

Integrating Programming and Cheminformatics into the Molecular Sciences Curriculum: Resources from the Molecular Sciences Software Institute

using  nanoHUB

Ashley Ringer McDonald

California Polytechnic State University
Department of Chemistry and Biochemistry
San Luis Obispo, CA, USA



education.molssi.org



nanohub.org/education



CAL POLY

The Molecular Sciences Software Institute (the MolSSI)

- National Science Foundation Center (launched 2016, renewed 2021)
- Collaborative effort by nine institutions in the United States
- Goals: Improve software, education, and training computational molecular science – a broad domain that includes quantum chemistry, computational materials science, and biomolecular simulation.

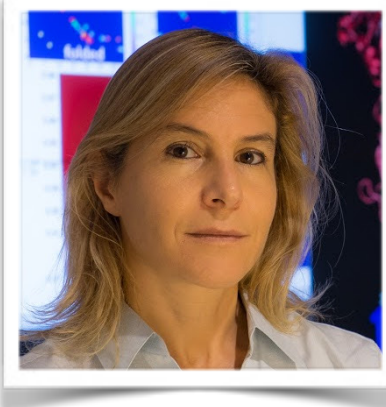


education.molssi.org



CAL POLY

MolSSI Board of Directors



Cecilia Clementi, Rice U., Co-Director for Biomolecular Simulation and International Engagement

T. Daniel Crawford, Virginia Tech, Director



Robert J. Harrison, Stony Brook U., Co-Director for Parallel Computing and Emerging Technologies

Teresa Head-Gordon, U.C. Berkeley, Co-Director for Laboratory, Industrial, and Academic Outreach and Education



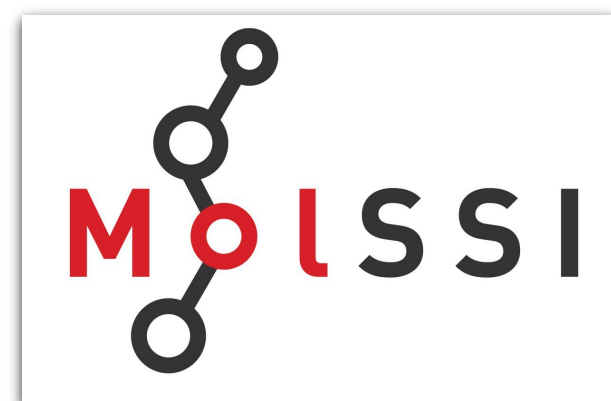
Shantenu Jha, Rutgers U., Co-Director for Software Engineering Process, Middleware, and Infrastructure

Anna Krylov, U. Southern California, Co-Director for Quantum Chemistry



Theresa Windus, Iowa State U., Deputy Director and Co-Director for Code and Data Interoperability

Dominika Zgid, U. Michigan, Co-Director for Materials Science

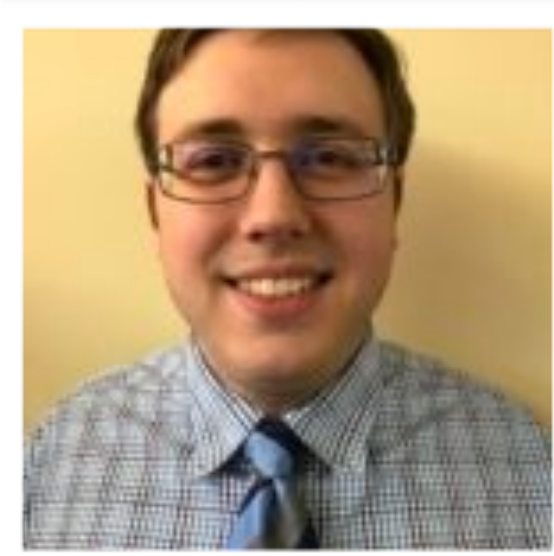


Ashley McDonald, Cal Poly SLO., Co-Director for Education, Training, and Faculty Development

