THE UNIVERSITY OF CHICAGO
DATA SCIENCE
INSTITUTE

Winter 2026 Data Science Clinic

Clinic Overview

The Data Science Clinic is a project-based course where students work in teams as data scientists with real-world clients under the supervision of instructors. Students are tasked with producing deliverables such as data analysis, research, and software along with client presentations and reports. Through the clinic course, Affiliate members gain access to undergraduate or graduate student teams to work on data science projects and explore proof of concepts while identifying top student talent. Projects are tailored and scoped to address company objectives with all deliverables overseen by the Clinic Director.

These unique collaborations allow Affiliate members to supplement their internal data science teams with outside support and perspectives, enlarging their capacity to experiment with new ideas. They also give students a window into a data science career, learning how companies build and use these tools internally.

Clinic Structure

Data Science Clinic runs during Fall, Winter and Spring quarters. Clinic projects are generally scoped to run for two full quarters. Each student works between 10 to 15 hours a week. Each team has a weekly 1-hour meeting with their assigned mentor and must submit a weekly progress report. Mentors are drawn from research staff, postdoctoral fellows and the faculty, subject to availability, interest and needs of the project. The mentor provides intellectual guidance, direct feedback to students and serves as a sounding board for both challenges and direction. The mentors will also provide support and guidance on any gaps in data science knowledge by providing literature and resources. Regular meetings are scheduled as it suits the client needs and to provide feedback to students.

<u>What does the ⚙ mean?</u>

If you look at the project descriptions below you will see that many of them have a gear/cog icon. These projects require a deeper knowledge of computing and preference will be given to those students who have demonstrated that capability.

Project List

Accountability Counsel

AI for Human Rights

Background:

Many international finance organizations maintain “accountability mechanisms”—systems for individuals and communities who have observed violations (environmental, labor, human rights) committed by internationally financed companies to submit a formal complaint to the finance organization. This offers communities an avenue to advocate for their rights, but the accountability mechanisms can be complex and difficult to utilize. Accountability Counsel’s mission is to help local communities craft and submit successful complaints against human rights and environmental violations.

Over the years, Accountability Counsel has collected data on thousands of complaints, including the text of the complaint and whether the complaint was deemed eligible and ultimately successful, and they’re looking to use the data they’ve collected to develop tools to help others write eligible complaints and advocate successfully. This project aims to develop AI models to characterize the likelihood that a given complaint will pass eligibility criteria, given the text of the complaint and relevant context about the accountability mechanism.

Mentor:

David Jacobson is the Data & Engineering Manager for the Data Science Institute Core Facility team. David specializes in leading machine learning and data science teams to develop practical solutions to complex real-world problems. His work is focused on data science projects for social good, especially related to climate, agriculture, human rights, and global health.

Apollo Care

Estimation of Deductibles

Background:

Apollo Care is a Chicago-based healthcare technology company that specializes in developing patient access and analytics solutions for pharmaceutical manufacturers, helping make prescription medications more affordable and accessible through innovative technology platforms. The company operates from downtown Chicago and serves pharmaceutical manufacturers as its primary market, offering integrated services including copay programs, hub solutions, and commercial analytics that optimize costs and improve patient outcomes.

Project:

Pharmaceutical companies often give discounts on prescription medicines to patients who meet certain economic or other criteria. One input into these discounts is the deductible, or the amount a patient has to pay before their insurance kicks in. The purpose of this project is to estimate the deductible for patients using medical payment data.

Having accurate deductible estimates for patients will enable pharmaceutical companies to more efficiently use discounting to provide better service to patients. Building an accurate model based off of medical claim data is the first step toward increasing this efficiency.

Data and Tools:

This project will utilize Python and machine learning methods on a set of historical anonymized patient claim data. Students will be expected to build and test different models on the claims data in order to achieve specific accuracy measures. This project will utilize commonly used machine learning techniques.

Mentor:

Austin Nafziger is the Senior Vice President of Product Strategy at Apollo Care, overseeing the firm's innovation roadmap. Over his career, Austin has helped launch several successful startups that have focused on improving efficacy of patient access programs; combatting fraud, waste, and abuse of traditional copay programs; and developing new tools and technologies to help manufacturers create strategic access for their products. He has also contributed to developing and managing pharmacy networks, hub programs, and distribution solutions. Austin holds a Bachelor of Science from the University of Michigan.

