

2024 Summer Internships in Parallel Computational Science (SIParCS) Technical Project Descriptions

Undergraduate

- Project 1. Investigate the workflow and scaling of organizational CI/CD server
- Project 2. Improving data center visibility with AI
- Project 3. High Performance Data Assimilation
- Project 4. Geospatial Analysis of Machine Learning for Hydrometeorological Hazards, Uncertainties, and Impacts
- Project 5. Creating geoscience data analysis and visualization workflow examples in Python
- Project 6. Analysis of the Slingshot High Speed Network
- Project 7. Optimizing ensemble data assimilation performance for coupled Earth System models
- Project 8. Exploring NCAR's Campaign Store with Elasticsearch

Graduate

- Project 7. Optimizing ensemble data assimilation performance for coupled Earth System models
- Project 8. Exploring NCAR's Campaign Store with Elasticsearch
- Project 9. UX Research: Understanding user needs, preferences, and pain points with advanced data visualization software tools
- Project 10. Utilization of modern linear algebra libraries in the photolysis rate calculator
- Project 11. Unlocking Cloud Insights: Scaling HOLODEC's Holographic Cloud Particle Analysis with GPU Parallelism
- Project 12. OpenIoTwx Mesonet Edge Computing and Cyber Infrastructure Integration
- Project 13. Development of High-Performance & Scalable Data Analysis Routines for Unstructured Grids
- Project 14. Developing Advanced Machine Learning Architectures for Weather and Climate Problems

Note: Students may apply to up to two (2) SIParCS projects.

Non-Technical Project

Graduate

- CISL Outreach Diversity Education (CODE) Intern

Project 1. Investigate the workflow and scaling of organizational CI/CD server

Areas of Interest in order of relevance: Software Engineering, Development and operation

Description: CI/CD self-hosted runners were successfully deployed in our previous SIParCS projects. However, it is not convenient to exercise it across the organization, since it is only implemented at the user level, which requires every user to set up a self-hosted runner and set the repository to private for security purposes. So we need to further investigate the workflow and scalability of an organizational runner server. With this in mind, we want to research a centralized CI/CD server - similar to the Exascale Computing Project's CI/CD system - with advanced workflow, batch executors, and a downscoping pipeline. After prototyping such a server, we can collect information about the scalability, usage limitation, and hardware requirements etc. for future proposals. This project will familiarize ourselves with how to start a real-time CI/CD server within an organization and have a more automated workflow for our scientific simulation models.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Familiarity with Linux environment is required.
- Strong motivation to learn new skills and resolve issues in a team is preferred.
- Experience with high performance computing (HPC) cluster, CI/CD, Bash script is optional.

Project 2. Improving data center visibility with AI

Areas of Interest in order of relevance: Supercomputer Systems Operations, Machine Learning, Data Science

Description: The goal of this project is to evaluate and extend where possible, a monitoring framework (AIOPS), that utilizes machine learning and AI to provide insight into datacenter infrastructure operations. The intern will work closely with HPE, the vendor for NCAR's flagship supercomputer Derecho, as well as NCAR HPC professional staff to analyze real-time datacenter metrics. The objective is to provide feedback and improvements to enhance tooling to assist staff in identifying infrastructure issues quickly to ensure reliable operation of Derecho and prevent damage to the equipment.

Applicants should have a strong knowledge of AI and/or statistics. There will likely be opportunities to work with HPC professional staff from NREL as well.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Familiarity with Linux particularly shell scripting
- Familiarity with Artificial Intelligence and Machine Learning in particular Statistical methods for univariate anomaly detection models, machine learning and deep learning methods such as forecasting-based and reconstruction-based anomaly detection for both uni-variate and multi-variate models.

Project 3. High Performance Data Assimilation

Areas of Interest in order of relevance: Application Optimization/Parallelization, Software Engineering

Description: The Data Assimilation Research Testbed (DART) is a widely used community software facility for data assimilation. One application of data assimilation is improving numerical weather prediction. To do this, DART ingests a group of model forecasts, say 80 predictions of weather in the United States, and uses statistics to combine these model forecasts with observations to produce a better estimate of the weather.