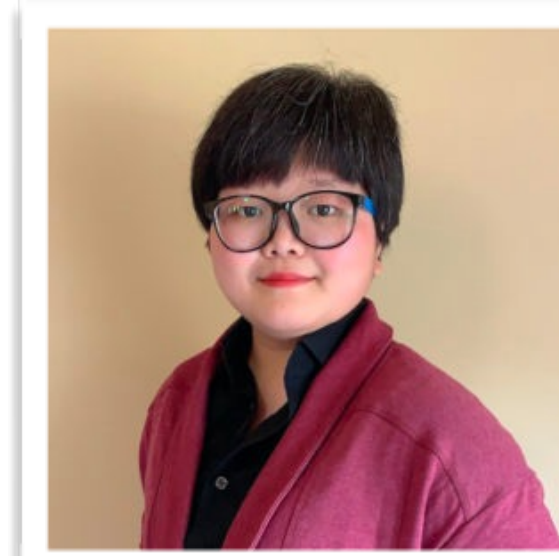
MolSSI Software Scientists



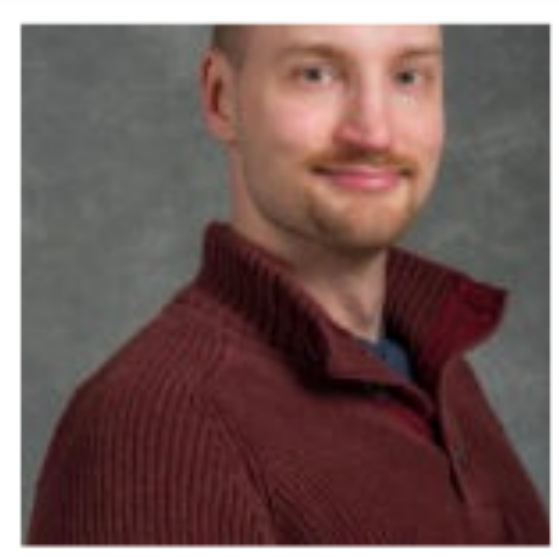
Paul Saxe
Lead Software Scientist



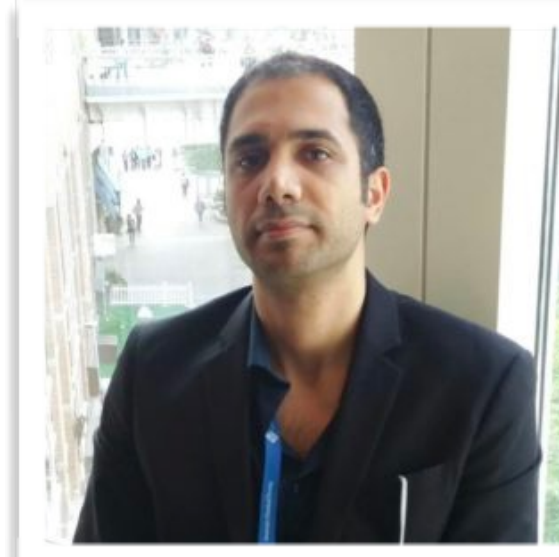
Taylor Barnes



Jing Chen



Susi Lehtola



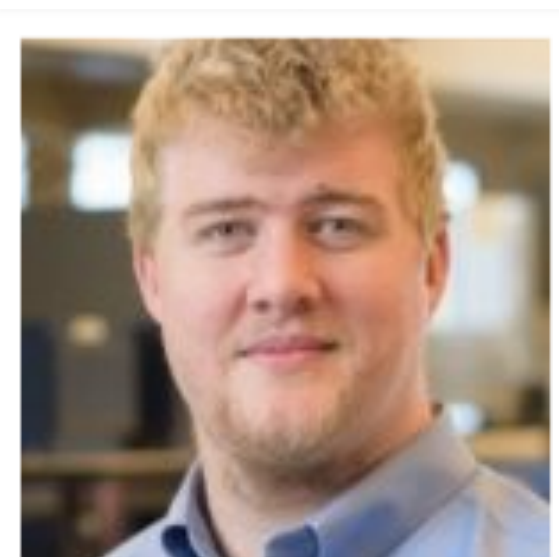
Sina Mostafanejad



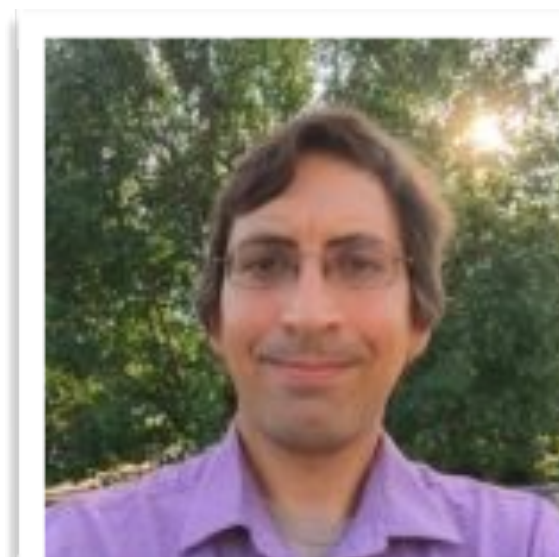
Sam Ellis



Jessica Nash
Education Lead



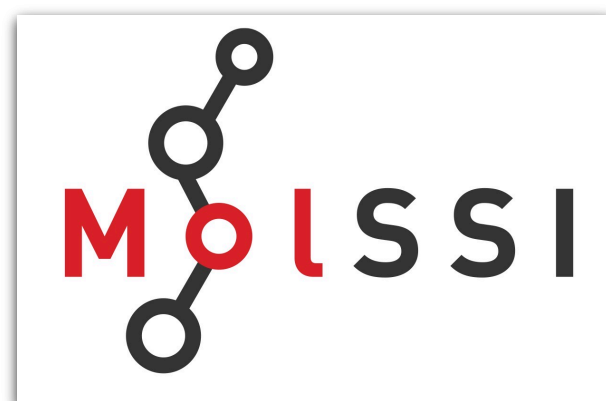
Levi Naden



Jonathan Moussa



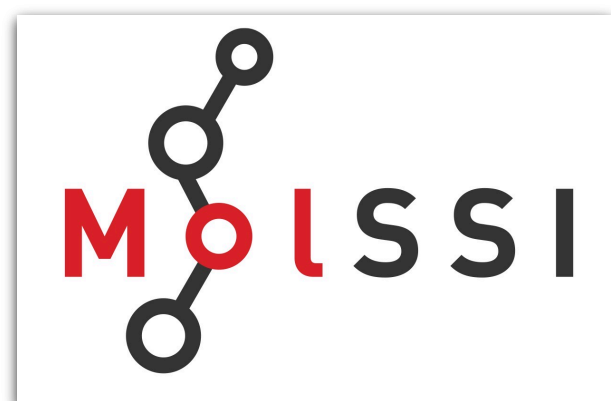
Ben Pritchard



MolSSI Education's Mission

To Provide Education and Training...

- MolSSI serves as an education and outreach nexus for the worldwide CMS community.
- MolSSI organizes summer schools, targeted workshops, high-school and undergraduate training programs, and on-line resources and classes to provide current and future CMS students with a modern and complete set of programming skills.
- MolSSI provides faculty development to help faculty upskill their own computational skills and develop discipline-specific curriculum.



MolSSI Education

Since 2017, over 2000 students have participated MolSSI Education workshops.

Our curriculum

Programming languages and tools...



...in the context of computational molecular science.

