



MACHINE LEARNING FOR ASTROPHYSICS

2ND EDITION CATANIA, 8-12 JULY, 2024

Contribution ID: 229

Type: **Oral Presentation**

[INVITED] Expediting Astronomical Discovery with Large Language Models: Progress, Challenges, and Future Directions

Thursday, 11 July 2024 09:00 (40 minutes)

The vast and interdisciplinary nature of astronomy, coupled with its open-access ethos, makes it an ideal testbed for exploring the potential of Large Language Models (LLMs) in automating and accelerating scientific discovery. In this talk, we present our recent progress in applying LLMs to tackle real-life astronomy problems. We demonstrate the ability of LLM agents to perform end-to-end research tasks, from data fitting and analysis to iterative strategy improvement and outlier detection, mimicking human intuition and deep literature understanding. However, the cost-effectiveness of closed-source solutions remains a challenge for large-scale applications involving billions of sources. To address this issue, we introduce our ongoing work at AstroMLab on training lightweight, open-source specialized models and our effort to benchmark these models with carefully curated astronomy benchmark datasets. We will also discuss our effort to construct the first LLM-based knowledge graph in astronomy, leveraging citation-reference relations. The open-source specialized LLMs and knowledge graph are expected to guide more efficient strategy searches in autonomous research pipelines. While many challenges lie ahead, we explore the immense potential of scaling up automated inference in astronomy, revolutionizing the way astronomical research is conducted, ultimately accelerating scientific breakthroughs and deepening our understanding of the Universe.

Presenter: SEN-TING, Yuan

Session Classification: Cosmology & Simulations