Peiyang Song

1200 E California Blvd, Pasadena, CA
✉ psong@caltech.edu
🖥 peiyang-song.github.io

Education

6/2026 (expected)	**California Institute of Technology**	Pasadena, CA

B.S. in Computer Science & Minor in Robotics

Advisors: Prof. Steven Low & Prof. Günter Niemeyer.

GPA: **4.2/4.0** (some courses occasionally offer A+ grade = 4.3)

Research Interests

My research focuses on **LLM reasoning**, **agentic AI**, and **neuro-symbolic AI**. I aim to advance intelligent agents capable of rigorous reasoning by combining the strengths of neural and symbolic paradigms, with one central theme and two adjacent directions:

- **Central theme:** Integrating LLM reasoning, agentic AI, and neuro-symbolic methods by combining *neural* models (LLMs) with *symbolic* systems (e.g., Lean) to advance LLM-based agents for formal reasoning in mathematics and code.

- **Broader LLM reasoning:** Extending from formal to informal reasoning, exploring how LLMs can better handle reasoning in natural language contexts, informed by *cognitive science* principles and studies of *human-like reasoning* processes.

- **Broader neuro-symbolic AI:** Applying neuro-symbolic approaches to broader AI challenges beyond formal reasoning, including the design of energy-efficient machine learning systems.

Research Experience

6/2025 – Present	**University of California, Berkeley**	Berkeley, CA

Researcher @ Berkeley AI Research (BAIR) Lab and Sunblaze Group

Advisors: Prof. Dawn Song (UCB), Dr. Jingxuan He (UCB)

Directions: Verifiable Code Agents, Generalizable LLM Reasoning

6/2024 – Present	**Stanford University**	Palo Alto, CA

Researcher @ Stanford AI Lab (SAIL) and Computation & Cognition Lab

Advisors: Prof. Noah Goodman (Stanford), Dr. Gabriel Poesia (Harvard)

Directions: AI-Assisted Formal Conjecturing, Cognition-Inspired LLM Reasoning

2/2023 – 2/2025	**California Institute of Technology**	Pasadena, CA

Research Fellow @ Anima AI+Science Lab

Advisors: Prof. Anima Anandkumar (Caltech), Dr. Kaiyu Yang (Meta)

Directions: Neural Theorem Proving, Prover Agents, Neuro-Symbolic Reasoning

11/2022 – 6/2024 **University of California, Santa Barbara**　　　　　　　Santa Barbara, CA

Researcher @ Computer Architecture Lab (ArchLab)

Advisors: Prof. Timothy Sherwood (UCSB), Dr. Jeremy Lau (Google)

Directions: Energy-Efficient ML Systems, Temporal Logic, Neuro-Symbolic AI

Selected Publications & Preprints

Refereed Publications

TMLR 2026 **Large Language Model Reasoning Failures**

Peiyang Song*, Pengrui Han*, Noah Goodman (* Equal Contribution)

Transactions on Machine Learning Research (TMLR), 2026, **Survey Certification**

NeuS 2025 **Lean Copilot: Large Language Models as Copilots for Theorem Proving in Lean**

Peiyang Song, Kaiyu Yang, Anima Anandkumar

Proceedings of the International Conference on Neuro-symbolic Systems, PMLR 288:144-169, 2025, **1.2k+ stars on Github, ranking 2nd after Mathlib4 among all Lean projects**

IEEE Micro 2025 **Delay Space Arithmetic and Architecture**

Rhys Gretsch, Peiyang Song, Advait Madhavan, Jeremy Lau, Timothy Sherwood

IEEE Micro 45, NO. 04 (2025): 87-94, **Top Pick Award**

ICLR 2025 **LeanAgent: Lifelong Learning for Formal Theorem Proving**

Adarsh Kumarappan*, Mo Tiwari*, Peiyang Song, Robert Joseph George, Chaowei Xiao, Anima Anandkumar

International Conference on Representation Learning, ICLR 73525-73564, 2025

TMLR 2025 **LeanProgress: Guiding Search for Neural Theorem Proving via Proof Progress Prediction**