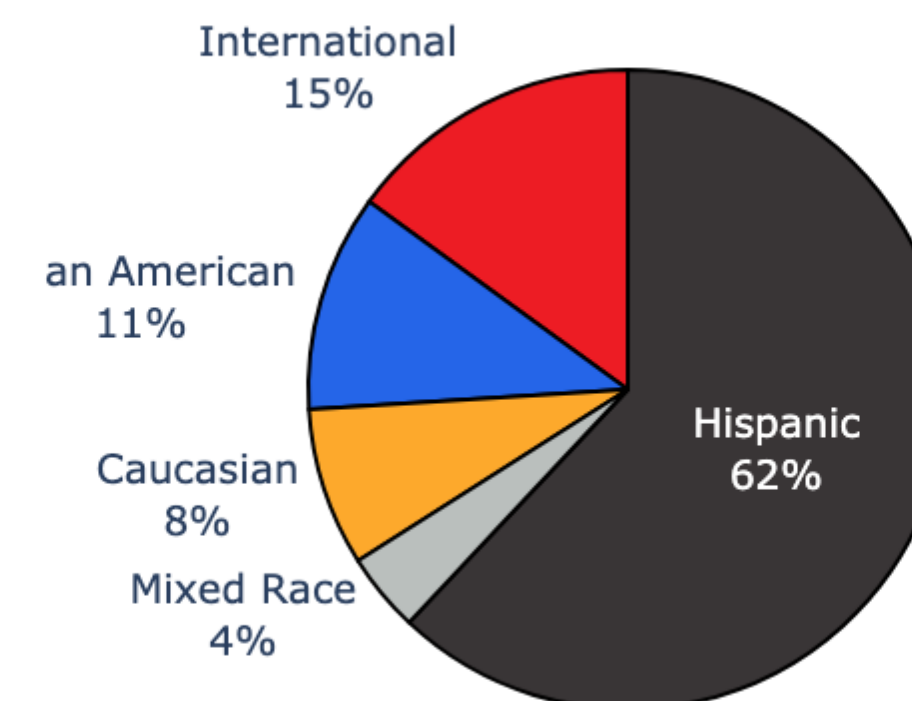
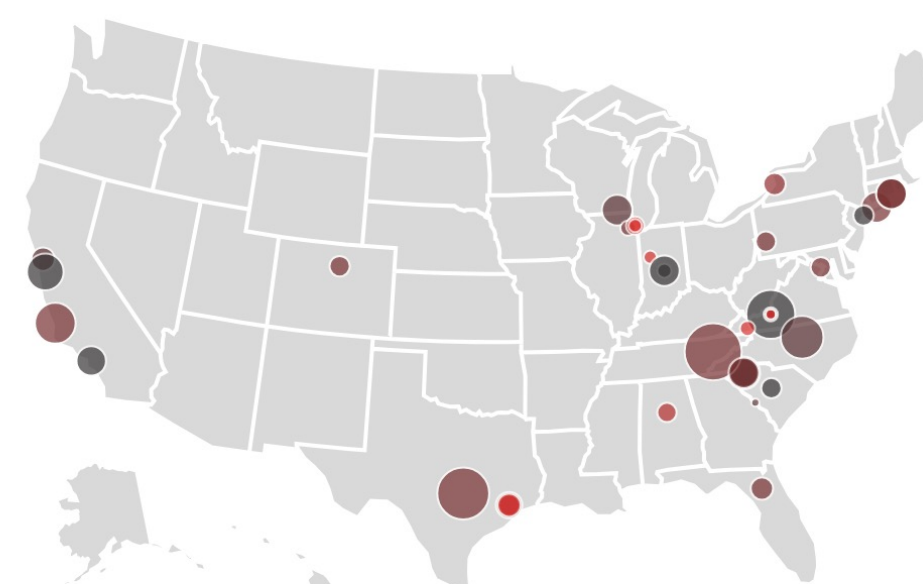
Online Lessons

<http://education.molssi.org/resources>



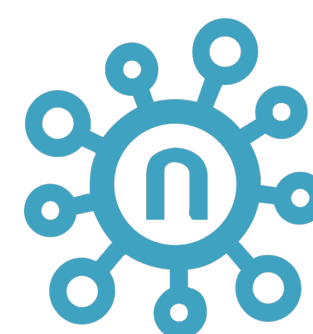
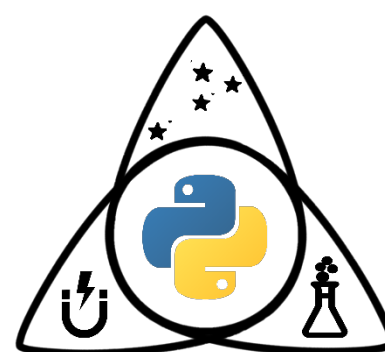
MolSSI Education engages audiences who are...

geographically and demographically diverse

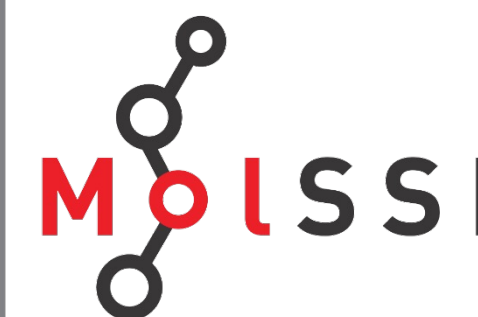


Student demographics for the 2021 MolSSI Workshop in collaboration with the Tapia Center at Rice University. Students in this program were **58% female.**

Partnerships and Collaboration



Partnership for Research and Education in Chemistry
PATHWAY TO DIVERSITY PROGRAM



RICE ENGINEERING
Richard Tapia Center for
Excellence and Equity

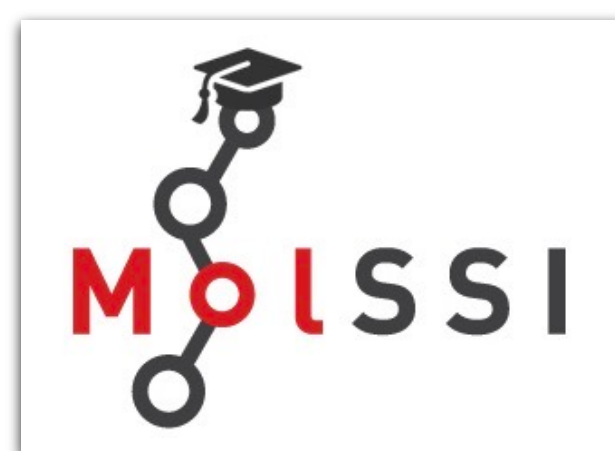
education.molssi.org

MolSSI Python Data and Scripting

Audience: Beginner to beginner+ programmers

How do I get started with Python?

- Jupyter notebooks
- Python syntax and control structures
- Reading and writing files
- File manipulation and parsing
- Analyzing and graphing data
- Writing functions
- Creating command line programs from Python scripts



education.molssi.org

A screenshot of a web page from MolSSI. The page title is "Python Scripting for Computational Molecular Science". It includes a description of the lesson, a list of prerequisites, and a schedule table. The schedule table lists seven topics with their respective durations and descriptions.

MolSSI Home

Python Scripting for Computational Molecular Science

This lesson by the [Molecular Sciences Software Institute \(MolSSI\)](#) teaches users fundamentals of scientific programming used in computational molecular sciences. The material is designed for undergraduate students, or other early career students, who have no prior programming experience. To see the full MolSSI's education mission statement, please see [here](#).