The project will start with profiling of the DART code on NCAR's Derecho supercomputer. The student will use the profiling results, and their own experience to guide the direction of the project into one of the the following projects:

Satellite observations are a vital source of data for weather forecasting. Previous work to successfully scale DART to 100,000 processors has focused on dealing with high resolution model forecasts, and relatively small observation counts. Satellite data has disrupted this assumption: billions of observations per day are expected and will become the norm for data assimilation. This project would involve redesigning the DART observation sequence structure with a focus on parallelism, memory storage, and IO.

Initial work on compressing out unwanted model state for the Red Sea Data Assimilation project was completed by a SIParCs student. The work was specific to the MITgcm ocean model interface for DART and provided huge efficiency gains for the project. This project would be to generalize this approach and allow general masking of the state, which is applicable for land models and ice models where the ensemble may not be complete at every 3D grid location. Currently the check for incomplete ensembles happens within the assimilation code, which is both costly and error prone. DART already distributes and manages model state across nodes. The goal of this project is to add an abstraction layer to separate the logical representation of the model state from the compressed/masked version in memory.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Experience with a compiled language
- Interest in parallel programming
- Interest in high performance computing
- Familiarity with Linux

Project 4. Geospatial Analysis of Machine Learning for Hydrometeorological Hazards, Uncertainties, and Impacts

Areas of Interest in order of relevance: Data Science, Geostatistics, Visualization

Description: Hydrometeorological hazards, such as flooding and winter storms, are among the most costly hazards in the U.S., with major impacts on people's well-being, livelihoods, and the environment. These are increasing with climate change, as evidenced by recent, devastating floods from atmospheric river events in California. For these reasons, it is more important than ever to understand the dynamics of hydrometeorological events, in order to better understand what changes in policies and practices might help. The impacts of these events are heavily influenced by meteorological, geographical, and societal factors, so impact modeling approaches must take all three into account. Machine learning approaches are a promising way to integrate these different factors but still require significant validation.

The student intern will analyze data and machine learning models for predicting hydrometeorological hazards as well as associated societal impacts and the uncertainties associated with both. The student will work with a team of data, social, and atmospheric science experts as part of the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES; ai2es.org) to perform this analysis. They will use a variety of software tools to validate the models and analyze what factors most influence the accuracy and uncertainty of the predictions. They may incorporate a wide range of social data sources, including population, mobility, vulnerability, and survey data to identify disproportionate impacts or disproportionate drops in model performance.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Python programming experience.
- Experience with base scientific Python libraries, such as numpy, scipy, pandas, matplotlib, and Jupyter notebooks.
- Introductory statistics, data science, and/or GIS coursework
- Experience working with geospatial data.

Project 5. Creating geoscience data analysis and visualization workflow examples in Python

Areas of Interest in order of relevance: Visualization, Software Engineering, Data Science

Description: The NCAR Command Language (NCL) has long been an essential tool for data analysis and visualization in the atmospheric, oceanic, and earth sciences, but has recently been sunset in favor of building upon the existing open-source scientific python ecosystem. The Geoscience Community Analysis Toolkit (GeoCAT) team aims to aid the scientific community in their transition from NCL to Python in support of community-driven open-science. During previous summers, SIParCS interns have contributed to this goal by expanding the GeoCAT-Examples gallery and by contributing to our computational and visualization software stack.

This summer, we are looking for an intern to build upon the existing work in GeoCAT-examples, GeoCAT-viz, and GeoCAT-comp to create NCL-to-Python demonstrational content. The student will work with earth science datasets to contribute to a new era of pythonic visualization and computational demonstrations with Jupyter Notebooks, using tools such as matplotlib and cartopy, numpy, xarray, scipy, metpy, and dask, in addition to GeoCAT functionalities. To create these notebooks, the student will dive into NCL functions and explore the scientific Python ecosystem to demonstrate equivalent functionality. This will involve research tasks to understand the inner workings of existing NCL functions and their potential equivalents in popular open-source python packages, learning about different computational and visualization techniques, and working with earth science datasets to make notebooks relevant to our user community.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Comfortable writing python code.
- Familiarity with jupyter notebooks.

