Annie S. Chen

asc8@stanford.edu
https://anniesch.github.io/

EDUCATION

Ph.D. in Computer Science, Stanford University *Sept. 2021 - Present*
Advisor: Chelsea Finn
Objective: Advance the capabilities and reliability of foundation models through research in reinforcement learning, data curation, and adaptive test-time reasoning.

M.S. in Computer Science, Stanford University *Sept. 2020 - June 2021*
Artificial Intelligence Specialization, GPA: 4.0

B.S. in Mathematics, Stanford University *Sept. 2017 - June 2021*
With Distinction

FUNDING & AWARDS

OpenAI Superalignment Fellowship (2024-2025)
NSF Graduate Research Fellowship (2021-2024)
Finalist, Citadel GQS PhD Fellowship (April 2024)
HAI Google Cloud Credit Grant (April 2024)
Stanford Community Impact Award (May 2024)
Microsoft Accelerate Foundation Models Grant (Sept. 2023)
HAI Google Cloud Credit Grant (Nov. 2022)
Women in Machine Learning (WiML) Travel Funding Award (Nov. 2022)
Stanford Mathematics Distinguished Service Award (June 2021)
University Distinction, top 15% graduating class, Stanford University (June 2021)
Honorable Mention, Computing Research Association (CRA) Undergraduate Researcher Award (Dec. 2020)
Outstanding Poster Presentation, Joint Mathematics Meetings (JMM) Undergraduate Poster Session (Jan. 2019)
J. Perry Bartlett STEM Scholarship (2017-2021), $40k for undergraduate studies

PREPRINTS & PUBLICATIONS

* denotes equal contribution

[21] Perry Dong*, Alec M. Lessing*, **Annie S. Chen***, Chelsea Finn. Reinforcement Learning via Implicit Imitation Guidance. *In Submission*, 2025.

[20] **Annie S. Chen**, Philemon Brakel, Antonia Bronars, Annie Xie, Sandy Huang, Oliver Groth, Maria Bauza, Markus Wulfmeier, Nicolas Heess, Dushyant Rao. Exploiting Policy Idling for Dexterous Manipulation. *International Conference on Intelligent Robots and Systems (IROS)*, 2025.

[19] **Annie S. Chen***, Alec M. Lessing*, Yuejiang Liu, Chelsea Finn. Curating Demonstrations with Online Experience. *Robotics: Science and Systems (RSS)*, 2025.

[18] **Annie S. Chen***, Alec M. Lessing*, Andy Tang*, Govind Chada*, Laura Smith, Sergey Levine, Chelsea Finn. Commonsense Reasoning for Legged Robot Adaptation with Vision-Language Models. *International Conference on Robotics and Automation (ICRA)*, 2025.

[17] Kaustubh Mani, Charlie Gauthier, Vincent Mai, **Annie S. Chen**, Samer B.

Nashed, Liam Paull. Safety Representations for Safer Policy Learning. *International Conference on Learning Representations (ICLR)*, 2025.

[16] **Annie S. Chen***, Govind Chada*, Laura Smith, Archit Sharma, Zipeng Fu, Sergey Levine, Chelsea Finn. Adapt On-the-Go: Behavior Modulation for Single-Life Robot Deployment. *Conference on Lifelong Learning Agents (CoLLAs), Oral*, 2025.

[15] Johnathan Xie*, **Annie S. Chen***, Yoonho Lee, Eric Mitchell, Chelsea Finn. Calibrating Fine-Tuned Language Models via Adaptive Temperature Scaling. *Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[14] Moritz Stephan, Alexander Khazatsky, Eric Mitchell, **Annie S. Chen**, Sheryl Hsu, Archit Sharma, Chelsea Finn. RLVF: Learning from Verbal Feedback without Overgeneralization. *International Conference on Machine Learning (ICML)*, 2024.

[13] **Annie S. Chen***, Yoonho Lee*, Amrith Setlur, Sergey Levine, Chelsea Finn. Project and Probe: Sample-Efficient Domain Adaptation by Interpolating Orthogonal Features. *International Conference on Learning Representations (ICLR), Spotlight (top 5%)*, 2024.

[12] Johnathan Xie, Yoonho Lee, **Annie S. Chen**, Chelsea Finn. Self-Guided Masked Autoencoders for Domain-Agnostic Self-Supervised Learning. *International Conference on Learning Representations (ICLR)*, 2024.

[11] Caroline Choi*, Yoonho Lee*, **Annie S. Chen**, Allan Zhou, Aditi Raghunathan, Chelsea Finn. AutoFT: Robust Fine-Tuning by Optimizing Hyperparameters on OOD Data. *NeurIPS DistShift Workshop*, 2023.

[10] **Annie S. Chen**, Yoonho Lee, Amrith Setlur, Sergey Levine, Chelsea Finn. Confidence-Based Model Selection: When to Take Shortcuts for Subpopulation Shifts. *NeurIPS DistShift Workshop*, 2023.

[9] Siddharth Karamcheti, Suraj Nair, **Annie S. Chen**, Thomas Kollar, Chelsea Finn, Dorsa Sadigh, Percy Liang. Language-Driven Representation Learning for Robotics. *Robotics: Science and Systems (RSS). Best Paper Finalist*, 2023.

[8] Yoonho Lee*, **Annie S. Chen***, Fahim Tajwar, Ananya Kumar, Huaxiu Yao, Percy Liang, Chelsea Finn. Surgical Fine-Tuning Improves Adaptation to Distribution Shifts. *International Conference on Learning Representations (ICLR)*, 2023.

[7] **Annie S. Chen**, Archit Sharma, Sergey Levine, Chelsea Finn. You Only Live Once: Single-Life Reinforcement Learning. *Neural Information Processing Systems (NeurIPS)*, 2022.