Argonne ⚙

Argo Research Agent: Generative AI-assisted Scientific Discovery at Argonne

Background:

Argonne is a multidisciplinary science and engineering research organization where talented scientists and engineers work together to answer the biggest questions facing humanity, from how to obtain affordable clean energy to protecting ourselves and our environment. The laboratory works in concert with universities, industry, and other national laboratories on questions and experiments too large for any one institution to approach alone.

Surrounded by the highest concentration of top-tier research organizations in the world, Argonne leverages its Chicago-area location to lead discovery and power innovation in a wide range of core scientific capabilities, from high-energy physics and materials science to biology and advanced computer science.

In late 2023, Argonne National Laboratory launched a secure, internal generative AI interface for the Argonne community called Argo. The initial implementation was a text-based conversational chatbot with access to OpenAI's GPT models hosted on private Azure instances. Since its debut, Argo now features a document upload tool, an API for programmatic access for science developers, multiple LLMs from OpenAI, Google, and Anthropic, and embedding models. We have also been working to integrate a retrieval-augmented generation (RAG)-based information retrieval mechanism supported by a knowledge graph to provide Argo users with direct access to Argonne-specific domain knowledge, such as operational policies, procedures, science records and data, and user facility and equipment documentation. While this feature is still under active development, its design was significantly supported by previous DSI student efforts. Much of their prior work, investigation, analysis, and prototype code are incorporated into our strategy and implementation design.

For 2025-26, we look forward to partnering again with DSI students to build on prior work with MRKL agents and other tooling to take our Argo solution to the next level: designing a deep research-like agent inside Argo to support scientists at Argonne with generative AI techniques that can enhance their capabilities to conduct research. This agentic extension will ingest and reason over a local corpus of Argonne-authored scientific publications, enabling advanced use cases such as structured literature synthesis, domain-specific Q&A, and multi-step reasoning. The prototype developed here will also serve as a candidate tool in a future user study (in collaboration with the UChicago AIR Lab) to investigate how generative AI tools impact human scientific capacity.

The University of Chicago DSI students will be challenged to learn, plan, and prototype the latest LLM-based engineering techniques into an existing generative AI interface that leverages vector and graph databases for supporting information retrieval. There may be multiple approaches to developing effective LLM-driven agents for this use case, including prompt-based iterative workflows, multi-agent systems, knowledge graph-assisted reasoning, reranking, and model context protocol (MCP) server architectures, which must be explored and evaluated to recommend the most appropriate techniques for the Argo implementation and future expectations. This project will directly support Argonne's mission to responsibly explore AI-assisted science and help position Argo as a next-generation research interface.

Mentor:

Matthew is an enterprise software engineer and Technical Lead for the AI for Operations initiatives at Argonne, with a Joint Appointment at UChicago. Matthew is also a Ph.D. student at the University of Illinois Chicago investigating advanced HPC management algorithms and LLMs for science and is an Adjunct Instructor in Computer Science at the University of Illinois Springfield.

AICE: AI for Climate ⚙

MonsoonBench

Background:

AI for Climate (AICE) is an interdisciplinary initiative that leverages advances in AI and expanding data availability to tackle critical challenges in climate prediction, impact modeling, and adaptation strategies. By uniting expertise from climate science, computer science, economics, physics, public health, and other fields, AICE develops novel tools—such as physics-informed predictive models and trustworthy datasets—to better understand climate dynamics and inform effective responses.

This project will transform a research benchmark for predicting the onset of the Indian monsoon into an open-source Python package. Accurate forecasting of the monsoon's spatiotemporal onset is critical for agriculture in India, yet the performance of new AI-based weather models on this task remains largely untested. Building on a methodology developed at the University of Chicago, the package will use precipitation forecasts to estimate onset dates, validate them against station-based rainfall data, and compute key skill metrics such as onset error, miss rates, and false alarms. It will also generate a visual scorecard to enable straightforward comparisons between AI-driven and traditional numerical weather prediction models. Deliverables include a PyPI package that implements the full benchmark using tools such as Xarray and Dask, providing researchers and stakeholders with a reproducible framework for evaluating monsoon prediction skill.

Mentor:

Pedram Hassanzadeh is an Associate Professor in Geophysical Sciences and Computational and Applied Mathematics at the University of Chicago and Faculty Director of AI for Climate. He leads the Climate Extremes Theory and Data Group, integrating theory, simulations, observations, and machine learning to study the dynamics and future of extreme weather. He earned his Ph.D. in geophysical turbulence and M.A. in applied mathematics from UC Berkeley and was a Ziff Environmental Fellow at Harvard. His honors include an NSF CAREER Award and an ONR Young Investigator Award.