Suozhi Huang, Peiyang Song, Robert Joseph George, Anima Anandkumar

Transactions on Machine Learning Research (TMLR), 2025

EMNLP 2024 **In-Context Learning May Not Elicit Trustworthy Reasoning: A-Not-B Errors in Pretrained Language Models**

Pengrui Han*, Peiyang Song*, Haofei Yu, Jiaxuan You (* Equal Contribution)

Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 5624-5643. 2024

EMNLP 2024   **Creative and Context-Aware Translation of East Asian Idioms with GPT-4**

Kenan Tang*, Peiyang Song*, Yao Qin, Xifeng Yan (* Equal Contribution)

Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 9285-9305. 2024

ASPLOS 2024   **Energy Efficient Convolution with Temporal Arithmetic**

Rhys Gretsch, Peiyang Song, Advait Madhavan, Jeremy Lau, Timothy Sherwood

Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2, pp. 354-368. 2024

NeurIPS 2023   **LeanDojo: Theorem Proving with Retrieval-Augmented Language Models**

Kaiyu Yang, Aidan Swope, Alex Gu, Rahul Chalamala, Peiyang Song, Shixing Yu, Saad Godil, Ryan Prenger, Anima Anandkumar

Advances in Neural Information Processing Systems 36 (2023): 21573-21612, **Oral Presentation**

Preprints

Preprint   **How and Why LLMs Generalize: A Fine-Grained Analysis of LLM Reasoning from Cognitive Behaviors to Low-Level Patterns**

Haoyue Bai*, Yiyou Sun*, Wenjie Hu, Shi Qiu, Maggie Ziyu Huan, Peiyang Song, Robert Nowak, Dawn Song (* Equal Contribution)

Preprint, 2025

Preprint   **Adaptation of Agentic AI**

Pengcheng Jiang*, Jiacheng Lin*, Zhiyi Shi*, Zifeng Wang, Luxi He, Yichen Wu, Ming Zhong, Peiyang Song, Qizheng Zhang, Heng Wang, Xueqiang Xu, Hanwen Xu, Pengrui Han, Dylan Zhang, Jiashuo Sun, Chaoqi Yang, Kun Qian, Tian Wang, Changran Hu, Manling Li, Quanzheng Li, Hao Peng, Sheng Wang, Jingbo Shang, Chao Zhang, Jiaxuan You, Liyuan Liu, Pan Lu, Yu Zhang, Heng Ji, Yejin Choi, Dawn Song, Jimeng Sun, Jiawei Han (* Equal Contribution)

Preprint, 2025

Preprint   **The Personality Illusion: Revealing Dissociation Between Self-Reports & Behavior in LLMs**

Pengrui Han*, Rafal D. Kocielnik*, Peiyang Song, Ramit Debnath, Dean Mobbs, Anima Anandkumar, R. Michael Alvarez (* Equal Contribution)

NeurIPS LAW Workshop: Bridging Language, Agent, and World Models, 2025, **Oral Presentation + Best Paper Honorable Mention Award**; *NeurIPS Workshop on LLM Persona Modeling, 2025*, **Oral Presentation**

Preprint **AI Impact on Human Proof Formalization Workflows**

Katherine M. Collins*, Simon Frieder*, Jonas Bayer, Jacob Loader, Jeck Lim, Peiyang Song, Fabian Zasier, Lexin Zhou, Shanda Li, Sam Looi, Jose Hernandez-Orallo, Joshua B. Tenenbaum, Cameron Freer, Umang Bhatt, Adrian Weller, Valerie Chen[†], Ilia Sucholutsky[†] (* Equal Contribution, [†] Equal Advising)

NeurIPS Workshop on Mathematical Reasoning and AI (MATH-AI), 2025

Preprint **Energy-Aware Temporal Function Approximation**

Peiyang Song, Rhys Gretsch, Jeremy Lau, and Timothy Sherwood

In Submission, Manuscript Available upon Request

Selected Awards

1/2026 **Survey Certification @ TMLR**

For paper "Large Language Model Reasoning Failures"