This lesson is under development, please report issues to the [GitHub repository](#)

Prerequisites

Students should be familiar with opening the Terminal window and creating and navigating files in bash.

Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction	What is the basic syntax of the python programming language?
00:45	2. File Parsing	How do I sort through all the information in a text file and extract particular pieces of information?
01:30	3. Processing Multiple Files and Writing Files	How do I analyze multiple files at once?
02:05	4. Working with Tabular Data	How do I work with numerical data presented in tables?
03:50	5. Plotting and Data Visualization	How do I visualize data by making graphs?
04:30	6. Writing Functions	How do I write include functions in my code?
05:10	7. Running code from the Linux Command Line	How do I move my code from the interactive Jupyter notebook to run from the Linux command line? How do I make Python scripts with help messages and user inputs using argparse?



CAL POLY

Data Analysis and Data Visualization

Audience: Beginner+

How do I use python to analyze data?

- Numpy arrays
- Pandas for data analysis
- SciPy and Data Fitting
- File manipulation and parsing

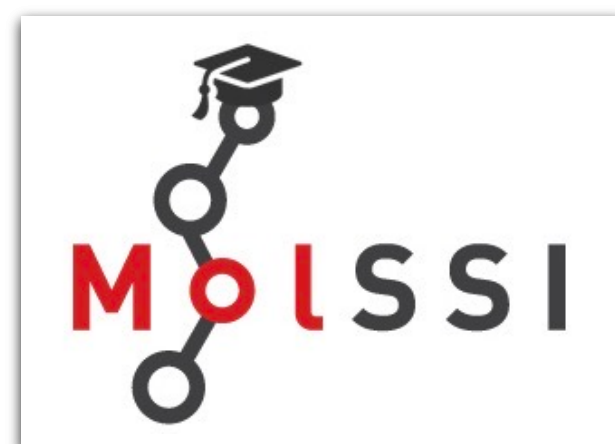
- Advanced data visualization
- Plots and subplots
- Multi-dimensional and interactive plotting

Schedule

	Setup	Download files required for the lesson
00:00	1. Working with Numpy Arrays	What are the differences between numpy arrays and lists? How can I use NumPy to do calculations?
01:05	2. Using pandas for data analysis	What is pandas? How do I access data in a pandas dataframe?
01:05	3. Using scipy for data fitting	How do I fit data to a specified function? How do I assess the quality of my fit? How do I determine the standard error for my fit parameters?



Scientific Visualization
Using Python



education.molssi.org



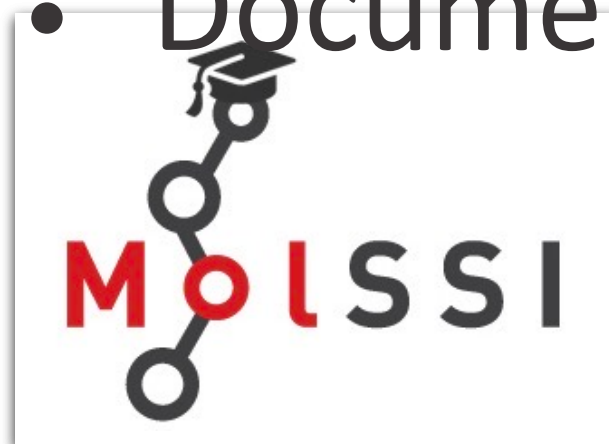
CAL POLY

MolSSI Best Practices in Python Development

Audience: Intermediate programmers

How do I share my code?

- Conda and Python environments
- How to structure your project?
- Version control using git
- Python Coding Style
- Code collaboration and repositories.
- Writing tests.
- Automatically running test.
- Documentation.



education.molssi.org

The screenshot shows the MolSSI website interface. At the top, there is a navigation bar with the MolSSI logo and the word 'Home'. Below the navigation bar, the main heading is 'Python Package Best Practices' in red. The text below the heading describes the lesson's purpose and provides a link to the MolSSI mission statement. It also mentions that the lesson is under continual development and encourages users to report issues or submit pull requests. A 'Schedule' section follows, listing eight topics with their corresponding start times and descriptions.

	Setup	Download files required for the lesson
00:00	1. Python Package Set-up	What is the layout of a Python package? How can I quickly create the structure of a Python package? What license should I choose for my project?
00:45	2. Intro to Version Control with Git	How do I use git to keep a record of my project?
01:20	3. Using GitHub	How do I use git and GitHub?
01:55	4. Python Coding Style	How can I write python code that is readable?
02:40	5. Deciding Package Structure	How should I break my code into modules? How can I handle imports in my package?
03:35	6. Code Collaboration using GitHub	How can others contribute to my project on GitHub? How can I contribute to the projects of others?
04:10	7. Python Testing	How is a Python module tested?
05:20	8. Continuous Intearation	How can we automate testing?



Finding MolSSI Tutorials Online



Home Resources Education Events Map Calendar Publications Workshop Requests

MolSSI Education Resources

MolSSI offers 1-2 day workshops as well as online tutorial materials. Most tutorials are hosted on GitHub in the MolSSI Education GitHub organization. Workshops and materials here may still be under development. Outside contribution is welcomed and encouraged!

Programming

Software
Development

Machine
Learning

High Performance
Computing

Molecular
Modeling

External
Resources

[Python Data and Scripting Workshop](#)

Description: The MolSSI Python Data and Scripting workshop is designed for students who are currently involved in, or planning to start computational chemistry research. This workshop is designed to help students develop practical programming skills that will benefit their undergraduate research, and will take students through introductory programming and scripting with Python to version control and sharing their code with others. NO prior programming experience is required.

► Workshop Topics

📄 [View Workshop Materials](#) | 🔄 [View GitHub Repository](#) | 📺 [View Workshop Recording](#)

[Python Data and Scripting for Biochemists and Molecular Biologists](#)

Description: The Python Scripting for Biochemistry and Molecular Biology Workshop is designed for students and faculty who are interested in getting started with coding as part of their teaching and research. The workshop provides hands-on python coding experience using examples relevant to biochemists. It includes parsing PDB files, data analysis, linear regression, nonlinear regression, and plotting data. No prior programming experience is necessary.