Adam Marchakitus is a researcher in the Climate Extremes Theory & Data Group at the University of Chicago, where he develops and evaluates data-driven models for weather and climate prediction. A DSI Clinic alumnus, he graduated from UChicago in 2025 with a B.S. in Environmental Science and Data Science.

Becker Friedman Institute for Economics (BFI), University of Chicago

Healthcare Employment Geography Data Studio

Background:
The Becker Friedman Institute for Economics (BFI) serves as a hub for cutting-edge analysis and research across the entire University of Chicago economics community, uniting researchers from the Booth School of Business, the Kenneth C. Griffin Department of Economics, the Harris School of Public Policy, and the Law School in an unparalleled effort to uncover new ways of thinking about economics. Inspired by Nobel Laureates Gary Becker and Milton Friedman, BFI works with the Chicago Economics community to turn evidence-based research into real-world impact by translating rigorous research into accessible and relevant formats and proactively disseminating it to key decision-makers around the world.

This project extends the recent BFI working paper “The Rise of Healthcare Jobs” by developing an interactive Data Studio dashboard that enables users to explore healthcare employment across 130 Metropolitan Statistical Areas from 1980–2022. Using a dataset covering roughly 75% of the U.S. population with 26 socioeconomic variables—including healthcare and manufacturing employment shares, education levels, earnings, demographics, and Medicare eligibility—the tool will allow users to examine regional healthcare job growth in relation to manufacturing decline and other economic indicators. Built as a web-based interactive visualization platform, the dashboard will provide journalists, researchers, and policymakers with the ability to filter by region, compare metropolitan trends, and analyze correlations between healthcare employment and broader socioeconomic shifts, making complex labor market dynamics more accessible for research and policy analysis.

Mentor:
Eric Hernandez is the Senior Digital Media Manager at the Becker Friedman Institute for Economics. He brings extensive experience in digital content strategy and marketing, having previously strategized and created content for Organizing for Action, a non-profit organization that advocated for former President Barack Obama's political agenda. He also worked for Argonne National Laboratory developing marketing campaigns for divisions within the laboratory, giving him valuable experience in communicating complex research to diverse audiences.

Chicago Blackhawks

Retail Demand Forecast and Order to Buy Optimization

Background:

The Chicago Blackhawks, one of the NHL's "Original Six" teams founded in 1926, are six-time Stanley Cup Champions and play at the United Center. Guided by values of integrity, curiosity, empathy, collaboration, and originality, the organization is dedicated to reimagining the potential of hockey and creating legendary fan experiences. Beyond the ice, the Blackhawks engage diverse audiences and invest in the Chicago community through initiatives like the Fifth Third Arena youth hockey facility, the Chicago Blackhawks Foundation, and the acquisition of the Rockford IceHogs. Their mission reflects a commitment to both competitive excellence and meaningful community impact.

This project develops improved demand forecasting models to optimize inventory planning for Blackhawks retail apparel, where long lead times make accurate forecasts critical. Building on a baseline model that relies on prior-season demand patterns, the analysis will incorporate four years of transaction-level sales data and current inventory levels to forecast demand across gameday, non-gameday, and offsite retail channels. Using statistical modeling, time series forecasting, Monte Carlo simulations, and optimization techniques, the project will produce refined forecasts paired with risk management strategies to account for uncertainty in demand. Deliverables include enhanced forecasting methodologies coded in Python, a risk management framework, and a dynamic "order to buy" tool that enables business users to interact with forecasts and determine optimal purchase orders for the upcoming season.

Mentor:

Paulino Diaz is the Director of Analytics at The Chicago Blackhawks. He has been working in analytics for 7+ years and previously worked in strategic communications and public relations. Paulino received his Master of Public Policy Analysis from The University of Chicago.

City of Chicago, Department of Technology and Innovation

Energy Transparency and Building Performance in Chicago

Background:
The Department of Technology and Innovation (DTI) is the City of Chicago's central IT agency. DTI manages the City's core technology infrastructure, digital services, data management, data analytics, and applied data science. Our data team applies data science and advanced analytics, including model development, forecasting, and pattern detection, to strengthen decision-making, improve service delivery, and build more accessible, resident-centered digital tools.

This project analyzes the impact of Chicago's Energy Rating Placards on building performance since their introduction in 2019, using benchmarking data from 2015–2024. The dataset includes Energy Star scores, Energy Use Intensity (EUI), greenhouse gas emissions, water use, and star ratings, enabling a longitudinal assessment of efficiency and emissions outcomes across building types and neighborhoods.