12/2025 **Best Paper Honorable Mention Award @ NeurIPS LAW Workshop**

For paper "The Personality Illusion: Revealing Dissociation Between Self-Reports & Behavior in LLMs"

10/2025 **Caltech FCC Appreciation Award**

In recognition of outstanding service in mentoring first-year Caltech students

5/2025 **ICLR Notable Reviewer Award**

In recognition of outstanding review service at ICLR 2025

4/2025 **George W. Housner Student Discovery Fund**

For paper "Lean Copilot: Large Language Models as Copilots for Theorem Proving in Lean"

2/2025 **IEEE Micro Top Pick Award**

For paper "Energy Efficient Convolutions with Temporal Arithmetic"

8/2023 **Early Research Scholarship**

In recognition of research work done in early undergraduate study

4/2023 **Caltech SURF Award**

For research work on machine learning for theorem proving in Lean

Selected Media

2025 **New Framework Simplifies the Complex Landscape of Agentic AI**
VentureBeat

2025 **This AI Paper Explains Why Most "Agentic AI" Systems Feel Impressive in Demos and then Completely Fall Apart in Real Use**

MarkTechPost

2025 **Researchers Discover "Personality Illusion" to Reveal a Profound Disconnect Between Language and Behavior in LLMs**

MIT Technology Review China

2024 **Mathematicians' Newest Assistants Are Artificially Intelligent**

Scientific American

2024 **LeanAgent: The First Life-Long Learning Agent for Formal Theorem Proving in Lean**

MarkTechPost

2024 **Lean Copilot: An AI Tool That Allows Large Language Models (LLMs) to Be Used in Lean for Proof Automation**

MarkTechPost

2023 **Can LLMs Generate Mathematical Proofs That Can Be Rigorously Checked?**

MarkTechPost

Invited Talks & Tutorials

LLM Reasoning for Math and Code

10/2025 Carnegie Mellon University L3 Lab

Tutorial: Neuro-Symbolic Theorem Proving with Lean

9/2024 3rd Neuro-Symbolic AI Summer School (NSSS)

Towards An AI Mathematician

12/2023 UC Santa Barbara NLP Lab

11/2023 CCS Research & Creative Activities Conference (RACA-CON)

8/2023 Caltech SURF Seminar Day

Teaching Experience

Spring 2026 **CS 38: Algorithms**

Teaching Assistant @ *California Institute of Technology*

Winter 2026 **ME/CS/EE 133B: Robotics – Planning**

Teaching Assistant @ *California Institute of Technology*

Fall 2025 **ME/CS/EE 133A: Robotics – Kinematics**

Teaching Assistant @ *California Institute of Technology*

Academic Services

Reviewer Conference on Neural Information Processing Systems (NeurIPS)
International Conference on Learning Representations (ICLR)
Conference on Language Modeling (COLM)
Association for Computational Linguistics Rolling Review (ARR)
Annual Meeting of the Association for Computational Linguistics (ACL)
Conference on Empirical Methods in Natural Language Processing (EMNLP)
International Joint Conference on Natural Language Processing (IJCNLP)
Asia-Pacific Chapter of the Association for Computational Linguistics (AACL)

NeurIPS Mathematical Reasoning and AI (MATH-AI) Workshop
NeurIPS Workshop on Deep Learning for Code (DL4C)
NeurIPS Workshop on Behavioral Machine Learning
ICLR VerifAI: AI Verification in the Wild Workshop
ICLR Workshop on Representational Alignment (Re-Align)
ICML AI for Math (AI4MATH) Workshop
ICML Workshop on LLMs and Cognition (LLM-Cognition)
ICML Workshop on Assessing World Models
ICML Workshop on on Models of Human Feedback for AI Alignment (MoFA)

Organizing Staff Agentic AI Summit 2025 @ UC Berkeley

On-Campus Services & Appointments

Caltech Admissions Ambassador @ Caltech Undergraduate Admissions Office
First-Year Caltech Connector (FCC) @ Student & Family Engagement Office

Languages

Programming Python, C++, Lean 4, Java, C, PASCAL, Rocq, OCaml, C#

Natural English (TOEFL 117/120), Chinese (Native)