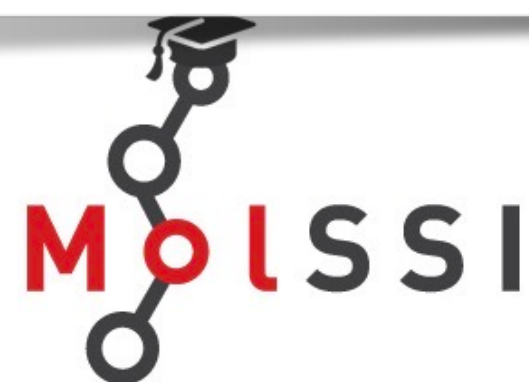
► Workshop Topics

📄 [View Workshop Materials](#) | 🔄 [View GitHub Repository](#)

We teach live workshops and webinars.

But, if you can't attend an event, our materials are available online.

Visit <http://education.molssi.org/resources>



Accessing MolSSI Resources through



education.molssi.org



nanohub.org/education



CAL POLY

nanoHUB

- Provides an online/cloud Jupyter Hub Platform.

This screenshot shows the nanoHUB interface for the tool 'Jupyter Notebook (202105)'. The page features a dark header with the nanoHUB logo and a 'MENU' button. Below the header, the tool title is prominently displayed. The main content area includes a 'Launch Tool' button, version information (Version 2.0, published on 24 Jan 2023), and a description: 'Starts the Jupyter notebook server in your home directory.' A sidebar on the left contains navigation options like 'About' and 'Tools'. A statistics box on the right shows 943 users, 0 citations, 0 questions, 0 reviews, and 0 wishes. At the bottom, there are tabs for 'About', 'Usage', 'Citations', 'Questions', 'Reviews', 'Wishlist', 'Versions', and 'Supporting Docs'.

This screenshot shows the nanoHUB interface for the tool 'Jupyter Lab (202105)'. The page features a dark header with the nanoHUB logo and a 'MENU' button. Below the header, the tool title is prominently displayed. The main content area includes a 'Launch Tool' button, version information (Version 1.0, published on 09 Feb 2023), and a description: 'An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook architecture.' A sidebar on the left contains navigation options like 'About' and 'Tools'. A statistics box on the right shows 96 users, 0 citations, 0 questions, 0 reviews, and 0 wishes. At the bottom, there are tabs for 'About', 'Usage', 'Citations', 'Questions', 'Reviews', 'Wishlist', 'Versions', and 'Supporting Docs'.



education.molssi.org

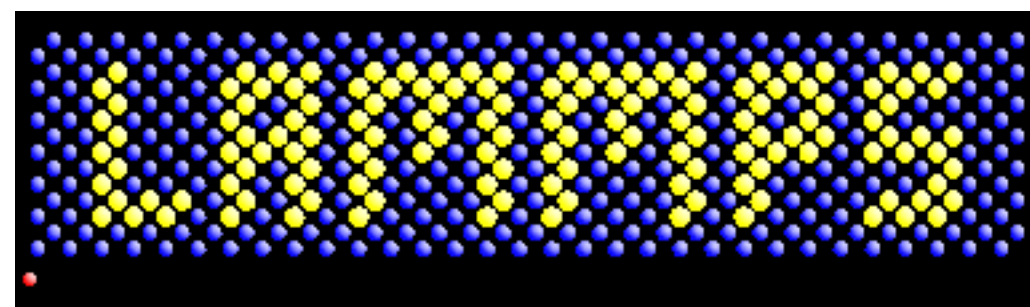


nanohub.org/education



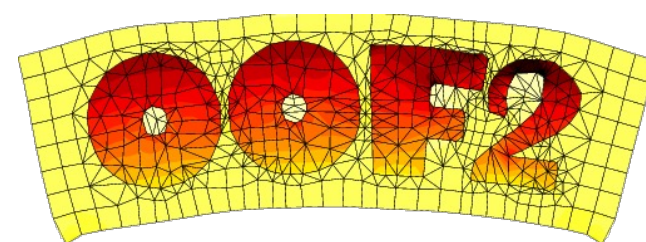
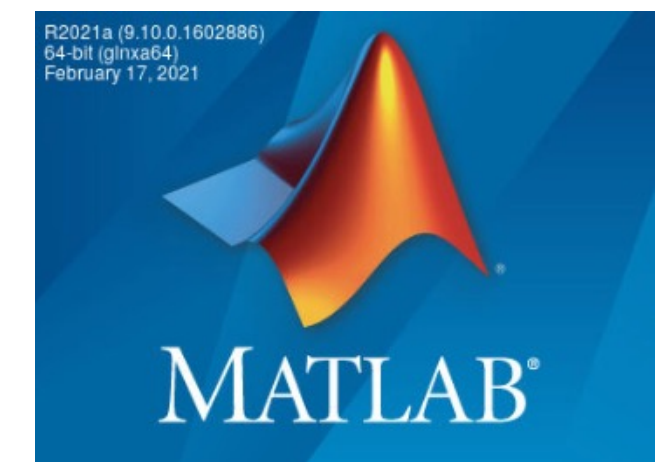
nanoHUB

- Has numerous chemistry and materials science packages already installed including:
 - Psi4 (molecular modeling)
 - LAMMPS (molecular dynamics)
 - PADRE (semi-classical device modeling)