Project 6. Analysis of the Slingshot High Speed Network

Areas of Interest in order of relevance: Supercomputer Systems Operations, Computer Networking

Description: NCAR's latest flagship computer, Derecho, uses the Hewlett Packard Enterprise (HPE) Slingshot interconnect to link its 2570 nodes. This new interconnect is based on a butterfly topology, significantly departing from traditional HPC network architectures based on fat trees and hypercubes. Slingshot additionally employs many advanced techniques like adaptive packet routing in an attempt to improve network performance. The increased bisection bandwidth and lower network diameter compared to other topologies should give Slingshot an advantage, however this comes at the cost of increased complexity, making intuiting about the network difficult.

The student will have the opportunity to investigate this cutting edge network technology, used by many of the Top500 systems like Frontier, helping to quantify its performance attributes and behavior. The student will benchmark the limits of the network, and establish the bottlenecks by correlating them with the deluge of metrics being collected by the system. This will include investigating the behavior of Slingshot's non deterministic congestion control properties and quantifying the impact of "noisy neighbors" and network variability on job performance.

This work will help to inform Derecho's network tuning and job placement policies, and hopefully result in a publication authored by the student.

Students: The project is open to undergraduate students.

Skills and Qualifications:

- Familiarity with Linux particularly shell scripting
- Familiarity with computer networking including the OSI networking model and TCP/IP

Project 7. Optimizing ensemble data assimilation performance for coupled Earth System models

Areas of Interest in order of relevance: Software Engineering, Application Optimization/Parallelization, Data Science

Description: Numerical models of the Earth System couple components representing the ocean, atmosphere, land surface, and cryosphere in order to simulate climate processes and make predictions on time scales from minutes to centuries. While these models are increasingly comprehensive, they still struggle to simulate many features of the highly complex system. For this reason, there is rapidly growing interest in using data assimilation (DA) – traditionally the province of weather modeling – to bring observations to bear on coupled climate models. However, the computational cost of data assimilation for high-dimensional physical models makes many applications of Earth System data assimilation prohibitive, even on large supercomputers.

In this project, we are seeking a student to improve the interface between a comprehensive Earth System model (NCAR's Community Earth System Model, CESM) and a state-of-the-art data assimilation tool. NCAR's Data Assimilation Research Testbed (DART) is an ensemble DA tool that has been heavily tested and used for weather forecasting, ocean prediction, climate projections, flood prediction, parameter estimation, and other applications. In previous applications, DART has relied on modifying so-called "restart" files – which must be written to disk – to influence the evolution of numerical models. The attending I/O and model stop and restart places substantial limitations on overall model-DA performance. The goal of this project is to build and test novel capability for DART to access the model state in memory using the National Unified Operational Prediction Capability (NUOPC) system, which is used to exchange information across Earth System components. Work will be carried out using NCAR's new HPE Cray EX cluster, Derecho, which is a 19.87-petaflops system. The student will join a dynamic team of researchers with diverse expertise in Earth System modeling and data assimilation, and have the opportunity to learn about Earth system modeling, data assimilation, and high performance computing.

Students: The project is open to undergraduate and graduate students.

Skills and Qualifications:

- Undergraduate student
 - Required
 - Familiarity with the Linux environment and git version control
 - Motivation to solve problems and learn new skills is essential!
 - Helpful but not required
 - Experience with physical science and/or Fortran
- Graduate student
 - Required

- Familiarity with the Linux environment and git version control
- Experience with collaborative code development
- Enthusiasm for mentoring and collaborating with an undergraduate intern
- Motivation to solve problems and learn new skills is essential!
- Helpful but not required
 - Experience with physical science and/or Fortran

Project 8. Exploring NCAR's Campaign Store with Elasticsearch

Areas of Interest in order of relevance: Software Engineering

Description: NCAR has large (120 petabyte) storage systems for scientific data. With such voluminous data it is difficult for some scientists to find scientific data of interest for their work. Our goal is to index certain curated directories inside of the NCAR Campaign Store to allow scientists at NCAR to find data of interest to them.

This project is focused on continued development of a Java-based web application that interfaces with Elasticsearch. This summer we're enhancing our current software by including but not limited to the following: Increased metadata harvested from files, improved reports from Kibana, improved infrastructure to store state between system restarts, audit of files, and adding checksums to help determine data duplication.

Students: The project is open to undergraduate and graduate students.

