Use Open-source LLMs with Python	Robert Tang
15:30 - 15:45	Coffee Break	
15:45 - 16:10	Code Generation in Bioinformatics	Robert Tang
16:10 - 16:35	Retrieval-Augmented Generation with Large Language Models	Qiao Jin
16:35 - 17:00	Querying PubMed with RAG to Answer Biomedical Questions with GPT-4	Qiao Jin

Tutorial Agenda: Part 1 - Tuesday, July 9, 2024

Part 2 (Tuesday, July 9, 2024)		
14:00 - 14:45	Large Language Models for Biomedicine: from PubMed Search to Gene Set Analysis	Zhiyong Lu
14:45 - 15:30	AI in Biomedicine: Developing Representations of Disease-Relevant Molecules	Mark Gerstein
15:30 - 15:45	Coffee Break	
15:45 - 16:10	Integrating Biomedical Data Database Development with LLMs	Hufeng Zhou
16:10 - 16:35	Database Query Generation with LLMs	Hufeng Zhou
16:35	Closing Remarks	Robert Tang