Padre 2.4E

ORCA



SeqQuest Electronic Structure Code



education.molssi.org

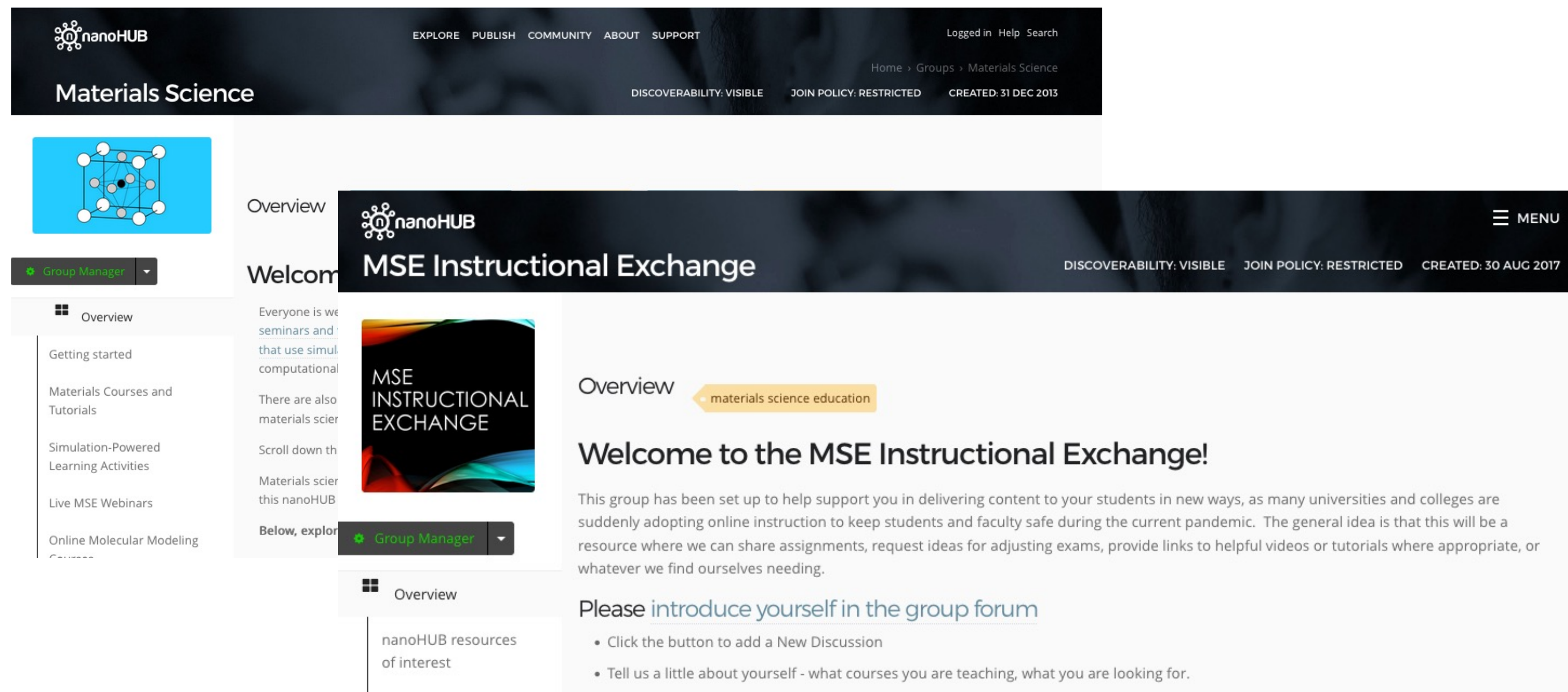


nanohub.org/education



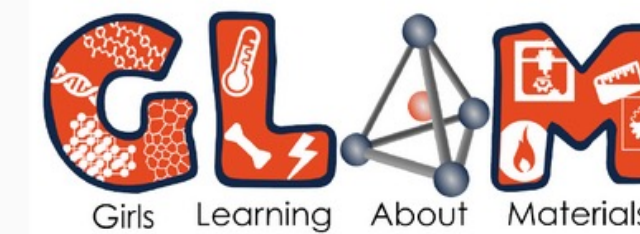
nanoHUB

- Allows the creation of groups, posting assignments and activities



The screenshot shows the nanoHUB interface for a Materials Science group. The top navigation bar includes 'EXPLORE', 'PUBLISH', 'COMMUNITY', 'ABOUT', and 'SUPPORT'. The group name 'Materials Science' is displayed, along with its discoverability (VISIBLE), join policy (RESTRICTED), and creation date (31 DEC 2013). A sidebar on the left contains a 'Group Manager' dropdown and a list of resources: Overview, Getting started, Materials Courses and Tutorials, Simulation-Powered Learning Activities, Live MSE Webinars, and Online Molecular Modeling. The main content area features a 'Welcome to the MSE Instructional Exchange' message, a 'Group Manager' dropdown, and a list of nanoHUB resources of interest.

Understanding Fracture Behavior in Materials Using Cheese



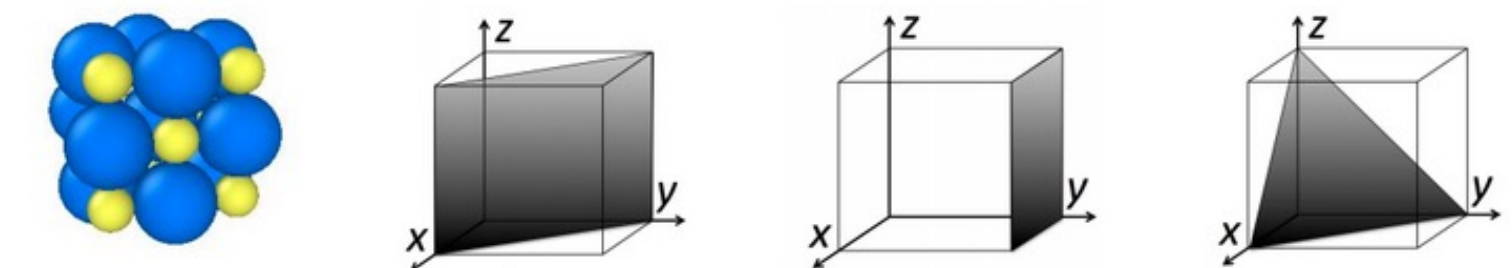
This is a resource from the Girls Learning About Materials (GLAM) outreach camp run at the University of Illinois at Urbana-Champaign. The learning objectives for this activity are:

1. Students will compare and contrast different types of fracture behavior (ductile/brittle).
2. Students will understand how to characterize fracture properties.
3. Students will apply their knowledge of fracture to predict how different

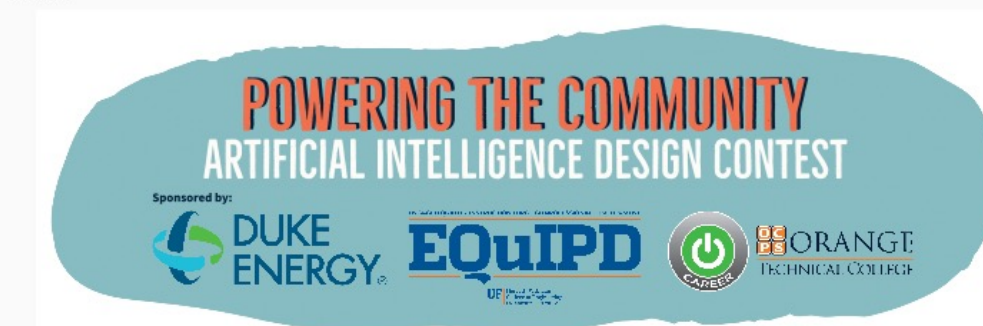
cheeses will fail.