[6] **Annie S. Chen**, Suraj Nair, Chelsea Finn. Learning Generalizable Robotic Reward Functions from "In-The-Wild" Human Videos. *Robotics: Science and Systems (RSS)*, 2021.

[5] Evan Z. Liu*, Behzad Haghgoo*, **Annie S. Chen***, Aditi Raghunathan, Pang Wei Koh, Shiori Sagawa, Percy Liang, Chelsea Finn. Just Train Twice: Improving Group Robustness without Training Group Information. *International Conference on Machine Learning (ICML)*, Long Oral (top 3%), 2021.

[4] **Annie S. Chen***, Hyunji Nam*, Suraj Nair*, Chelsea Finn. Batch Exploration with Examples for Scalable Robotic Reinforcement Learning. *International Conference on Robotics and Automation (ICRA)*, 2021.

[3] Rishi Bommasani, ..., **Annie S. Chen**, ..., Percy Liang. On the Opportunities and Risks of Foundation Models. *Report by the Center for Research on Foundation Models (CRFM)*, 2021.

[2] Bryce Cai*, **Annie S. Chen***, Ben Heller*, Eyob Tsegaye*. Limit Theorems for Descents in Permutations and Arithmetic Progressions in Z/pZ. *Outstanding Poster Presentation, Joint Mathematics Meetings Undergraduate Poster Session*, 2019.

[1] **Annie S. Chen**, T. Alden Gassert, Katherine E. Stange. Index Divisibility in Dynamical Sequences and Cyclic Orbits Modulo p. *New York Journal of Mathematics* 23, 1045-1063, 2017.

EXPERIENCE

Student Researcher *Sept. 2024 - March 2025*
Google DeepMind (London, UK)

- Hosted by Philemon Brakel and Dushyant Rao, working on self-improvement for robot behavior models

CS Researcher *Sept. 2019 - Present*
Stanford Artificial Intelligence Laboratory (SAIL), Computer Science Dept.

- Advised by Chelsea Finn.
- Developing scalable methods that improve model adaptability and robustness to distribution shifts.

Research Intern, Student Researcher *June 2021 - Jan. 2022*
Brain Robotics, Google, Inc.

- Mentored by Pete Florence and Andy Zeng
- Worked on improving action-conditioned visual dynamics models.

CS Research Assistant *Sept. 2018 - Apr. 2019*
Stanford AI for Human Impact Lab, Computer Science Dept.

- Advised by Prof. Emma Brunskill
- Studied the impact of shared learning autonomy systems in practice.

Math Researcher *June 2018 - Aug. 2018*
Stanford Undergraduate Research Institute in Math (SURIM)

- Mentored by Felipe Hernandez
- Improved bounds on the rate of convergence of various random variables locally to a Gaussian distribution.

INVITED TALKS

CVPR 3D VLMs for Robotics Manipulation *June 2025*
Post-training Robotic Foundation Models (joint with Chelsea Finn)

Nuro *May 2025*
Curating Demonstrations using Online Experience

Google DeepMind Robotics *December 2024*
Towards More Robust and Generalizable Robot Learning

Citadel Global Quantitative Strategies *April 2024*
Robustness and Adaptation Methods for Reliable Model Deployment

Stanford Robotics Seminar *January 2024*
Single-Life Robot Deployment: Adapting On-the-Fly to Novel Scenarios

MosaicML *August 2023*
Surgical Fine-Tuning Improves Adaptation to Distribution Shifts

ICLR TrustML Unlimited Workshop *May 2023*
Project and Probe: Sample-Efficient Domain Adaptation by Interpolating Orthogonal Features

ML Collective DLCT *March 2023*
Project and Probe: Sample-Efficient Domain Adaptation by Interpolating Orthogonal Features

NeurIPS Women in Machine Learning (WiML) Workshop *Nov. 2022*
You Only Live Once: Single-Life Reinforcement Learning

ICLR Self-Supervised Reinforcement Learning Workshop *May 2021*
Learning Generalizable Robotic Reward Functions from "In-The-Wild" Human Videos.

INVOLVEMENT

Course Assistant, CS224R (Deep Reinforcement Learning) *Spr. 2025*

Community Associate *Apr. 2023 - Aug. 2024*
Stanford Graduate Life Office

- Hold multiple events per quarter to foster community engagement and provide support to dorm residents

Mentoring Program Co-Organizer *Oct. 2021 - Aug. 2024*
Stanford CS Undergraduate Mentoring Program

- Match undergraduate students with graduate student mentors, aimed at increasing the participation of underrepresented minorities in CS

Peer Advisor *Sept. 2019 - June 2021*
Stanford University Mathematics Department

- Held weekly office hours to advise math majors and organized quarterly events

Teaching Assistant *Sept. 2018 - Dec. 2019*
Euler Math Circle

- Taught advanced proof techniques and topics, such as number theory, combinatorics, and analysis, to advanced middle and high school students

Board Member *Oct. 2017 - May 2020*
Stanford Women in Computer Science (WiCS)

- Co-led outreach workshops introducing coding to middle school girls

MENTORING

Undergraduate Students:
Alec Lessing, Oct. 2023 - Present
Andy Tang, Oct. 2023 - June 2024

Johnathan Xie, Jan. 2023 - Feb. 2024
Govind Chada, Sept. 2022 - June 2024

MISC.

Reviewing:
IEEE Robotics and Automation Letters (RA-L): 2021
International Conference on Learning Representations (ICLR): 2022, 2023, 2024
International Conference on Machine Learning (ICML): 2024
Conference on Robot Learning (CoRL): 2022, 2023, 2024
Neural Information Processing Systems (NeurIPS): 2022, 2023, 2025