The study will measure changes in energy efficiency and greenhouse gas intensity, compare performance trajectories of lower- versus higher-rated buildings, and apply machine learning models to identify building characteristics—such as size, age, use type, and energy source mix—most strongly associated with improvement.

Core outputs will include a report, interactive tool, or dashboard with metrics and visualizations showing trends by time, neighborhood, and building category. A stretch goal is to build predictive models that highlight which building types are most likely to achieve performance gains, supporting data-driven policy and management strategies.

Mentor:
Candice Stauffer, PhD, is a data scientist with extensive expertise in predictive analytics, machine learning, Python, and data visualization. She has nearly a decade of professional experience spanning academia, the private sector, nonprofit organizations, and government. Dr. Stauffer earned her doctorate in astrophysics from Northwestern University, where her research applied machine learning methodologies to model the behavior of astrophysical transients.

She currently serves as the Lead for Data Analytics at the City of Chicago, overseeing the development of data-driven strategies to enhance municipal services and inform decision-making processes. In addition, she is the founding president of Refactor 312, a civic technology nonprofit committed to advancing the use of data and technology to strengthen local communities.

City of Chicago, Department of Technology and Innovation

Affordable Housing and Short-Term Rental Restrictions

Background:

The Department of Technology and Innovation (DTI) is the City of Chicago's central IT agency. DTI manages the City's core technology infrastructure, digital services, data management, data analytics, and applied data science. Our data team applies data science and advanced analytics, including model development, forecasting, and pattern detection, to strengthen decision-making, improve service delivery, and build more accessible, resident-centered digital tools.

This project examines how short-term rental (STR) restrictions intersect with affordable housing in Chicago by combining datasets on affordable rental developments, prohibited STR buildings, foreclosures, and socioeconomic indicators. Using clustering techniques, spatial statistics, and descriptive analysis, the study will assess whether STR restrictions correlate with affordability outcomes and identify neighborhoods at risk of becoming housing "pressure zones." The core deliverable is an interactive map that overlays affordable housing with prohibited STR buildings, with filters for community area, housing density, and socioeconomic factors, paired with summary statistics and clustering results. A stretch goal is to integrate a predictive model—using methods such as logistic regression, random forest, or gradient boosting—into the map as a forecasting or risk-scoring layer, providing policymakers with a forward-looking tool to anticipate housing challenges.

Mentor:

Candice Stauffer, PhD, is a data scientist with extensive expertise in predictive analytics, machine learning, Python, and data visualization. She has nearly a decade of professional experience spanning academia, the private sector, nonprofit organizations, and government. Dr. Stauffer earned her doctorate in astrophysics from Northwestern University, where her research applied machine learning methodologies to model the behavior of astrophysical transients.

She currently serves as the Lead for Data Analytics at the City of Chicago, overseeing the development of data-driven strategies to enhance municipal services and inform decision-making processes. In addition, she is the founding president of Refactor 312, a civic technology nonprofit committed to advancing the use of data and technology to strengthen local communities.

Cook County Justice Advisory Council

Labeling the Law: Illinois Criminal Statute Data Enhancement

Background:

The Justice Advisory Council (JAC) coordinates and implements Cook County Board President Toni Preckwinkle's criminal and juvenile justice reform efforts and community safety policy development. The work of the Justice Advisory Council is guided by the county's Policy Roadmap, which identifies a central priority of building safe and thriving communities throughout Cook County. In collaboration with governmental and non-governmental stakeholders, the Justice Advisory Council devises, supports, and advocates for administrative reform within Cook County, as well as legislation which improves conditions and outcomes for individuals involved in the justice system. The JAC manages a portfolio of grants, primarily awarded to community-based organizations who work in geographic areas that have experienced historic disinvestment.

This project enhances the Administrative Office of the Illinois Courts (AOIC) "Offense Code Table" (OCT) by linking offense codes to the full statutory text from the Illinois Compiled Statutes (ILCS). The current OCT relies on truncated descriptions that require manual cross-referencing, limiting readability and analysis. By integrating OCT codes with complete ILCS descriptions, the project applies data science methods—such as text processing, clustering, and structured data integration—to produce clearer descriptions and analysis-ready groupings of offenses. This enables more robust examination of criminal legal trends, policy impacts, and charging practices across Illinois.

The enhanced dataset will reduce redundant lookups, establish standardized conventions for categorizing offenses, and support more precise studies of disparities, reforms, and legislative impacts. Designed for both operational and research use, it will provide court officials, attorneys, and policymakers with a machine-readable, scalable foundation for dashboards, models, and decision-support tools. At its core, the project creates a common analytical language for Illinois statutes, improving both clarity and reproducibility in criminal legal system research.