4. Students will discuss how geometry plays a role in fracture behavior.

Visualizing Crystal Structures: An interactive group classroom activity



Abstract



These notebooks are for students completing Activity 3: Estimating Power Load within the PTC: AI Design Challenge.



education.molssi.org



nanohub.org/education



MolSSI Python Data and Scripting in nanoHUB

https://nanohub.org/courses/MolSSI_Python

The screenshot shows the nanoHUB website interface for the course "MolSSI Python Data and Scripting". At the top, the nanoHUB logo is on the left, and navigation links (EXPLORE, PUBLISH, COMMUNITY, ABOUT, SUPPORT) and user options (Login, Sign Up, Help, Search) are on the right. Below the navigation bar, the course title "MolSSI Python Data and Scripting" is displayed, along with a brief description: "The MolSSI Python Data and Scripting hands-on course is designed for students who are currently involved in, or planning to start computational chemistry research." A series of tags (chemistry, computational chemistry, computational science/engineering, data science, MolSSI, plotting, Programming, python, python training) are shown below the description. To the right of the tags is the MolSSI logo. Below the tags, a box indicates the course is "Brought to you by: The Molecular Sciences Software Institute". At the bottom of the page, there are two tabs: "Overview" (selected) and "Offerings". The "Overview" tab contains a detailed description of the course, its goals, and prerequisites. On the right side of the page, there is a summary box with the following information: "Enrolled: 85", "Course length: 5 Sessions", and "Estimated Effort: 2-3 hours each session". Below this summary is an "Enroll" button. At the bottom right, there is a section titled "About the Instructors" featuring a profile picture and the name "Ashley Ringer McDonald" from "Cal Poly San Luis Obispo".



