Jaden Fiotto-Kaufman

jadenfk@outlook.com ❖ (508) 505-6938 ❖ Somerville, MA ❖ github.com/JadenFiotto-Kaufman

My career spans the research, application, and enabling mechanistic analysis for deep learning models. This has resulted in several projects empowering researchers, users, and developers to further their understanding of AI.

WORK EXPERIENCE

National Deep Inference Fabric (Northeastern University) Boston, MA
Sr. Principal Research Software Engineer Oct, 2023 – Present

- Directed engineering efforts on a $9M NSF grant for democratizing research access to large pre-trained models.
- Designed and developed NNSight, an open-source library for interacting with model internals.
- Architected NDIF, an inference service enabling research groups to study models exceeding 400B parameters.
- Built and supervised a team, establishing practices to meet NSF review criteria and advisory board expectations.
- Deployed a suite of services across two providers (AWS, Delta) to support scalable and performant inference.
- Guided the creation of workbench, a no-code interface for exploratory analysis of AI.
- Managed partnerships with industry leaders [Anthropic, Goodfire, Transluce, NCSA]

Raytheon BBN Technologies Cambridge, MA
Senior Scientist Dec, 2019 – Oct, 2023

- Involved in a wide range of projects, primarily at the intersection of software development and AI research.
- Position was highly independent and involved working with multiple sub-contractors and universities.
- Worked on the DARPA SAIL-ON program to create an AI agent that adapts to and characterizes novelty.
- Designed and developed a system for the ASTARTE project to use deep learning for aircraft path prediction.
- Participated in the Explainable Artificial Intelligence (XAI) project and subsequent analysis.

Boston Digital Boston, MA
.NET Developer Sep, 2017 – Dec, 2019

- Developed Angular fronts end with C# .NET RESTful API's to display and access custom client content.
- Joined a budding .NET team and helped lay the groundwork for the development and CI/CD processes.

PUBLICATIONS & PRESENTATIONS

[1] R. Gandikota, J. Materzynska, J. Fiotto-Kaufman, and D. Bau, "Erasing concepts from diffusion models," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023.
[2] J. Fiotto-Kaufman, et al., "NNsight and NDIF: Democratizing access to open-weight foundation model internals," arXiv preprint arXiv:2407.14561, 2024.
[3] A. Mueller, A. Geiger, S. Wiegreffe, D. Arad, I. Arcuschin, A. Belfki, Y. S. Chan, J. Fiotto-Kaufman, et al., "MIB: A mechanistic interpretability benchmark," arXiv preprint arXiv:2504.13151, 2025.
[4] I. Grabe, J. Fiotto-Kaufman, et al, "Hidden layers: An interactive installation for exploring the neural semantics of image synthesis," in Adjunct Proc. 6th Decennial Aarhus Conf.: Computing X, 2025.
[5] W. Ferguson, D. Batra, R. Mooney, D. Parikh, A. Torralba, D. Bau, D. Diller, J. Fiotto-Kaufman, et al., "Reframing explanation as an interactive medium: The EQUAS (Explainable Question Answering System) project," Appl. AI Lett., vol. 2, no. 4, e60, 2021.

- NEMI 2025 – Talk regarding the NNsight and NDIF projects.
- CVPR AI Art Showcase – Demonstration of Patch Explorer, a tool for understanding diffusion model internals.
- NEMI 2024 – Talk regarding the NNsight and NDIF projects.

EDUCATION

Northeastern University (Khoury College of Computer Science) Boston, MA
MS, Artificial Intelligence April, 2023

- Master's thesis exploring semantic representation in Stable Diffusion for interpretability. 3.7/4.0 GPA
- Created an interactive web app demonstration for MEMIT, a technique to update the memories in LLMs.
- Courses: Artificial Intelligence, AI for Human Computer Interaction, Machine Learning, Pattern Recognition & Computer Vision, Programming Design Paradigm, Algorithms, Master's project & thesis

Wentworth Institute of Technology Boston, MA
BS, Computer Science (Minor in Mathematics) April, 2018

- Graduated Cum Laude with four times on the Dean's List 3.58/4.0 GPA
- Collaborated with researchers from Tufts in order classify if a subject was performing a trivial or strenuous task.
- Created a service to analyze Twitter accounts and their behavior to estimate the likelihood they were bots.

SKILLS, LANGUAGES & INTERESTS

- **Skills:** Mechanistic interpretability, machine learning operations, management, team lead, human computer interaction, diffusion models, LLMs, computer vision, full-stack development, PyTorch, RESTful API, Websockets, AWS, React, Vue, Huggingface, Git, Jira
- **Languages** *(in order of proficiency)***:** Python, English, C#, Java, JavaScript, C++, C, SQL, R, Assembly