Skills and Qualifications:

- All candidates
 - Ability to interact with diverse mentors and peers in a friendly, professional manner that supports collaboration and inquiry.
 - Good problem solving skills.
 - Good oral and written communication skills.
 - Willingness to learn and use software development tools and programs.
 - Curiosity to explore new things.
- Undergraduate students
 - Basic understanding of software development programming.
 - Basic experience in languages such as Java, Javascript, XML, HTML, CSS, and a query language like SQL.
 - Basic understanding of web services and web based user interfaces.
- Graduate students
 - Intermediate understanding of software development programming.
 - Intermediate experience in languages such as Java, Javascript, XML, HTML, CSS, and a query language like SQL.
 - Intermediate understanding of web services and web based user interfaces.

Project 9. UX Research: Understanding user needs, preferences, and pain points with advanced data visualization software tools

Areas of Interest in order of relevance: Visualization, Human Computer Interaction, User experience

Description: The Visualization Services and research group (ViSR) supports the earth science research community with their data visualization workflows, through the development of software tools and applications. VAPOR is one such software application designed to help meet the analysis and visualization needs of the geoscience research communities. It is an open-source, community-driven, interactive, 3D visualization tool which operates primarily on 3D arrays of time-varying, gridded data derived from numerical simulations. Throughout the past decade, VAPOR has steadily gained popularity within the geosciences field and has amassed a dedicated and engaged user community.

In order to extend VAPOR's reach and impact within the geosciences, it is imperative to understand our user needs, preferences, and pain points, both with VAPOR as well as their interactions with advanced data visualization software tools in a broader context. This project focuses on using user experience (UX) research methods to achieve this. As a SIPARCS intern, you would be collaborating with the ViSR staff to conduct exploratory-primary research (such as interviews, surveys) and secondary research (such as literature reviews and competitive analysis). In addition to gaining hands-on experience in the entire research process, from defining the research goals, creating a research plan, choosing the research methods, data collection and analysis, you'll also assume a critical role as an advocate for UX Research. This internship aims to empower the intern with practical experience in UX research, making a meaningful contribution to the continued success and utility of 3D advanced visualization software.

Students: The project is open to graduate students.

Skills and Qualifications:

- Enrolled in a relevant graduate degree program with a focus on Human Computer Interaction (HCI), UX research, Psychology, Data Visualizations, or a related field.
- Proficiency in UX methodologies, including practical experience (professional or academic) with user interviews, surveys, and usability testing.
- Strong verbal and written communication skills, including empathetic listening, to effectively engage with users, and stakeholders, ensuring a deep understanding of their needs, concerns, and feedback in the context of UX research.
- Excellent analytical skills and the ability to interpret research data, identify trends, and draw actionable conclusions.
- Familiarity with data visualization software such as VAPOR or ParaView is a plus.

Project 10. Utilization of modern linear algebra libraries in the photolysis rate calculator

Areas of Interest in order of relevance: Software Engineering, Numerical Methods, Application Optimization/Parallelization

Description: Atmospheric chemistry is a key component in an atmospheric model that interacts closely with many other processes such as aerosol microphysics, deposition and radiation. The complexity and non-linearity of the chemical system (e.g., gas phase, aqueous phase, photolysis) typically makes chemistry a computationally expensive part of atmosphere modeling. In particular, the calculation of photolysis rates is critical for atmospheric chemistry modeling. Estimations of photolysis rates using lookup tables are often used for their low computational cost. More accurate treatments that account for the effects of aerosols and other components of the atmospheric system, such as the TUV-x package (<https://github.com/NCAR/tuv-x>), are available but time-consuming to run on CPUs. These costs have prevented scientists from doing more complex simulations that assess the impacts of the changing climate and potential mitigation strategies.