Mentor:

Whitney Key Towey, PhD, MPH, MSW – Director of Data and Research, JAC
Nico Marchio – Associate Director, Office of the Cook County Public Defender

Data, Policy, and Innovation Centre

AI-Enabled Grievance Redressal

Background:

The Data, Policy & Innovation Centre (DPIC) is a first-of-its-kind partnership between the University of Chicago and the Odisha government in India. Odisha is a coastal state in eastern India with a population of over 45 million. DPIC, funded by the state government and located in the state capital, Bhubaneswar, leverages large administrative datasets and cutting-edge data science and research to address complex development challenges and drive evidence-based governance. DPIC's work spans the full data-to-policy cycle - from curating high-quality datasets, extracting analytical insights, conducting rigorous research and building data capabilities within government.

This project applies data science and natural language processing (NLP) techniques to modernize *Janasunani*, the Government of Odisha's grievance redressal platform, by improving efficiency in handling the 1.8 million citizen complaints filed between April 2021 and June 2025. The clinic will develop workflows to digitize unstructured English-language documents, extract entities such as schemes, departments, and locations, automatically summarize long-form grievances, and classify complaints into thematic categories. The outputs will include a structured dataset of summarized and categorized complaints and reproducible code pipelines. Beyond reducing manual entry and administrative burden, the work will generate actionable insights at scale, turning citizen feedback into a tool for governance improvement and demonstrating how data science can strengthen trust between governments and citizens.

Mentor:

Dr. Urmila Chatterjee is the Executive Director at DPIC and former Research Director at the Energy Policy Institute at the University of Chicago in India (EPIC India). Previously, she served over a decade as Senior Economist at the World Bank, leading policy dialogue, research, and lending projects across South Asia on topics including human development, firm growth, fiscal policy, energy reform, and climate adaptation. She has managed large teams, mentored junior colleagues, and published widely in policy journals, reports, and newspapers. Earlier in her career, Urmila worked at the Indian Institute of Management and Citigroup. She holds a Ph.D. in Economics from UC Berkeley, an M.A. in Economics from the University of Mumbai, and is a Chartered Financial Analyst.

Ecdysis ⚙

Deep Learning to Characterize Species Diversity

Background:
Ecdysis Foundation's mission is to support the evolution of a regenerative food system using science, education, and demonstration. The Ecdysis 1000 Farms Initiative partners with farmers to characterize the chemical and physical properties of soil, water dynamics, and species diversity on their land. To further expand this program, Ecdysis is utilizing AI and deep learning algorithms to characterize species diversity using audio and image data.

Currently Ecdysis collects data on avian species diversity on a small number of farms via visual observation—a.k.a. bird-watching. This process is too resource intensive to scale to 1000+ farms, so they want to develop a deep learning pipeline for identifying bird species and characterizing overall abundance and diversity based on hour-long audio recordings taken on the farms. The audio recordings contain a mix of human and natural sounds, including bird calls. This project aims to build a pipeline that will separate the relevant from irrelevant sounds, match bird calls to species, and characterize the likely abundance of each species.

Mentor:
David Jacobson is the Data & Engineering Manager for the Data Science Institute Core Facility team. David specializes in leading machine learning and data science teams to develop practical solutions to complex real-world problems. His work is focused on data science projects for social good, especially related to climate, agriculture, human rights, and global health.

Fermi National Accelerator Laboratory ⚙

Graph Neural Networks for Liquid Argon Time Projection Chambers

Background:

Fermilab is America's particle physics and accelerator laboratory, hosting numerous experiments and international collaborations dedicated to solving the mysteries of matter, energy, space, and time.

Neutrinos are the lightest known matter particles, interacting only through the weak nuclear force, which makes them difficult to study but central to questions about matter–antimatter asymmetry and possibly Dark Matter. Fermilab, in partnership with the University of Cincinnati, is developing NuGraph, a Graph Neural Network (GNN) for particle reconstruction in Liquid Argon Time Projection Chamber (LArTPC) detectors. NuGraph classifies charge measurements on detector wires by particle type, achieving 94% accuracy and 97% cross-plane consistency. The project will expand GNN capabilities with new decoders for clustering nodes from the same particle, optimize information flow and resource use, and tune hyperparameters. Models will be trained on multiple datasets, including public MicroBooNE data, to improve generalization and reduce dependence on specific detector features.

