Hayden Housen

haydenhousen.com
github.com/HHousen
linkedin.com/in/HHousen

hayden@haydenhousen.com • (845) 518-1380 • Ithaca, NY

Education

Cornell University — Computer Science, BS — Aug 2021 – May 2025

- GPA: 3.826, Dean's Honor List
- Coursework: Object-Oriented Programming and Data Structures, Discrete Structures, Linear Algebra, Probability and Statistics

Work Experience

Machine Learning Intern — Ada Support (Remote) — May 2022 – Aug 2022

- Conducted **>60 experiments** and trained >110 models to increase the accuracy of Ada's intent classification pipeline.
- Experimented with knowledge transfer, unsupervised learning of sentence embeddings, multi-task learning, and contrastive losses in the context of transformers and SVMs.
- My technique **improved production accuracy by 8%** using only 3% of production data.

Undergraduate Researcher — Cornell University AI — Sept 2021 – May 2022

- Overcame bias in paraphrase identification by using transformers & out-of-distribution detection techniques. Advised by Dr. Sernam Lim at Meta AI. Published in ICML 2022, Principles Of Distribution Shift Workshop.
- Experimented with zero-shot image classification via OpenAI's CLIP model.

Machine Learning Intern — Ada Support (Remote) — May 2021 – Aug 2021

- Led the discovery and experimentation phases of a project to enable Ada chatbots to better understand non-English languages.
- Wrote a data processing pipeline to efficiently clean and analyze **9 billion** chat messages for machine learning models.
- Researched novel techniques in multilingual intent prediction and cultivated skills in PyTorch, transformers, and pandas.

Projects

AI Lecture Notes Generation — lecture2notes — Sept 2019 – Jan 2022

- Conducted scientific research and created a state-of-the-art system to summarize classroom lectures using deep learning, computer vision, and NLP. Source on GitHub.
- Named a **top 300 scholar in the 2021 Regeneron Science Talent Search**, the nation's oldest and most prestigious science and math competition for high school seniors.
- Deployed ML pipeline in production via a full-stack website powered by Docker, Flask, Celery, Bootstrap, and Stripe.

Neural Summarization Library — TransformerSum — Mar 2020 – Oct 2020

- Furthered research in neural-network text summarization models with a focus on long document summarization.
- 4.45x smaller than the state-of-the-art model but 94% as accurate at release. 10+ pre-trained models available.
- Designed with code readability and thorough documentation as top priorities. **300+ stars on GitHub**.

AI Snow Day Prediction — Will I Have A Snow Day.com — Dec 2019 – Sept 2020

- Created an AI-powered automatic snow day predictor website that improves itself over time using user feedback. Powered by scikit-learn, Materialize.css, SendGrid, and Flask. Source on GitHub.
- Correlated **100GB+** of weather data from NOAA with school closings to select good features and train a gradient boosting classifier.

Cybersecurity Challenges — CTFs — Sept 2019 – Current

- Placed in **top 3%** on average in the PicoCTF 2019/2021/2022 competitions. Wrote detailed writeups to document my learning and help others.
- Continuously practicing ethical hacking skills by solving HackTheBox.com machines and publishing writeups.

Technologies and Languages

Languages	Python (& PyTorch), Java, JavaScript, HTML/CSS, SQL
Web	Flask, Bootstrap, jQuery, Materialize.css, web scraping
Other	API design, relational databases, MongoDB, Docker, Kali Linux, Metasploit