The iterative solver for radiation in TUV-x uses a number of functions derived from LINPACK (<https://www.netlib.org/linpack/>). These linear algebra functions account for a significant portion of the computational time in the TUV-x radiation solver. LAPACK (<https://www.netlib.org/lapack/>) is designed to supersede LINPACK and perform more efficiently on modern architectures. Moreover, LAPACK can be connected to MAGMA (<https://icl.utk.edu/magma/>), which is a linear algebra library that supports heterogeneous computing (e.g., GPU). Evaluating the performance of LAPACK and MAGMA against LINPACK is likely to greatly enhance the attraction and adoption of these modern standard linear algebra libraries in this research community. In addition to the use of GPU-based linear algebra functions, the TUV-x calculations are parallelizable across wavelength bands and vertical columns that compose the 3-D model grid. Exploiting this parallelizability using GPUs has the potential to dramatically reduce the computational cost of online photolysis rate calculations.

The goal of this 2024 summer internship is to develop a LAPACK and parallelized MAGMA version of TUV-x (at least for some computations identified as hot spots). The student's primary focus will be 1) replacing the LINPACK functions by the corresponding LAPACK and MAGMA ones and 2) documenting the procedures, success or known issues. The student will also run various scenarios on the linux cluster at NCAR to verify the correctness of implementation and evaluate the performance.

Students: The project is open to graduate students.

Skills and Qualifications:

- Familiarity with Linux and any compiled language is required.
- Strong motivation to learn new skills and resolve issues in a team is required.
- Experience with high-performance computers, LINPACK, LAPACK or MAGMA is desirable.

Project 11. Unlocking Cloud Insights: Scaling HOLODEC's Holographic Cloud Particle Analysis with GPU Parallelism

Areas of Interest in order of relevance: Application Optimization/Parallelization, Machine learning, Physics

Description: This project is dedicated to harnessing state-of-the-art techniques employed in training Large Language Models (LLMs) and applying them to process extensive field-campaign datasets utilizing powerful neural networks distributed across hundreds of GPUs. At its core, a "U-Net" style neural network is currently utilized for particle recognition within holograms, following computational refocusing. The primary aim of this project is to facilitate the scaling and testing of the "distributed data parallel" (DDP) and "fully sharded data parallel" (FSDP) algorithms, both are cutting-edge parallelization methods, on our recently acquired Derecho supercomputer. This will enable the efficient processing of holographic data obtained through NCAR's HOLODEC instrument.

Over the summer, the successful applicant will work with scientists in the Computation and Information Systems Lab (CISL) and the Earth Observing Lab (EOL) toward writing a deployment-ready implementation of pyTorch's DDP and FSDP algorithms. Upon successful completion of the framework, the applicant will then scale the algorithm across hundreds of A100 GPUs available on the Derecho supercomputer. Depending on time, there may be opportunities to explore other potential pipeline frameworks on GPUs as well as designing new large neural networks that more effectively leverage the GPUs.

This project is developed jointly by CISL and EOL and is intended for graduate students.

Students: The project is open to graduate students.

Skills and Qualifications:

- Prior experience with machine learning, data analysis, and pyTorch.
- Prior experience with Python, Jupyter notebooks, and at least basic UNIX/LINUX proficiency required.
- HPC and GPU experience is desired but is not required.
- Fundamental knowledge of optics is desired.

Project 12. OpenIoT Mesonet Edge Computing and Cyber Infrastructure Integration

Areas of Interest in order of relevance: IoT (internet-of-things), Data Science, Supercomputer Systems Operations

Description: Low cost Internet-of-things (IoT) technologies have the potential to increase accessibility of communities to co-design and deploy sensor networks of environmental monitoring equipment. The open IoT project uses low cost IoT based electronics and 3D printed parts to create accessible open source instrumentation. openIoT can operate in both “small (low bandwidth) data” and “big data” configurations. Small and low bandwidth data configurations include standard measurement nodes such as digital rain, wind, and air (T,P,RH, air quality, etc). Big data modes may include high density, high frequency data such as those produced by a LIDAR (Laser Imaging, Detection, and Ranging) instrument, which could generate over 100,000 cloud point data per second or data from ultra high-resolution video or images from super-HD cameras.

