ANELISE NEWMAN

(818)279-4468 ◇ anelisenewman@icloud.com ◇ AneliseNewman.com ◇ LinkedIn

SUMMARY

I build human-in-the-loop ML systems for evaluating and improving GenAI quality. I combine technical depth in data pipelines and platform infrastructure with experience designing human-facing experiments and extracting signals from noisy data. My recent work includes owning the data infrastructure for an evaluation platform supporting 40+ studies weekly, leading a zero-downtime platform productionization effort for a 10-person evaluation team, and partnering across product, research, and legal to build preference datasets from customer feedback. I work cross-functionally to turn human data into dependable signals for model improvement.

EXPERIENCE

Machine Learning Engineer, GenAI Evaluation, Adobe Firefly *October 2023 - Present*

- Led the productionization of Firefly's evaluation tooling into a scalable internal platform. Coordinated 10 interdiscplinary colleagues to deliver features, debug issues, migrate historical data, and transition evaluation workflows with 0 downtime. Our platform supports 40 studies and 500k annotations/week.
- Owned data pipelines, analysis, and results reporting for GenAI model evaluation. Presented model quality insights to senior leadership and built self-serve analysis endpoints that cut 10–20 hrs/week of team effort, reduced the codebase by 80%, and enabled flexible, on-demand results retrieval.
- Enabled use of customer feedback data for model improvement. Built pairwise post-training datasets through user analysis and advocated for cross-team logging changes that surfaced millions of daily signals.
- Designed and evaluated a RAG-based IP detection module for internal guardrails, defining reference-image curation, detector selection, and retrieval logic for IP violation detection.

Machine Learning Engineer, Stitch Fix *April 2022 - October 2023*

- Technical lead defining and executing the evaluation strategy for Stitch Fix's inventory purchasing model, a company-critical system driving millions of dollars in purchasing decisions.
- Set priorities for impactful evaluation metrics, built backtesting infrastructure, and surfaced the team's first daily model performance metrics. Implemented data-backed model improvements to unblock rollout.
- Trained computer vision models to predict the performance of new merchandise and productionized the company's first internal tooling for image-based deep learning.

PhD Student, Stanford Computer Science *September 2020 - March 2022*

Advisors: Nick Haber and Maneesh Agrawala

- Investigated methods for improving human–agent collaboration with minimal human data, training hundreds of reinforcement-learning agents and running a human-subject study on Overcooked.
- Found that population training fails to improve coordination with humans; led a research team spanning faculty, graduate, and undergraduate collaborators.

Researcher, MIT CSAIL *January 2019 - June 2020*

Advisor: Aude Oliva *Undergraduate Researcher September 2017-December 2018*

- Researched the intersection of computer vision and human perception, leading to peer-reviewed publications in top venues (ECCV, CVPR, CHI).
- Built a computational model of video memorability, deploying an interactive web platform to collect 10K human judgments and training a model to predict memory decay over time. [8]
- Developed novel datasets, interfaces, and models for human visual attention, including the first multi-duration saliency model and a toolbox of interfaces for crowdsourcing attention data. [2, 3]

Software Engineering Intern, Applied Intuition *June 2019 - August 2019*

AV Simulation Startup in Sunnyvale, CA

- Measured and reduced the domain gap between real and simulated data using current techniques in domain adaptation and image-to-image translation.
- Integrated neural networks and traditional computer vision techniques into production software.

Software Engineering Intern, Google *June 2018 - August 2018*
Kirkland, WA

- Implemented automatic message transcription for Duo, Google's video calling app (Android).

Software Engineering Intern, GrokStyle *June 2017 - August 2017*
Computer Vision Startup in San Francisco, CA (acquired by Facebook)

- Wrote a data ingestion SDK for clients, created a client-facing website to view uploaded data (Django), and built an analytics pipeline to monitor website interactions.

Front-End Development Intern, PlayStation *June 2016 - August 2016*
San Francisco, CA

- Took on responsibilities of a full-time developer building a social toolbar for PlayStation.com.

EDUCATION

MIT Computer Science, Master of Engineering *January 2019 - May 2020*
MIT Computer Science, Bachelor of Science *September 2015 - June 2019*

HONORS AND AWARDS

National Science Foundation Graduate Research Fellowship (NSF GRFP) *2020-2025*
Prestigious research fellowship awarded to 15% of applicants
1st place MIT MEng Thesis Award in AI and Decision Making *July 2021*
Top Masters of Engineering thesis in Artificial Intelligence and Decision Making
Robert M. Fano UROP (Undergraduate Research Opportunities) Award *May 2019*
For outstanding undergraduate research in computer science

PUBLICATIONS

1. **Newman, A.***, Fosco, C.*, Casser, V., McNamara, B., Lee, A., Oliva, A. "Multimodal Memorability: Modeling Effects of Semantics and Decay on Video Memorability." *ECCV*, 2020.
2. Fosco, C.*, **Newman, A.***, Sukhum, P., Zhang, Y.B., Zhao, N., Oliva, A., Bylinskii, Z. (2019) "How much time do you have? Modeling multi-duration saliency." *CVPR*, 2020.
3. **Newman, A.**, McNamara, B., Fosco, C., Zhang, Y.B., Sukhum, P., Tancik, M., Kim, N.W., Bylinskii, Z. "TurkEyes: A Web-Based Toolbox for Crowdsourcing Attention Data." In *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2020.
4. **Newman, A.***, Fosco, C.*, Casser, V.*, McNamara, B., Oliva, A. "To Decay or not to Decay: Modeling Video Memorability Over Time." SVRHM Workshop at *NeurIPS*, 2019.
5. Fosco, C.*, **Newman, A.***, Sukhum, P., Zhang, Y.B., Zhao, N., Oliva, A., Bylinskii, Z. (2019) "How many glances? Modeling Multi-duration Saliency." SVRHM Workshop at *NeurIPS*, 2019.
6. Bylinskii, Z., **Newman, A.**, Tancik, M., Madan, S., Durand, F., Oliva, A. "ZoomMaps: Using Zoom to Capture Areas of Interest on Images." *Journal of Vision*, 19. 149. 10.1167/19.10.149, 2019.
7. **Newman, A.**, Bylinskii, Z., Haroz, S., Madan, S., Durand, F., Oliva, A. "Effects of title wording on memory of trends in line graphs." *Journal of Vision*, 18. 837. 10.1167/18.10.837, 2018.

THESIS

8. **Newman, A.**, "Human-Computer Perception: Modeling Visual Perceptual Attributes". MIT MEng Thesis in Electrical Engineering and Computer Science. 2020.

SKILLS

Python, ML modeling, inference, and evaluation (PyTorch, vllm), Human-in-the-loop data systems (Prolific, Appen, MTurk, custom React UIs), Back-end systems (FastAPI, SQL, MongoDB, S3, Docker), Distributed data processing (Databricks/Spark, multiprocessing)