Eric Anthony Mitchell

eric.mitchell@cs.stanford.edu / ericmitchell.ai

Education

Stanford University, Stanford, CA   2019 - 2024 (Expected)
Ph.D. Candidate, Computer Science
Advisors: Chelsea Finn & Christopher D. Manning

Princeton University, Princeton, NJ   2014 - 2018
B.S.E., Computer Science (Highest Honors)
Advisor: H. Sebastian Seung

Professional Experience

DeepMind (AlphaCode team)   *London, England*
Research Scientist Intern [hosts: Junyoung Chung, Nate Kushman, Aäron van den Oord]   *2022*

Samsung Research America   *New York City, New York*
Research Intern, part-time   *2019 - 2020*
Machine Learning Research Engineer   *2018 - 2019*

Princeton Neuroscience Institute   *Princeton, New Jersey*
Research Engineer   *2018*

Google   *Mountain View, California*
Software Engineering Intern   *2017*

TapTools LLC   *Winston-Salem, North Carolina*
Founder, iOS Developer   *2012 - 2016*

Selected Works

*indicates equal contribution

Katherine Tian*, **Eric Mitchell***, Huaxiu Yao, Christopher D. Manning, Chelsea Finn.   2023
Fine-tuning Language Models for Factuality. *Preprint.*

Eric Mitchell, Rafael Rafailov, Archit Sharma, Chelsea Finn, Christopher D. Manning. An   2023
Emulator for Fine-Tuning Large Language Models using Small Language Models. *Preprint.*

Rafael Rafailov*, Archit Sharma*, **Eric Mitchell***, Stefano Ermon, Christopher D. Manning,   2023
Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model.
Oral Presentation (<1% of submissions), *Neural Information Processing Systems (NeurIPS).*

Nathan Hu*, **Eric Mitchell***, Christopher D. Manning, Chelsea Finn. Meta-Learning Online   2023
Adaptation of Language Models. *Empirical Methods in Natural Language Processing (EMNLP).*

Katherine Tian*, **Eric Mitchell***, Allan Zhou, Archit Sharma, Rafael Rafailov, Huaxiu Yao,   2023
Chelsea Finn, Christopher D. Manning. Just Ask for Calibration: Strategies for Eliciting
Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback.
Empirical Methods in Natural Language Processing (EMNLP).

Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, Chelsea Finn.   2023
DetectGPT: DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability
Curvature. **Oral Presentation (2% of submissions)**, *International Conference on Machine
Learning (ICML).*

Peter Henderson*, **Eric Mitchell***, Christopher D. Manning, Dan Jurafsky, Chelsea Finn.   2023
Self-Destructing Models: Increasing the Costs of Harmful Dual Uses in Foundation Models.
Honorable Mention, Best Student Paper. *AAAI/ACM Conference on Artificial Intelligence,
Ethics, & Society.*

Eric Mitchell, Joseph J Noh, Siyan Li, William S Armstrong, Ananth Agarwal, Patrick Liu, Chelsea Finn, Christopher D Manning. Enhancing Self-Consistency and Performance of Pretrained Language Models with NLI. **Oral Presentation (4% of submissions)**, *Empirical Methods in Natural Language Processing (EMNLP)*.　2022

Eric Mitchell, Charles Lin, Antoine Bosselut, Christopher D. Manning, Chelsea Finn. Memory-Based Model Editing at Scale. *International Conference on Machine Learning (ICML)*.　2022

Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, Christopher D. Manning. Fast Model Editing at Scale. *International Conference on Learning Representations (ICLR)*.　2022

Eric Mitchell, Rafael Rafailov, Xue Bin Peng, Sergey Levine, and Chelsea Finn. Offline Meta-Reinforcement Learning with Advantage Weighting. *International Conference on Machine Learning (ICML)*.　2021

Eric Mitchell, Selim Engin, Volkan Isler, Daniel D Lee. Higher-Order Function Networks for Learning Composable 3D Object Representations. *International Conference on Learning Representations (ICLR)*.　2020

Tarik Tosun*, **Eric Mitchell***, Ben Eisner, Jinwook Huh, Bhoram Lee, Daewon Lee, Volkan Isler, H. Sebastian Seung, Daniel D Lee. Pixels to Plans: Learning Non-Prehensile Manipulation by Imitating a Planner. *IEEE International Conference on Intelligent Robots and Systems (IROS)*.　2019

Academic Awards

Stanford Knight-Hennessy Graduate Fellowship	2019 - 2022
Princeton CS Outstanding Undergraduate Thesis Prize	2018
Princeton CS Best Undergraduate Research Poster Prize	2017
Srixon/Cleveland Golf All-America Scholar	2017
Tau Beta Pi Engineering Honor Society Vice President	2016 - 2017
Princeton Shapiro Prize for Academic Excellence	2016

Service

NeurIPS Socially Responsible Language Modeling Workshop Reviewer	2023
NeurIPS Conference Reviewer	2020 - 2021, 2023
Stanford CS Ph.D. Admissions Committee Student Representative	2020 - 2022
ICLR Conference Reviewer	2020 - 2021
NeurIPS Meta-Learning Workshop Reviewer	2020 - 2021
AI4ALL Summer Camp Instructor	2018

Other Activities

PADI Open Water SCUBA certification	2023
Princeton NCAA Division I Varsity Golf Team	2014 - 2018
HackPrinceton (1st Place Team, Software)	2015
Guitar, Singing, Songwriting	