Nicholas Bai

nbai@caltech.edu | (858) 790-2418 | github.com/nick01as

EDUCATION

California Institute of Technology, Pasadena, CA — *Expected June 2029*
B.S. in Computer Science — GPA: 3.8/4.0

RESEARCH EXPERIENCE

Undergraduate Researcher, Anima AI + Science Lab (Prof. Anima Anandkumar) — Dec. 2025 – Present
California Institute of Technology — Carl F. Braun SURF Fellow (2026) — *Pasadena, CA*

- Design diffusion-prior and particle-based samplers for Bayesian inverse PDE problems, recovering full multimodal posteriors where standard maximum-a-posteriori methods collapse to a single mode.

Independent Student Researcher — June 2025 – Present
California Institute of Technology — *Pasadena, CA*

- Analyzed geometric representations and failure modes of counting tasks in large language models, improving high-count accuracy by up to 20%; **first-authored** paper presented at the **ICML 2026 Workshop** on Compositional Learning.
- Built curvature-adaptive transformers with token-wise mixture-of-experts routing across three geometric attention branches, yielding a 10% improvement in Hits@10 compared to fixed-geometry baselines; co-authored paper presented at the **NeurIPS 2025 Workshop** on Non-Euclidean Foundation Models and Geometric Learning.

Research Intern, Trustworthy AI Lab (Prof. Tsui-Wei Weng) — May 2023 – Dec. 2025
University of California San Diego — *San Diego, CA*

- **First-authored** a language-model-based framework for explaining individual neurons in vision networks, preferred over prior state-of-the-art baselines by more than 2× in human evaluations; published in **TMLR (2025)**, selected as an **ICML 2024 Mechanistic Interpretability Workshop Spotlight Paper**.
- Devised concept-vector methods for controllable diffusion-model editing, improving edit fidelity by ∼11% while cutting training time by more than 20×.

SELECTED PUBLICATIONS

N. Bai, A. Mehrotra. "Count Me If You Can: Geometric Failure Modes in Language Model Counting." *ICML 2026 Workshop on Compositional Learning: Safety, Interpretability and Agents.*

N. Bai, et al. "Interpreting Neurons in Deep Vision Networks with Language Models." *TMLR*, 2025. Extended version of "Describe-and-Dissect," *ICML 2024 Mechanistic Interpretability Workshop*, **Spotlight**.

R. Y. Lin, S. Ojha, **N. Bai**. "CAT: Curvature-Adaptive Transformers for Geometry-Aware Learning." *NeurIPS 2025 Workshop on Non-Euclidean Foundation Models and Geometric Learning.*

LEADERSHIP & ACTIVITIES

Co-Director, SPARC Labs — May 2026 – Present
California Institute of Technology — *Pasadena, CA*

- Co-founded Caltech's interdisciplinary AI research club; organize speaker events and coordinate student-led research groups, with members presenting work at **seven ICML 2026 and ICRA 2026 workshops**.

Organizer, Hacktech — Fall 2025 – Present
California Institute of Technology — *Pasadena, CA*

- Direct university outreach and corporate sponsorships for Caltech's flagship 2026 hackathon.

NCAA Division III Men's Swimming & Diving, Caltech — *2025 – Present*
Team **Newcomer of the Year**, 2025–26 season.

SELECTED HONORS

- **National Merit Finalist** — $2,500 NVIDIA Corporate Scholarship.
- **USA Computing Olympiad, Platinum Division** — top 350 competitors nationally.
- **Overall 1st Place, UCSD TritonHacks 2023** — ML tool forecasting nighttime stargazing conditions.

TECHNICAL SKILLS

Languages: Python, C/C++, Java, MATLAB **Tools & Frameworks:** PyTorch, NumPy, SciPy, Linux, Git, LaTeX