Mentor:

Giuseppe Cerati received his Ph.D. in Physics and Astronomy from Università degli Studi di Milano – Bicocca in 2008. At Fermilab since 2016, he is currently a Scientist and head of the Data Science and AI department. He collaborates on neutrino experiments such as MicroBooNE, ICARUS, DUNE and on collider experiments such as CMS, with focus on physics analysis and data processing algorithms (both traditional and artificial intelligence).

The Impact Project

NLP for Federal Government Impact Mapping

Background:

The Impact Project collects and synthesizes data on how government change affects communities, overlaying information on local economies, industries, services, and demographics. At a time of rapid shifts in federal policy and funding, there is a critical need to track and explain how these changes ripple through states, districts, and neighborhoods. Reliable, timely information enables policy debates grounded in evidence and helps local leaders and advocates respond effectively.

This quarter, students will develop models to extract structured information about federal government funding, workforce, and policy changes from news articles, public reports, and other unstructured sources. The resulting data will expand the Impact Project's "Impact Map," allowing users to trace government decisions to their local consequences. Students may also investigate the trends in this impact database — identifying sectors most affected, geographic disparities, or emerging themes in government activity.

Mentor:

From The Impact Project:

- Abby André, Director @ The Impact Project
- Jonathan Gilmour, Data Lead @ The Impact Project

Invenergy

Community Sentiment Analysis for Renewable Energy Development

Background:

Invenergy, North America's largest privately held renewable energy developer, is partnering with a DSI team to analyze community perceptions of proposed renewable projects. Students will design pipelines to process diverse sources of text data, including public comments, town hall transcripts, and local media coverage. Using natural language processing techniques such as sentiment analysis, topic modeling, and large language models, the team will identify recurring themes, key stakeholders, and drivers of both support and opposition. The analysis will also explore methods for constructing a "community risk score" that can provide Invenergy with an evidence-based framework to anticipate challenges and strengthen community engagement strategies.

This project offers students hands-on experience at the intersection of data science and renewable energy development. Students will gain practical skills in web scraping, text preprocessing, and unstructured data analysis, while applying advanced machine learning methods to real-world policy and industry questions. The project will also emphasize clear communication of findings through dashboards and visualizations, giving the team the opportunity to deliver actionable insights that could directly influence the future of renewable energy deployment.

Mentor:

Sophie Logan is a Senior Data Scientist on the Enterprise Analytics and AI team at Invenergy, a leading renewable energy developer. She leads projects that apply advanced analytics, machine learning, and large language models to accelerate the integration of clean energy onto the grid. Her work spans wind, solar, storage, and transmission, informing decisions from site selection to long-term operations.

Before joining Invenergy, Sophie worked at the World Bank as a climate finance specialist, developing models of countries' environmental metrics, and founded a data science consulting company serving climate technology sectors from building decarbonization to sustainable fashion. She holds a BA in Economics and German from Bard College and an MS in Computer Science and Public Policy from the University of Chicago.

Sophie is passionate about harnessing data and AI to tackle complex challenges in the energy transition and excited to mentor the next generation of data scientists eager to drive innovation at the intersection of technology and climate solutions.

Karczmar Lab, Department of Radiology, University of Chicago ⚙

Predicting Breast Cancer Treatment Response Using Blood Vessel Networks

Background:

The Karczmar Lab in the University of Chicago Department of Radiology is a leading research group specializing in advanced MRI techniques for cancer imaging and treatment prediction. Led by Dr. Gregory Karczmar, Director of MRI Research, the lab has pioneered innovative MRI methods that are now used in clinical breast cancer screening. The lab combines expertise in MRI physics, medical imaging analysis, and computational methods to develop novel approaches that directly improve cancer detection and patient care.

This project aims to develop a novel approach to predict how breast cancer patients will respond to treatment by analyzing the patterns of blood vessels around tumors. Blood vessels form natural networks that change in response to treatment, and we hypothesize that these vascular patterns can serve as early markers of treatment effectiveness. The ultimate goal is to enable personalized treatment decisions and potentially develop risk assessment tools for earlier cancer detection.

The core innovation involves representing each patient's blood vessel structure as a mathematical graph, where vessel branch points become nodes and vessel segments become edges, annotated with properties like length, width, and curvature. Students will implement Graph Neural Networks (GNNs) to analyze these vascular networks, as GNNs can capture both local structural changes and global network patterns critical for understanding patient-specific responses. The project will involve building a complete pipeline from MRI data to graph representations, training and evaluating GNN models against traditional approaches, and identifying which vascular features are most predictive of treatment outcomes.

