# Danielle Yang

dsyang@caltech.edu | danielleyang.com | github.com/dsyang1219 | linkedin.com/in/danielle-yang-0013b82a0 | Pasadena, CA

## Education

**California Institute of Technology** Pasadena, CA
B.S. in Applied and Computational Mathematics. GPA: 3.8 Expected June 2029
**Relevant Coursework:** Machine Learning and Data Mining (CS 155), Applied Linear Algebra (ACM/IDS 104), Discrete Mathematics (Ma 6a), Differential Equations (Ma 2), Data Structures and Systems Programming in C (CS 1 to 3), Numerical Computing in MATLAB (ACM 11)
**Winter 2027:** Probability and Statistics (Ma 3), Methods of Applied Mathematics (ACM 95a), Discrete Mathematics II (Ma 6b)

## Experience

**Linde Center for Science, Society, and Policy, Caltech** Pasadena, CA
**SURF Fellow (2026 Citadel Global Fixed Income SURF Fellow), Voting Technology Project, advised by Prof. R. Michael Alvarez** July 2026 – Present

- Built an NLP pipeline over 194 political podcasts (31,191 episodes): batched faster-whisper transcription on lab GPU servers, LLM-based ad removal, and segmentation into passages.
- Modeled host ideology from transcripts with LDA topic models, keyATM and sentence embeddings; recovered ideology at $R^2 \approx 0.46$ against campaign-finance scores and showed it is a secondary axis driven by topic choice.
- Linked show-level language measures to Kettering-Gallup and Pew survey data; found that direct address to listeners predicts less-educated audiences (0.21 SD lower), replicated on two Pew surveys and Apple reviews.
- Developing an embedding-based method for faster speaker diarization, to separate host speech from guest speech as the corpus expands to 50,000 episodes.
- Co-authored the final SURF report, “Who Are They Talking To? Listener-Directed Address and the Audiences of Political Podcasts,” and present results at weekly lab meetings.
- Finalist in Caltech’s Perpall SURF Speaking Competition (ongoing), the biggest competition for scientific communication at Caltech.

**Observatory on Social Media, Indiana University** Bloomington, IN
**Student Researcher** August 2024 – Present

- Co-developed Prefill-Guided Thinking (PGT), which prefills a vision-language model’s response with a task-aligned phrase to guide its reasoning in zero-shot detection of AI-generated images.
- Improved Macro F1 by up to 24% across three open-source vision-language models on three benchmarks covering 16 image generators.
- Showed that prefills counteract early overconfidence in some models, akin to mitigating the Dunning-Kruger effect, which leads to better detection.
- Co-authored the paper, accepted as one of seven oral presentations at the NeurIPS 2025 Workshop on Generative and Protective AI for Content Creation; submitted to ACL Rolling Review.

**Storm Shield LLC**
**Co-founder and Technical Lead; Samsung Solve for Tomorrow National Winner and Global Ambassador** August 2024 – Present

- Co-founded an assistive wearable for hard-of-hearing athletes: a headband, mobile app and sensor circuit that protects hearing aids from sweat, rain and contact and gives directional vibration alerts for spatial awareness.
- Led technical development from an Arduino and Python prototype to second-generation embedded firmware in C on STM32, with a Bluetooth companion app in progress.
- Selected as 1 of 3 national winners of Samsung Solve for Tomorrow (3,000+ teams, $125,000 award) and 1 of 10 global ambassadors among 70+ national winners.
- Built the next prototype with Samsung engineers and presented it at the 2026 Winter Olympics to 80+ stakeholders, including the IOC president. Featured on PBS NewsHour.
- Working with ImpactSquare and the International Olympic Committee in a pilot phase through 2028 that provides mentoring and funding to develop the product further.

**Cambridge Center for International Research** Cambridge, UK
**Future Scholar, Data-Driven Astronomy** August 2023 – April 2024

- First-authored a paper on real-time classification of astronomical transients with a recurrent autoencoder and gradient-boosted classifier (ROC-AUC 0.94), published in the Stanford Undergraduate Research Journal.
- Presented at the International Conference on Machine Learning in Astrophysics; admitted to a program with a roughly 15% acceptance rate.

## Leadership and Activities

- **Caltech Varsity Track and Field:** hurdler; school record holder in the 4x200m relay and top 10 all-time in school history.
- **Caltech Student Investment Fund:** member.
- **Fleming House:** leadership team; help run the house athletics program.

## Skills and Interests

**Technical:** Python, C, MATLAB, R, Java, SQL, LaTeX
**Machine Learning and NLP:** gradient-boosted trees, SVMs, neural networks, matrix factorization (PyTorch, TensorFlow, scikit-learn), sentence embeddings, topic modeling (LDA, keyATM), speech-to-text pipelines (faster-whisper), statistical modeling (regression, time series, Bayesian)
**Tools:** Git, Linux and SSH, GPU servers (CUDA), STM32, Arduino
**Language:** English (fluent), Korean (fluent), Spanish (intermediate)