Emma Rafkin

emmarafkin.com• emmarafkin@gmail.com • (312) 636-9908

Education

Georgetown University — Washington, DC
M.S. Computational Linguistics. GPA 4.0 — 2023-2026
Master's Research Paper: Investigating the Cross-Lingual Transferability of NLP Tasks

Dartmouth College — Hanover, NH
B.A. Major: Computer Science, Minor: Linguistics. GPA 3.74, Cum Laude — 2017-2021

Experience

Johns Hopkins University Applied Physics Laboratory — Laurel, MD
Natural Language Processing Researcher — Aug 2021 – Present

- Created tools for U.S. Government that inject domain knowledge into language models and extract critical information from text using state of the art NLP methods and Domain Ontologies
- Awarded 8 grants to pursue research in AI transparency, concept erasure, summarization, entity disambiguation, LLM domain adaptation, AI red teaming, and multimodal anomaly detection
- Led research teams of 3-6 people, presented results to internal and external stakeholders

DALI Lab — Hanover, NH
Software Developer — Jan 2019 – Jun 2021

- Produced full-stack web and mobile applications for external stakeholders every 10 weeks
- Created and led classes on web design and development
- Developer lead, Women in Science Program lead and core executive board member for 2020-2021

Alarm.com — Tysons, VA
Software Engineer Intern — Jun 2020 – Aug 2020

- Created client-facing automated reports for property managers about all devices in their units

Cambly Inc. — San Francisco, CA
Software Engineer Intern — Jan 2020 – Mar 2020

- Enhanced Android application to include English as a Second Language courses for adults.

Dartmouth College Computer Science Department — Hanover, NH
Teaching Assistant — Sept 2018 – Dec 2018

- Graded coursework and tests for Object Oriented Programming course, taught weekly review class to 10-15 students

Featured Research Projects

Using Task Arithmetic for Cross-Lingual Transfer in LLMs: Meaning Transfers Better than Form: Explores the cross-lingual transfer of tasks in LLMS as distance increases between source and target languages. (Paper under review)
Task Arithmetic with Support Languages for Low-Resource ASR: Incorporates information from genetically related high-resource languages to improve performance on low-resource automatic speech recognition models. [System Paper], to be published in an ACL workshop.
CaTE: Guidance on data curation methods for trustworthy AI development. [Full Report] [Paper on domain knowledge elicitation]

Skills

ML/NLP: PyTorch, Tensorflow, Hugging Face, fine-tuning, LoRA, activation probing & steering, RAG
Software: Python, JavaScript, TypeScript, React, Java, Kotlin, SQL, ElasticSearch, MongoDB, Git