While “low bandwidth data” mode can present challenges as the size of a sensor network scales up, the “big data” mode quickly creates more substantial challenges requiring creative edge computing architectures, communication, and cyber-infrastructure solutions. The goal of this project is to develop IoT edge computing protocols and integration with cyber infrastructure via CISL’s supercomputing resources or the NSF ACCESS resources. If we lived in a world where all stations can be deployed with high-speed multi-gigabit internet access then all of the data would simply be transmitted to a cloud based location for rapid analysis and modeling. However, in the case of atmospheric sensor networks (especially in remote or indigenous communities often out of reach of this infrastructure), we must 1) decide what data to store and what data to send within very limited communication constraints and 2) development of strategies and tools to post-process large scale IoT datasets using available cyberinfrastructure such as NCAR’s supercomputer or NSF ACCESS resources. The challenge, therefore, is to balance edge computing in the field with more advanced analysis and large scale sensor network management at the HPC infrastructure level. An important outcome of this project is to develop architectures and protocols for pre-processing and filtering down the data streams at the edge using SBCs (Single Board Computers) such as a Raspberry Pi 5 or NVIDIA Jetson Nano device attached on the same network as an openIoT station.

Students: The project is open to graduate students.

Skills and Qualifications:

- Strong Arduino and Python, MQTT or other communications protocols, LoRa WAN, RS485 protocols, IPFS and Jupyter Notebooks.
- Raspberry Pi (or similar SBC platforms) or Jetson Nano experience for edge computing algorithms would be helpful
- Big data analysis including downscaling, unstructured point cloud data streams, filtering and data transforms, image analysis, tensorflow, and experience with HPC systems.

Project 13. Development of High-Performance & Scalable Data Analysis Routines for Unstructured Grids

Areas of Interest in order of relevance: Data Science, Application Optimization/Parallelization, Software Engineering

Description: Project Raijin was awarded by NSF EarthCube in order to develop community-owned, sustainable, and scalable tools for performing standard data analysis and visualization routines on unstructured (ie. not regular lat/lon grids) climate and global weather meshes at global storm resolving resolutions. The development of Project Raijin leverages the Scientific Python Ecosystem (SPE), particularly the open development Xarray and Dask packages, and the Pangeo community. Our work is conducted under an open development model that encourages participation of the community in all aspects of the project's development. As a result of this, the majority of Raijin's development is centered around the UXarray Python Package, which is an Xarray-like package for directly working with and analyzing unstructured meshes that reside on a sphere (i.e. climate model output)

During this internship, you will work directly with the engineers and scientists behind the UXarray package to research, develop, and implement high-performance data analysis routines for unstructured grids on a sphere, such as computational operators, visualization methods, or efficient data structures. You will also learn high performance computing (HPC) principles through use of NCAR's HPC clusters as well as through researching and leveraging parallelization and optimization libraries such as Dask, Numba, CuPy, and Datashader. Most, if not all, of the student's work will be made publicly available through our open development model, which will in turn help this intern create a strong Python portfolio in data analysis and software engineering. The Project Raijin team is excited to provide the student with an in-depth experience of working within a professional software engineering team; therefore, the student will participate in all project development activities such as regular team meetings, morning standups, hackathons, cross-team discussion/collaboration meetings, debugging and bug-fixing, pair programming, documentation, etc.

Students: The project is open to graduate students.

Skills and Qualifications:

- Significant Experience with Python Programming
- Experience working with one of more of the following packages:
 - Xarray, NumPy, SciPy, Dask, CuPy, Matplotlib, Cartopy, Datashader, Holoviews, GeoViews, ScikitLearn, Numba
- Familiarity with Object Oriented Principles, either through Python or another programming language (C++, Java, etc.)
- Conceptual or Applied understanding of High Performance Computing
- Conceptual understanding of Climate Models
- Conceptual understanding of how Geoscience data is represented
- Comfortable with working with the following Mathematical concepts

- Representing data in terms of Vertices, Edges, & Faces
 - Working with Spherical Coordinates
- Optional Skills and Qualifications:
 - Familiarity with unstructured grids/meshes

Project 14. Developing Advanced Machine Learning Architectures for Weather and Climate Problems

Areas of Interest in order of relevance: Machine Learning, Data Science, Application Optimization/Parallelization

Description: Machine Learning numerical weather prediction models/digital twins have showcased dramatic improvements in forecast accuracy in the past year, showing competitive performance with the top physics-based models. However, these models exhibit artifacts and biases in their predictions, especially when integrated forward in time. Some of these issues may be addressed by changing the way these kinds of models are composed, trained, and deployed to make predictions. The intern will work with the NCAR Machine Integration and Learning for Earth Systems (MILES) group to test more advanced architectures, including variants of transformers and graph neural networks, on a relevant large-scale weather or climate problem that incorporates auto-regressive prediction. They will test different architectures, loss functions, and ways to roll out the model and will evaluate the accuracy, physical consistency, and uncertainty associated with the predictions. They will get to work with NCAR's new Derecho supercomputer, which features 100s of Nvidia A100 GPUs. The implementation of the project can be customized based on the intern's experience and interest.

