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, Rawlings Presidential Research Scholar
- 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

- **Improved Ada's production accuracy by 8%** using only 3% of production data by developing a novel intent classification pipeline.
- Conducted **>60 experiments** and trained >110 models to determine the most accurate methodology.
- Experimented with knowledge transfer, unsupervised learning of sentence embeddings, multi-task learning, and contrastive losses in the context of transformers and support vector machines.

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

- Overcame bias in paraphrase identification by using **transformers** & out-of-distribution detection techniques: "GAPX: Generalized Autoregressive Paraphrase-Identification X." Published in NeurIPS 2022 (3rd author). Advised by Dr. Sernam Lim at Meta AI.
- 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

- Created a state-of-the-art system to summarize classroom lectures using PyTorch, transformers (BERT), optical character recognition, speech to text, and convolutional neural networks. Source on GitHub. Learn more in the research paper.
- 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 with a focus on long document summarization. **310+ stars on GitHub**.
- 4.45x smaller than the state-of-the-art model but 94% as accurate at release. 10+ pre-trained models available.
- Rewrote researchers' code with enhanced performance and a focus on code readability and thorough documentation.

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 XGBoost, scikit-learn, Materialize.css, SendGrid, and Flask. Source on GitHub.
- Scraped school closings and reprocessed **100GB+** of weather data from NOAA to build a snow day dataset.
- Trained a gradient boosting classifier after extensive data exploration and feature engineering.

Cybersecurity Challenges — Capture The Flag — Sept 2019 – Current

- Placed in **top 3%** on average in the PicoCTF 2019/2021/2022 competitions. Learned ethical hacking skills including web exploitation, cryptography, reverse engineering, and binary exploitation. Worked with popular tools included in Kali Linux.
- Wrote technical guides with **over 94,000 views** to document my learning and help others.
- Continuously practicing cybersecurity principles by solving HackTheBox.com machines and publishing writeups.

Technologies and Languages

Languages	Python, Java, JavaScript, HTML/CSS, SQL, C
Machine Learning	PyTorch, transformers (BERT), scikit-learn, Lightning, pandas, OpenCV, Spacy, NumPy
Web	Flask, Bootstrap, jQuery, React, web scraping, API design
DB and DevOps	MongoDB, PostgreSQL, MySQL, Docker, AWS, CI/CD, Git