Mat•AI: Team building through scientific communication

Datalab: Materials data management for humans and machines

Joshua D. Bocarsly
Department of Chemistry & TcSUH, University of Houston

Thursday, November 7th 4-5 pm, Houston Science Center Room 102

Abstract:
In the field of materials science, researchers use a wide variety of synthesis and characterization techniques to develop materials that can address society's grand challenges. As the quantity and diversity of our data grows, there is an increasing need for tools that can help us to keep track of this valuable information and make the best use of it for scientific discovery. Storing the various raw data files is generally not enough: experiments must be linked to their metadata and context in order to be useful. With these challenges in mind, we have been developing an open-source, web-based data management platform, *datalab* (github.com/datalab-org/datalab). This platform strives to make it as easy as possible for researchers to store all of their experimental data with full metadata and dynamic links between starting materials, samples, and devices. It also provides various tools to quickly visualize results and perform simple analysis. For example, a battery researcher can view the charge/discharge cycling performance of a certain battery cell they have made in the lab, and link that performance to all of the characterization that has been performed on the electrodes and electrolytes that were used to fabricate that cell.

This type of contextualized digital data management affords many advantages: researchers can quickly find and visualize their own data for analysis, sharing, and collaboration. At the same, the data is accessible in machine-friendly formats and can be accessed using a python API for automated analysis and use with machine learning. The recent rise of powerful multimodal large language models (MLLMs) (e.g. *claude 3.5 Sonnet* and *gpt-4o*) is particularly promising in this regard, allowing us to directly act on the diverse, denormalized collections of text and images that are collected in real working materials labs. This talk will discuss our development of *datalab* and associated AI agents, and give a perspective on future opportunities for the use of AI agents in accelerating research data management and analysis tasks.

Bio: Dr. Joshua Bocarsly is an Assistant Professor in Chemistry at the University of Houston and a Robert A. Welch Foundation Professor at the Texas Center for Superconductivity (TcSUH). His research focuses on discovering, characterizing, and controlling advanced inorganic materials for applications in batteries, computing, and energy conversion. He is also engaged in advancing digital laboratory data management to enable greater reproducibility and sharing of scientific data.

After completing his undergraduate in Chemistry from Princeton University in 2015, Dr. Bocarsly received his Ph.D. in Materials in 2020 from UC Santa Barbara. In Santa Barbara, he studied magnetocaloric materials and skyrmion host materials under the joint mentorship of Profs. Ram Seshadri and Stephen Wilson. He then moved to the University of Cambridge to complete a postdoctoral fellowship studying next-generation battery cathodes in the group of Prof. Clare Grey. He joined the UH faculty in Fall 2023.

Please RSVP @ https://UH-MatAI.eventbrite.com for food and drink