

Cara Van Uden

email: cara@cs.stanford.edu
github: @caravanuden

Education

MS in Computer Science

Stanford University | Sept 2021 - June 2023 (expected)

- **Concentration:** Artificial Intelligence
- **Teaching assistant:** Data Management and Data Systems (1 term)

BA in Computer Science & BA in Cognitive Science

Dartmouth College | Sept 2015 - June 2019

- **Thesis:** “Comparing brain-like representations learned by vanilla, residual, and recurrent CNN architectures”. Oral defense with high honors.
- **GPA:** 3.87/4.00, summa cum laude
- **Teaching assistant:** Introduction to Computer Science (6 terms), Foundations of Applied Computer Science (1 term)

Papers

- **Van Uden, C. E.,** Nastase, S. A., Connolly, A. C., Feilong, M., Hansen, I., Gobbin, M. I., & Haxby, J. V. (2018). “Modeling semantic encoding in a common neural representational space.” *Frontiers in Neuroscience*.

Projects

- Data quality validation and cleaning with Deequ
- Visual similarity product clustering for browse
- Image classification API with ResNet50
- Decoding neural representations of emotion in resting-state fMRI data

Awards

- Wayfair Hackathon Finalist (2021)
- Wayfair Hackathon Winner (2019)
- Phi Beta Kappa (2019)
- High Honors Thesis in Computer Science (2019)
- Neukom Award for Outstanding Undergraduate Research in Computational Science (2019)
- Academic Award in Cognitive Science (2019)

Skills

- **Languages:** Python, C++, C, Scala, R
- **Frameworks:** PyTorch, PySpark, SparkML, Keras, Docker, Airflow
- **Data:** SQL, Hadoop/Hive, BigQuery
- **Cloud:** Google Cloud Platform

Work Experience

(Incoming) Research Scientist Intern

whiterabbit.ai | Summer 2022

Machine Learning Engineer II

Wayfair | Feb 2020 - Aug 2021

- Built and deployed machine learning pipelines for computer vision, NLP, and matching at scale. Redesigning and moving pipelines from on-prem to GCP.
- Pipelines preprocessed data, extracted features from text and imagery, and performed model prediction for product matching. Applications included product deduplication and competitor price matching.
- Was sole technical contributor for first three months; onboarded rest of the team. Collaborated with data scientists and other engineering teams.
- Pipelines generated annualized \$180M in GRS during tenure on team.

Data Scientist I

Wayfair | Aug 2019 - Feb 2020

- Worked on the computer vision team. Incorporated noisy environmental imagery into the “visually similar” product recommendation pipeline.
- Reduced product coverage gap by 50%. Decreased runtime by 20%.

Translational Data Science Intern

Celgene | Summer 2016

- Built an exploratory data analytics and visualization tool for analyzing gene expression and drug response data.
- Used site-wide by scientists for exploratory target deconvolution/validation in translational drug development for blood and bone marrow cancers.

Research Experience

Research Intern, Stanford AI Lab, AIMI, and BMI

Stanford University | Oct 2021 - present

- Various multimodal (patient EHR and images) representation learning and few-shot learning projects. Deploying one such model in a hospital setting for pneumonia screening.

Research Intern, Computational Cognitive Neuroscience

Dartmouth College | Jan 2018 - June 2019

- Used fMRI data alignment and forward encoding models to predict neural responses to naturalistic video stimuli across people. Models demonstrated improved spatial specificity and model performance compared to previous single- and between-subject methods (**paper**).
- Compared the representations learned by different CNN architectures to those of the human brain’s ventral visual stream. Found preliminary evidence that recurrent and deep residual CNNs learn more brain-like representations than feedforward models. Achieved state-of-the-art neural response prediction performance in late-stage visual areas (**thesis**).

Research Intern, Biomedical Data Science

Dartmouth College | Jan 2016 - June 2017

- Developed a CNN/LSTM ensemble for estimating high-risk substance use from Instagram data. Built pipelines that extracted word and sentence embeddings from captions and comments.
- Ensemble was able to estimate the risk of alcohol abuse, and found social media data characteristics associated with high-risk alcohol use.