Mentor:

Dr. Gregory Karczmar is a Professor of Radiology and Medical Physics and Director of MRI Research, with more than 30 years of experience developing new MRI techniques now used in breast cancer screening. Dr. Milica Medved is a Research Associate Professor of Radiology with over 20 years of experience developing advanced MRI methods for cancer detection and risk prediction. Dr. Zhen Ren is a Research Assistant Professor of Radiology specializing in advanced imaging techniques, including dynamic contrast-enhanced MRI and image reconstruction for breast cancer research.

Metropolitan Water Reclamation District of Greater Chicago ⚙

Oak Park Roof Analysis for Stormwater Management

Background:

The MWRD manages wastewater and stormwater in Cook County, Illinois. Established in 1889, it operates one of the largest wastewater treatment systems in the world, serving over 5 million residents. The MWRD's mission is to protect public health and the environment by treating and reclaiming water, managing stormwater, and reducing pollution. It also plays a key role in flood control and water quality improvement.

This project uses high-resolution satellite imagery, machine learning, and GIS to identify and analyze the largest rooftops in Oak Park. By generating a distribution of roof sizes, the project will help answer practical questions such as how many roofs exceed a certain area and highlight properties suitable for downspout disconnection. The analysis will be integrated with property records and environmental data, creating a comprehensive view of stormwater management opportunities. Students involved will gain hands-on experience in data analysis, imaging technology, and environmental science. The outcomes include a software tool, detailed maps, and a final report with recommendations that can be scaled across the Metropolitan Water Reclamation District, advancing sustainable urban development and climate resilience.

Mentor:

Richard Fisher is a Principal Civil Engineer in the Stormwater Division of the MWRD’s Engineering Department. He oversees stormwater master planning, pilot studies, and various stormwater programs. With over 30 years of experience, he has planned, designed, and managed numerous public and private capital improvement projects.

Mothers Out Front

School Lookup Tool for Environmental Action

Background:

Mothers Out Front is a grassroots movement of parents and caregivers organizing for healthy, climate-safe schools and communities. Local teams often confront fragmented, technical information spread across federal and state databases, district websites, and mapping tools—slowing advocacy and putting at a disadvantage schools with fewer resources. Volunteers need school-specific answers to practical questions: Are there grants or rebates for electrification or solar? Has the district adopted a climate or sustainability plan? Is there known lead risk in drinking water? What is the rooftop solar potential? By assembling these signals in one place and pairing them with clear next steps, the tool will help community members move quickly from data to action—supporting equitable, evidence-based campaigns at the school and district levels.

Project Objective:

Students will build an interactive lookup tool (search by school name, district or location) that assembles an actionable environmental profile for each school/district, including information such as:

- Documents describing climate-related policies and incentives, which could be parsed and categorized using LLMs.
- Drinking water lead/copper results from EPA ECHO.
- Potential for environment-friendly facilities improvements, including rooftop solar potential indicators and possible HVAC and electrification upgrades.
- Electric school bus adoption and opportunities.
- Public support in your district for climate-friendly investment.

The tool should give parents and advocates the information they need to prioritize and activate for meaningful change in their school and district.

Mentor:

- Camille Greer - Co-Executive Director @ Mothers Out Front
- Jenny Zimmer - Co-Executive Director @ Mothers Out Front

Promega Corporation ⚙

Biological Organoid Development

Background:
Promega is a global biotechnology company, headquartered in Madison, Wisconsin, that provides products and solutions for life science research, drug discovery, and human identification. The company manufactures reagents, enzymes, and other biochemicals used in fields like genomics, protein analysis, cellular analysis, and forensics, serving academic, government, and industrial researchers worldwide. Promega is also committed to sustainability, employee development, and community engagement, striving to create innovative, science-driven solutions for real-world challenges.

This project leverages data science and AI methods to predict the suitability of organoids for scientific use by analyzing early-stage characteristics. Using a dataset of organoid images, chemical composition information, and survey data collected by Promega, the study aims to build supervised learning models that classify organoids as “good” or “bad” after 30 days of growth. The approach will begin with tabular data to establish baseline predictors and then extend to image analysis with computer vision techniques, enabling the identification of key morphological and chemical signals associated with successful development. The resulting models will improve efficiency in organoid research by reducing time, cost, and experimental waste.

Mentor:
Liya is a Data Scientist at the University of Chicago and Thomas Kirkland is a Senior Scientific Investigator at Promega.