Students: The project is open to graduate students.

Skills and Qualifications:

- Pursuing a PhD in computer science, physical science, math, statistics, or engineering.
- Advanced scientific Python programming experience
- Experience with Pytorch, Keras, Tensorflow, and/or Jax, including developing custom ML models.
- Experience with HPC systems preferred.
- Experience with geospatial data analysis.

Non-Technical Project:

CISL Outreach, Diversity, and Education (CODE) Intern

Areas of Interest in order of relevance: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

Description: *This is a paid student intern position.* As the CODE Intern, you will provide administrative support to the SIParCS Program office and affiliated programs. You will also assist with planning and preparation for education and outreach programs to occur during the 2024 - 2025 school year. This is a full-time (40 hours/week) student intern position that runs from **May 13 - August 2, 2024**. End date may be flexible.

Responsibilities

Student intern support:

- Be an active participant on the SIParCS team to provide support and mentoring for students.
- Live-in at the suite-style apartments with the interns, and plan and participate in after-hours team building activities.
- Keep program leadership informed of any issues that arise.
- May assist students/participants with special needs.
- May travel to assist with intern recruitment during fall months.
- Attend the Rocky Mountain Advanced Computing Consortium (RMACC) with the SIParCS program.

Summer program logistics:

- Assist with summer program support including planning and running events. Events include orientation, professional development workshops, field trips, and other learning opportunities for interns.
- Assist with apartment move-in and move-out logistics, distributing and collecting student supplies, and coordinating with apartment administration.
- Help write and edit SIParCS Annual Report.
- Update SIParCS program alumni tracking documents for program assessment and evaluation purposes.

General administrative support:

- Maintain program databases and updating web pages.
- Edit and compile information for documents, posters, brochures and newsletters.
- Provide multimedia support at program events (photos, video) and keep inventory of program pictures.
- Provide meeting support and prepare meeting materials.

School-year program support:

- Update presentation slides and other educational materials for school-year programs.

- May write and or edit educational materials such as text updates for HPC-TV and/or Meteo-AR.
- May test and evaluate new educational materials.
- May provide input to the new exhibit space at NWSC
- The above statements describe the primary work being performed. Additional tasks and professional development opportunities may be assigned according to the intern's skill set and interests.

Students: The project is open to graduate students in Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

Skills and Qualifications:

What you need:

- Must have attained a bachelor's degree at the time of application.
- Must be enrolled in a graduate level college, university or accredited professional program during the normal school year.
- Ability to work 40 hours/week from **May 13 - August 2, 2024** (dates to be confirmed). End date may be flexible.
 - NCAR has 3 unpaid holidays for interns during the summer internship (Memorial Day, Juneteenth, 4th of July).
- Experience with word processing, database, and spreadsheet applications in a Windows environment (Microsoft Word, PowerPoint, Excel, Access) as well as a Google Apps environment (Sheets, Docs, etc).
- Proven ability to organize, prioritize, and follow through on multiple tasks, with close attention to detail.
- Good written and verbal communication skills with the ability to convey information to interns, visitors, and partners in a welcoming and professional manner.
- Interest and sensitivity in working in diverse settings that include interns from a broad spectrum of underrepresented groups.
- Proven ability to work in a team and individually. A strong work ethic.
- Patience and adaptability.
- Discretion in handling confidential information.
- On-site participation in Boulder.

Preferred:

- Experience working with undergraduate and graduate students.
- Currently enrolled in masters level program in: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work, or related program.

The above qualifications describe the ideal skill set we look for in candidates. We encourage you to apply even if you do not possess all of the listed qualifications.

More information:

<https://www2.cisl.ucar.edu/siparcs/cisl-outreach-diversity-and-education-code-intern>