

Harnessing Scientific AI for Knowledge Discovery in the Open Research Knowledge Graph

Course at: 2nd European Summer School on Artificial Intelligence (ESSAI 2024), Athens, Greece

Date of event: 22nd to 26th July 2024

Presented by: Jennifer D'Souza (Junior AI Research Group Lead) and Allard Oelen (Postdoc, ORKG Group)

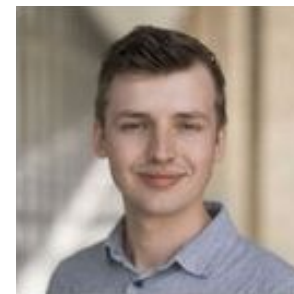
Contact: <https://www.linkedin.com/in/jennifer-l-dsouza/>, <https://www.linkedin.com/in/allard-oelen/>

Plan for the course

- **Topic 1:** Harnessing Scientific AI for Knowledge Discovery in Virology (Monday, July 22nd)
- **Topic 2:** Harnessing Large Datasets for Large Language Models (Tuesday, July 23rd)
- **Topic 3:** Harnessing Large Language Models as Research Assistants (Wednesday, July 24th)
- **Topic 4:** The Open Research Knowledge Graph (Thursday, July 25th)
- **Topic 5:** ORKG Ask – An AI-powered Next Generation Scientific Search System (Friday, July 26th)



Jennifer D'Souza, PhD
AI Research Group Lead



Allard Oelen, PhD
Researcher and ORKG Frontend Dev Lead

Plan for the course



- **Topic 1: Harnessing Scientific AI for Knowledge Discovery in Virology (Monday, July 22nd)**
 - **What?** Broad overview of various scientific AI objectives w.r.t. the Virology domain
 - **Why?** Much of the recent emphasis has been on the uptake of generative AI technologies, i.e. LLMs. This talk aims to look more broadly at scientific AI and the various objectives it entails.
- **Topic 2: Harnessing Large Datasets for Large Language Models (Tuesday, July 23rd)**
 - **What?** A look under the hood at the large-scale datasets that power the generative AI abilities of LLMs
 - **Why?** The final talk of this series presents a scientific search system powered by LLMs. This talk aims to set the stage to better comprehend how LLMs do what they do given the data they have seen. The takeaway insight expected is: if generic LLMs need to be tailored for scientific purposes how could their underlying data be potentially enhanced such that they are better suited to comprehend scientific objectives.
- **Topic 3: Harnessing Large Language Models as Research Assistants (Wednesday, July 24th)**
 - **What?** A tool to leverage ChatGPT for various research tasks including to build an Open Research Knowledge Graph
 - **Why?** Researchers are already using generative AI to support research tasks. This talk specifically visits various objectives by which LLMs can assist scientists including building an ORKG and even across 3 domains in Science.

Plan for the course



- Topic 4: The Open Research Knowledge Graph (Thursday, July 25th)
 - **What?** A scholarly knowledge graph that uses Human-AI collaboration to create and curate data
 - **Why?** Scholarly articles are mainly published in a document-based form, which makes the knowledge presented within those articles difficult to access for machines. Structured representations of the knowledge make it possible to make scholarly knowledge more findable, accessible, interoperable and reusable (FAIR).
- Topic 5: ORKG Ask – An AI-powered Next Generation Scientific Search System (Friday, July 26th)
 - **What?** ORKG Ask is a scholarly search and exploration system and serves as the next generation scholarly information retrieval systems.
 - **Why?** Traditional scholarly search systems mainly use keywords-based retrieval, ranking documents based on their relevance to the keywords. ORKG Ask is using vector search and LLM extraction to find more relevant documents and extract key information automatically from paper abstracts and full-texts.

Let's get started!

Topic 1: Harnessing Scientific AI for Knowledge Discovery in Virology (Monday, July 22nd)