Sustainable Fisheries Partnership

Descriptive Models to Characterize Fishery Sustainability

Background:

Sustainable Fisheries Partnership (SFP) maintains an extensive database of fishing stock profiles, with scores and descriptions characterizing the sustainability of fishing operations, called FishSource. FishSource consolidates and summarizes the main scientific and technical information needed by seafood buyers to gauge the sustainability of the fisheries they are sourcing from and take actions to help improve them. To collect the data for FishSource, SFP analysts manually review thousands of fishing stock reports to extract relevant information and summarize the sustainability practices (or lack thereof) of each fishery.

To improve the quality and maintainability of their data, FishSource would like to develop an AI-enabled pipeline to generate descriptions to characterize the strengths and weaknesses of fisheries sustainability practices. This will involve developing natural language models to read stock reports, incorporate relevant statistics, and write meaningful descriptions to inform consumers about sustainability practices.

Mentor:

David Jacobson is the Data & Engineering Manager for the Data Science Institute Core Facility team. David specializes in leading machine learning and data science teams to develop practical solutions to complex real-world problems. His work is focused on data science projects for social good, especially related to climate, agriculture, human rights, and global health.

University of Chicago Library

Optimizing Library Storage

Background:

In support of free inquiry and expression, the University of Chicago Library is transforming the global knowledge environment to be open, accessible, and equitable. We enable the University of Chicago and our greater community to create a better world through effective information services, a comprehensive connected collection, and a culture of innovation, respect, and partnership.

This project will develop predictive models to help the Library decide which single-volume print monographs should remain in on-site storage. Using ten years of anonymized circulation data combined with bibliographic metadata such as subject classifications, publication dates, and call numbers, students will analyze patterns of past usage and estimate likely future demand. The core deliverable is a tool that produces optimized on-site storage lists for different collection sizes (e.g., 30,000 to 3 million volumes) and allows adjustments by class or subclass to reflect curatorial priorities. If time allows, the team will also create a report that explores distinctive patterns of usage and collection strength within selected Library of Congress subclasses, highlighting areas of unusual demand or uniqueness.

Mentor:

David Bottorff is the Collection Management & Circulation Services Librarian for the University of Chicago Library and is leading the Library's strategic priority of developing a comprehensive collection management and storage plan for its physical collections.

University of Chicago Transportation

CTA 171/172 Service Change Analysis

Background:

The University of Chicago's transportation infrastructure relies heavily on both Chicago Transit Authority (CTA) public bus routes and the university's UGo shuttle system to connect the Hyde Park community to campus. With potential service cuts to CTA routes 171 and 172, there is an urgent need to understand how these changes would affect community access to campus and identify mitigation strategies. These routes serve as critical transportation arteries for students, faculty, staff, and community members who depend on public transit to reach university facilities.

This quarter, students will conduct a comprehensive geospatial analysis to map transportation vulnerability across Hyde Park and develop data-driven recommendations for service optimization. The project will produce heat maps identifying blocks most impacted by route cuts—specifically areas lacking easy alternatives through existing UGo shuttle connections. Additionally, the team will analyze current UGo shuttle routes to propose strategic adjustments that could minimize disruption, while quantifying the expected ridership impacts on remaining CTA services. The end goal is to provide university leadership with actionable insights to inform both advocacy efforts with the CTA and internal shuttle service planning.

Mentor:

Beth Tindel is the Director of Transportation & Parking Services at the University of Chicago. In this role, Beth oversees University-wide commuter, parking, traffic, and transit functions. She is responsible for parking functions in the University's parking structure and surface lots and she manages the University's agreement with the CTA, the daytime and NightRide shuttles, charter buses, and the University's various alternative transportation programs. Beth earned a Bachelor of Arts in Studio Arts from the University of Georgia.

University of Northern Iowa

Emotional Progression in Young Adult Novels

Background:

The University of Northern Iowa (UNI) is a public university that offers more than 90 majors with a total enrollment of about 9000 students. UNI's main reputation is for its rich history in teacher preparation.

This project applies Natural Language Processing (NLP) techniques to a curated corpus of award-winning young adult novels to visualize emotional trajectories across texts authored by majority and diverse population writers. By modeling the progression of eight major emotions throughout each novel, the analysis will compare average emotional patterns between groups to identify distributional differences in how emotions are expressed. The resulting visualizations will highlight systematic variation in emotional progression linked to author background, providing quantitative insight into literary expression and representation in young adult fiction.

Mentor:

Taraneh Matloob is an Associate Professor of children's literature at the University of Northern Iowa. She teaches multicultural children's literature as well as doctoral courses. Her scholarly interests are focused on multicultural children's literature, sentiment analysis, virtual reality, and augmented reality