

# Knowledge Discovery Interest Group

Chair: **Raffaele D'Abrusco**<sup>1</sup>, Vice-chair: **Yihan Tao**<sup>2</sup>

<sup>1</sup> Center for Astrophysics | Harvard & Smithsonian

<sup>2</sup> National Astronomical Observatories, Chinese Academy of Science

*"Knowledge Discovery is the task of processing and analyzing astronomical datasets with the aim of **extracting new knowledge**. This endeavor spans multiple disciplines including *visualization, data access and exploration, machine learning, statistical methods and workflow orchestration.*"\**

*"Knowledge Discovery is the task of processing and analyzing astronomical datasets with the aim of **extracting new knowledge**. This endeavor spans multiple disciplines including *visualization, data access and exploration, machine learning, statistical methods and workflow orchestration.*"\**

\* Facilitating the deployment and application of KD methods **where astronomical data can be found** is a crucial task

# Goals

- ML-proofing existing and future science platforms
  - Are existing astronomy science platforms compatible with ML methods?
  - Investigate whether science platforms can access tabular and non-tabular data through VO interfaces.
  - Building libraries of well-established pre-trained models and integrating them in science platforms.
  
- Collecting resources for the "ready to ML in astronomy" kit
  - Library of datasets: collecting data for standard DM application for testing and benchmarking of KD models ("Iris datasets" for astronomy)
  - Looking at different "astronomy data challenges": can they represent the starting point for the library of astronomical datasets?
  - Collecting and describing methods for standardization/normalization of data used in astronomy, with reference data to test different implementations of the methods

## Artificial Intelligence-powered chatbots

- ◇ The basics of the ML methods at the core of these tools
- ◇ Practical applications to astronomical research
- ◇ Commercial (gargantuan, general-purpose, available now) vs open (agile, curated, in the making) tools
- ◇ Extension to non-language outputs

*“We don’t know exactly where we are going, but we are going somewhere”*

# KD-IG @2023 Spring InterOperability Meeting

## KD-IG session: Tuesday 05/09, 11:00 AM (CEST)

- Sandor Kruk: *Exploring astronomy data archives at large scales using deep learning and crowdsourcing*
- Mini-session on generative language models and AI-powered tools
  - Y. Tao: *Foundation models for Astronomy*
  - A. Schaaff: *AI in querying astronomical data services*
  - R. Galarza-Martinez: *Intro to Transformers*
  - Ioana Ciucă: *Galactic ChitChat: Using Large Language Models to Engage with Astronomy Literature*
  - Adrian Damian: *Discover IVOA with ChatGPT*
  - Discussion about generative language models

## Other (very) relevant sessions

- Science Platforms sessions I and II
- Grid and Web Services (GWS) Sessions I and II