education.molssi.org

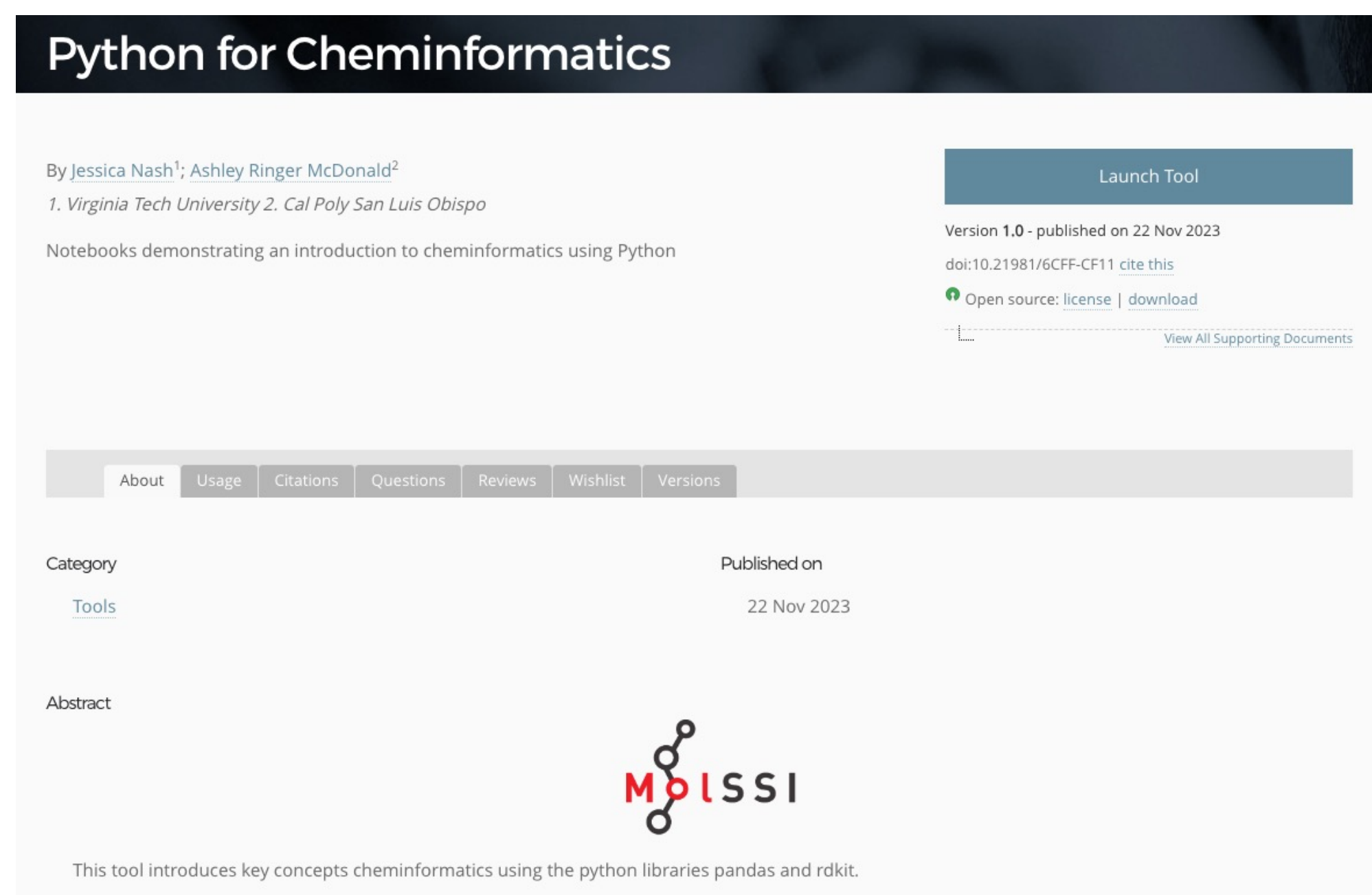


nanohub.org/education



New Cheminformatics workshop available as a nanoHUB tool

<https://nanohub.org/resources/38065>



The screenshot shows the nanoHUB interface for the tool 'Python for Cheminformatics'. At the top, the title 'Python for Cheminformatics' is displayed in white on a dark background. Below this, the authors are listed as 'By Jessica Nash¹; Ashley Ringer McDonald²' with affiliations '1. Virginia Tech University 2. Cal Poly San Luis Obispo'. A description states 'Notebooks demonstrating an introduction to cheminformatics using Python'. A prominent blue 'Launch Tool' button is visible. To the right, it shows 'Version 1.0 - published on 22 Nov 2023', a DOI 'doi:10.21981/6CFF-CF11', and links for 'Open source: license | download' and 'View All Supporting Documents'. A navigation bar includes 'About', 'Usage', 'Citations', 'Questions', 'Reviews', 'Wishlist', and 'Versions'. The 'Category' is 'Tools' and the 'Published on' date is '22 Nov 2023'. The 'Abstract' section features the 'MOLSSI' logo and the text 'This tool introduces key concepts cheminformatics using the python libraries pandas and rdkit.'



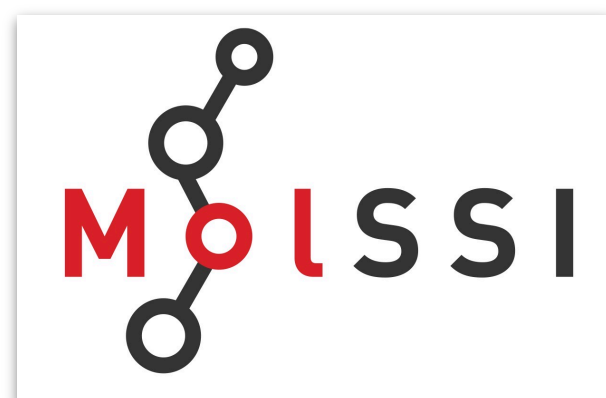
education.molssi.org

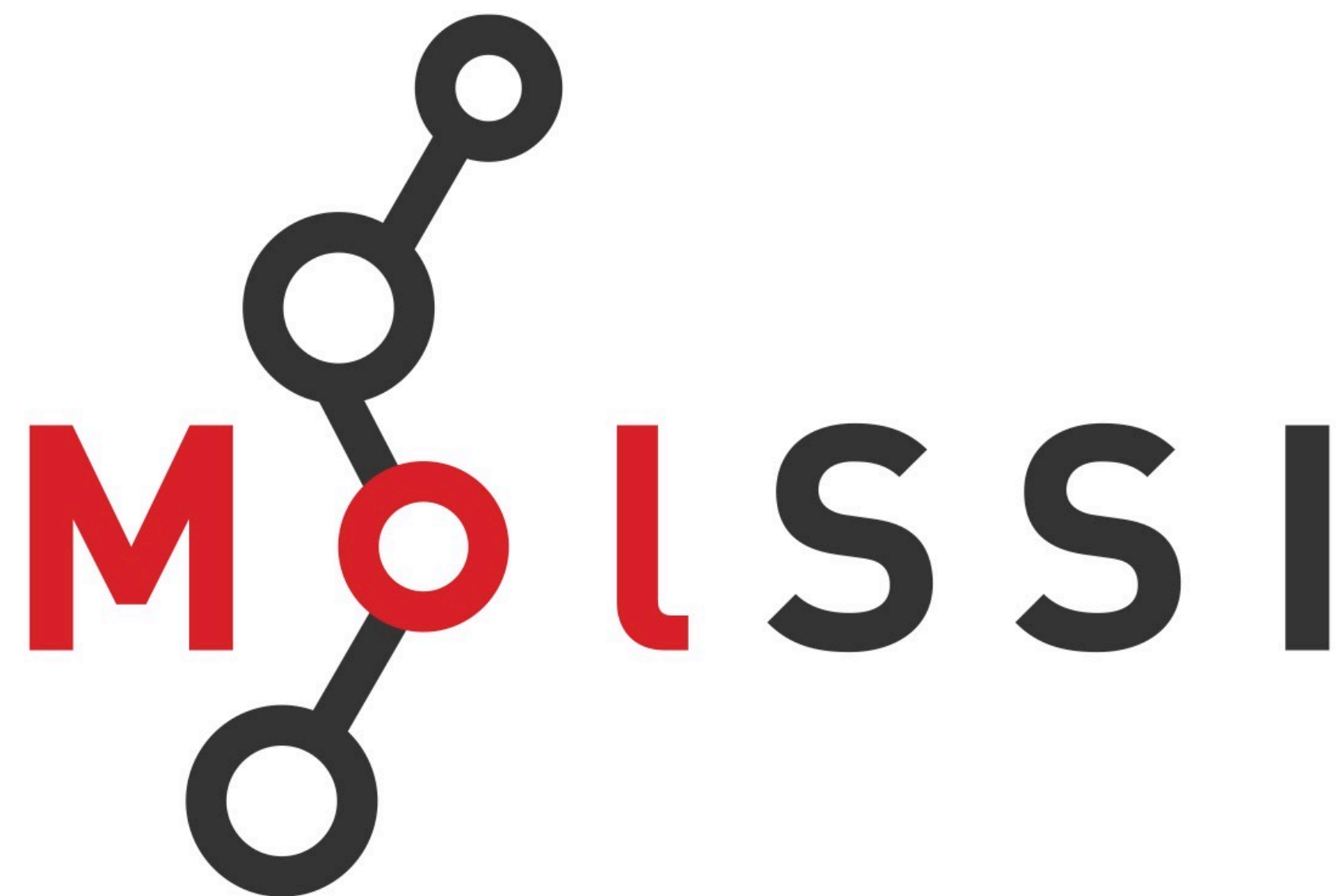


CAL POLY

Opportunities with the MolSSI

- MolSSI Software Fellows Program
 - Fellowship for graduate students and post-docs at US institutions who work on software development projects
 - Stipend, software development training, and mentorship from a MolSSI software Scientist
 - Applications open Feb. 15, 2024
- MolSSI Faculty Fellows Program
 - Fellowship for faculty at US institutions who are interested in incorporating programming and computation in their curriculum
 - Stipend, curricular design and assessment training, computational skills professional development, and mentorship
 - Applications open Jan. 29, 2024





education.molssi.org
NSF Award CHE-2136142



nanohub.org
NSF Award EEC-1227110