Knowledge Discovery Interest Group summary

Themes of the KDIG

The emergence of science platforms

→ ◇ Multiple astronomical science platforms have become or will be shortly available to the community

→ ◇ Some have use cases including ML tasks, but most primarily cater to standard processing/analysis of data

→ ◇ Exciting technologies (stemmed from IVOA standards) that could make access to catalog data in ML applications more effective

Artificial Intelligence-powered chatbots

→ ◇ The models used by LLMs **are already used in astronomy**

→ ◇ Practical applications to astronomical research **of current commercial tools require careful training and validation of results**

→ ◇ Commercial (gargantuan, general-purpose, available now) vs open (agile, curated, in the making) tools, **or learn how to feed data to commercial tools**

KDIG session

- Sandor Kruk: *Exploring astronomy data archives at large scales using deep learning and crowdsourcing*

  ML methods and crowdsourcing helped discover asteroids in archival HST observations, opening new options at the onset of massive astronomical datasets
  - R. Martinez-Galarza: *Intro to Transformers*

    Introductions to the Transformers, the model at the basis of the LLMs: the importance of "attention" vs persistence, and applications to linear astronomical data.
  - Y. Tao: *Foundation models for Astronomy*

    Bringing chatGPT into context: foundation models are trained on large training sets with high capability to response to broad queries. Examples in astronomical research.
  - A. Schaaff: *AI in querying astronomical data services*

    The CDS chatbot as an example of AI-powered chatbot, possible integration with Rasa conversational AI agent comes with financial and policy concerns.
  - Ioana Ciucă: *Galactic ChitChat: Using Large Language Models to Engage with Astronomy Literature*

    Talking with GPT-4 when it's given sound, accurate context produces surprisingly good responses in a specific field!
  - Adrian Damian: *Discover IVOA with ChatGPT*

    Even if ChatGPT-4 progresses, answers can still be really wrong: disruptive value for documentation with additional training needed

Goals

☐ Seize the moment
- Take advantage of the built-in flexibility of IGs to pursue potentially interesting topics
- Act as liaison between the VO community, astronomical organizations/missions and the world at large
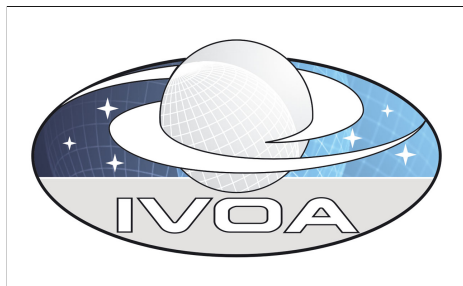
☐ Ride the momentum
- Lobbying activity within the IVOA
- Enhance communication (Slack channel #kdd)
- Regular, focused, "in-between InterOps" meetings

☐ Documents are what makes the (VO) world go around
- Draft one or more *notes* on distinct topics
- Aim for an *endorsed note*

Contacts

- Chair: Raffaele D'Abrusco
- Vice-chair: Yihan Tao

Staying in touch
→ ■ E-mail: *kdd@ivoa.net*
→ ■ Slack: *IVOA#kdd*
→ ■ Webpage: https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaKDD