

# Sean Kulinski

🏠 seankulinski.com

✉ smkulinski@gmail.com

☎ (773) 844-4447

📍 San Francisco, California  
(U.S. Citizen)

## EDUCATION

Aug. 2019 – Dec 2023

### PhD of Engineering in Computer Engineering

Purdue University, West Lafayette, Indiana

- *Area of Study:* Machine Learning and Artificial Intelligence
- Some Foci: Robust ML, Generative AI, Efficient AI

Aug. 2015 – May 2019

### Bachelor's of Science in Electrical Engineering

Purdue University, West Lafayette, Indiana

- *Area of Study:* Automatic Controls

## RESEARCH EXPERIENCE

May 2023 – Sept 2023

### Research Intern ( Deep Learning; Efficient AI)

**Microsoft Research** (*Advised by: Ankur Mallick and Kevin Hsieh*), Redmond, Washington

- Created a forecasting model that can predict the future performance of large machine learning models (e.g., large language models or other foundation models) that are deployed on high dimensional streaming data – enabling predicting model failure *before* it happens.
- Developed a compute+data efficient finetuning algorithm that can mitigate ML performance degradation due to distribution shifts or applications to novel domains.
- Worked with a research + product team to integrate these works into the Azure Machine Learning monitoring toolbox as well as patent the algorithms.

Aug. 2019 – Present

### Ph.D. Research (Machine Learning; Natural Language Processing; Computer Vision)

**Purdue University** (*Advised by: Dr. David Inouye*), West Lafayette, Indiana

- Creating a causally-grounded generative AI model that generates counterfactual examples that answer the question “What would this look like if X had happened instead of Y?” (e.g., what my chest x-ray look would like if I had it imaged at Hospital B instead of Hospital A.)
- Derived methods for interpretable domain translation for the purposes of explaining distribution shifts to a human operator which can be used for system monitoring or knowledge discovery .
- Constructed a new large-scale multi-agent computer vision (CV) dataset based on human matches of StarCraft II that exhibits complex and shifting agent behaviors yielding 1.8 million images with multiple data representations such as ones that can be used as a drop-in replacement for CIFAR10 and MNIST.
- Created a light-weight machine learning algorithm that uses deep density models to detect shifts in distributions and localize the shift to specific problematic feature(s) that caused the shift–allowing for online monitoring with little additional overhead .

May 2018 - Aug. 2019

### Lead Undergraduate Researcher for Prosthetic Haptic Interfaces Project

**Purdue University** (*Advised by: Dr. Hong Tan*), West Lafayette, Indiana

- Began international collaborative research project with Centro di Ricerca “E. Piaggio” Lab from University of Pisa, Italy.
- Lead experiment to test effects of proprioceptive feedback of prosthesis delivered via skin stretching for upper limb prostheses.
- Developed GUI for data collection for human subjects with automatic data parsing.
- Designed and manufactured wearable motorized device to encode the amount a robotic hand is opened.

## PUBLICATIONS

Present

- **Sean Kulinski**, Kevin Hsieh, Mayukh Das, Chetan Bansal, David Inouye, Ankur Mallick, “Prediction and Preemptive Mitigation of Failures in Large Machine Learning Models” *Under Review*

Present

- **Sean Kulinski\***, Zeyu Zhou\*, Ruqi Bai\*, Murat Kocaoglu, David Inouye, “Towards Characterizing Domain Counterfactuals For Invertible Latent Causal Models” *Under Review* (Preprint Available)

- July 2023 • **Sean Kulinski**, David Inouye, "Towards Explaining Distribution Shifts", *International Conference on Machine Learning (ICML)*. PMLR, 2023.
- June 2023 • **Sean Kulinski**, Nicholas Waytowich, James Zachary Hare, David Inouye, "StarCraftImage: A Dataset For Prototyping Spatial Reasoning Methods For Multi-Agent Environments" *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (CVPR) 2023*
- April 2022 • **Sean Kulinski**, Saurabh Bagchi, David Inouye, "Towards Explaining Image-Based Distribution Shifts", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition VDU Workshop. (CVPR), 2022.*
- Dec 2020 • **Sean Kulinski**, Saurabh Bagchi, David Inouye, "Feature Shift Detection: Localizing Which Features Have Shifted Via Conditional Distribution Tests", *Neural Information Processing Systems (NeurIPS)*, Dec 2020.

## WORK EXPERIENCE

---

May 2022 – Aug 2022 **Data Scientist Intern** (Natural Language Processing; Search)

**Microsoft**, Seattle, Washington

- This work with Microsoft365 Research studied using generative language models (e.g., LLMs) to improve enterprise search results in Microsoft Apps by adding related search terms to the user's search query.
- We used Natural Language Processing (NLP) models (e.g., BERT, GPTx) to generate related search terms for a given query and designed an additional NLP model to evaluate the relevance of the generative additions – which together greatly improved the query alternations.
- Identified and explored bridging the gap between web search methods (e.g., Google search or Bing search) and enterprise search methods (e.g., Outlook search or Teams search).

Oct 2021 – May 2022 **Machine Learning Scientist** (Computer Vision)

**AbbVie**, Remote [San Francisco, California]

- Led the design and development of a novel computer vision model for processing histopathological images for the purpose of cancer detection and downstream diagnosis.
- Developed robust high-performance pipeline for continuous analysis of whole slide images for deployment to consumers.
- Assisted in deploying a consumer-facing ML model and custom viewer+annotator web-app for displaying mappings and meta-statistics generated by the model.

May 2020 – Aug 2020 **Research Intern** (Computer Vision)

**Lawrence Livermore National Laboratory**, Livermore, California

- Research position for identifying issues in state-of-the-art computer vision frameworks for detection of COVID-19.
- Built computer vision models to conquer some of these issues, such as being robust to spatial distribution shifts. The models were trained using Livermore's Sierra HPC system.
- Used Natural Language Processing techniques on parsed Material Science publications to create an interpretable deep model to aid in the discovery of new nanostructures and nanomaterials.

Jan. 2019 – Aug. 2019 **Software and Embedded Electrical Engineer**

**Indiana Microelectronics**, West Lafayette, Indiana

- Developed Genetic Algorithm to automate and optimize design of transmission zero filters for Lockheed Martin.
- Designed automated testing of temperature drift for a closed-loop linear piezoelectric motor.
- Oversaw testing, calibration, and reworks for a phased-array filter system.

Aug. 2019 – Dec. 2019 **Graduate Teaching Assistant for ECE Senior Design**

**Purdue University**, West Lafayette, Indiana

- Directed a lab of 16 students through the process of bringing a hardware/software project from idea to product.
- Acted as a Senior Engineer and Manager to four different projects simultaneously.
- Held external workshops in various ECE skills such as Python, PCB design, and device testing.

*May. 2017 – Aug. 2017***Robotics R&D Testing Engineer****Stryker Corp.**, Fort Lauderdale, Florida

- Created dynamic power consumption measuring tool for loads on an arm-assisted surgical robot.
- Designed, laid out, and printed circuit board for high frequency and large duration measurements.
- Created tool for automatic formatting and parsing of raw input data for the purpose of immediate analysis.

**TALKS AND LECTURES**

---

*Feb 2023***Understanding Distribution Shifts and How We Can Learn from Them**

Invited talk to Laboratory of Data Science to discuss my research on Distribution Shifts covering:

- What is Distribution Shift and why is it significant?
- Whether distribution/domain shifts can be seen as “a feature rather than a bug”?
- How can we use distribution shift explanations to learn actionable environmental information?
- How can training on real-world distribution shifts lead to better generalizable machine learning models?

(Recording of talk [here](#))*Feb 2022 - Mar 2022***Current Methods in Domain Generalization**

Led five lectures on the current approaches in Machine Learning for Domain Generalization covering:

- What is Domain Generalization, and it's main thrusts?
- Why do we want domain invariance, how is IRM different than ERM, and what are people doing now?
- Showed how data augmentation can be seen as a tool for domain generalization, with an example using *Mixup*.
- Additional methods in domain generalization such as Meta-Learning, Domain disentanglement, etc.

*Nov 2020 - Dec 2020***Primer on using Optimal Transport in Machine Learning**

Led introductory lectures on applying Optimal Transport in Machine Learning with four sessions:

1. What is Optimal Transport?
2. How can Optimal Transport be computed?
3. Walking through toy problems of exact computation of Wasserstein distances, visualization with McCann interpolations, and approximation using entropic regularization and Sinkhorn iterations.
4. Optimal Transport applications in Machine Learning such as WGAN and domain adaptation.

*Sep 2019***ECE Design Skills Workshop**

- Walked through how to use Python for automation and control of simple motorized and visual systems using a Raspberry Pi, stepper motors, and LCD screen.
- Showed process of laying out a printed circuit board design using Altium and how to output GERBER files for fabrication.
- Walked through PCB assembly process and showed suggested debugging practices for both hardware and software.

**SKILLS**

---

**Programming Languages and Frameworks**

Python, Pytorch, Tensorflow, AzureML, LaTeX, Bash, Batch, Git, Numpy, Pandas, Scikit Learn, Scipy, PyOT, OpenCV, Excel, Linux, Jupyter

**More Skills**

Working with generative models, Large Language Models, A/B testing, leading and working within teams, strong mathematics background, statistics, data analysis, development using high-performance computing systems, deployment of ML systems, product research and design

**LEADERSHIP AND  
ACTIVITIES**

---

2015 - Present Co-leader of Purdue's ML Reading Group (Present), Conference Reviewer (NeurIPS, ICML, AIS-TATS) (Present), Led lectures involving Python and Altium (2019), Elected Executive Board Member of Alpha Tau Omega Leadership Development Fraternity (2017), Volunteer at Natalie's Second Chance Dog Shelter (Present), Eagle Scout and Avid Backpacker (Present)