# Albert Tam

altam@mit.edu ◆ github.com/tam-albert ◆ tam-albert.github.io

## Education

**Massachusetts Institute of Technology** *Expected graduation:* May 2026
*B.S. in Computer Science, B.S. in Mathematics* GPA: 5.0/5

- **Courses**: Stochastic Processes, Abstract Algebra, Deep Generative Models, Natural Language Processing, Sensorimotor Learning, Efficient Deep Learning, Sublinear Time Algorithms, Computer Systems Engineering, Machine Learning

## Work Experience

*Machine Learning Research Intern* | **Phonic** • *Text-to-Speech* Jun 2024 – Aug 2024

- Post-trained text-to-speech models to follow natural language descriptions, improving key quality measure from **60%** to **74%**.
  - Created a **3,000-hour** training dataset of speech samples, annotated with synthetic voice descriptions.
  - Developed test set and quantitative methodology to evaluate quality of generated samples and adherence to text prompts.
  - Designed custom architectures to adapt large audio diffusion transformers for conditioning on text inputs.
- Researched contrastive learning methods with synthetic data to further improve TTS controllabillity.
  - Achieved **5x** performance on an internal retrieval benchmark over existing benchmarks.

*Researcher* | **Madry Lab, MIT Computer Science and Artificial Intelligence Laboratory** • *Interpretability* Jun 2023 – May 2024

- Researched interpretability methods to attribute image segmentation model predictions to training data [NeurIPS '23].
- Designed dataset curation method that achieves **3x** data efficiency in training robust, generalizable segmentation models.
- Improved quality of popular image datasets (e.g. MS COCO) by systematically identifying mislabeled training examples.

*Software Engineer Intern* | **Inkeep (YC W23)** • *LLMs and RAG* Jan 2024 – Feb 2024

- Leveraged information retrieval ranking algorithms and OpenAI API to improve recall of relevant coding documentation in a retrieval-augmented generation (RAG) product for developers.
- Developed query router to improve quality of answers for complex user queries, while hitting critical latency targets.

## Projects

**Natural Language Terminal Assistant (TreeHacks 2024)** Feb 2024

- Built a developer tool in Zig using a local, fine-tuned Mixtral 8x7B model to transform natural language into shell commands.

**Constructing Defenses Against Adversarial LLM Jailbreak Attacks** Nov 2023

- Fine-tuned a LLaMa-7b model with soft, continuous prompts to defend against published LLM jailbreaks to improve model safety.

**Conversational Language-Learning Chatbot (LAHacks 2023)** Apr 2023

- A language-learning chatbot for users to practice conversations out loud with, using OpenAI and GCP APIs.

## Leadership & Awards

**Director, HackMIT** Sep 2022 – Present

- Leading a team of 40 to organize one of the largest hackathons in the US, with 1,000+ participants annually and a budget of $300K+.
- Managed 4 subteams, overseeing in-house app development, brand design, corporate sponsorships, and event logistics.

**USA Biology Olympiad National Finalist, Top 12 (2021)**: top 0.5% of high school students. Invited to attend most prestigious biology summer program in the US, and placed top 12 among invited finalists.

**American Invitational Mathematics Examination, 5x Qualifier (2018 – 2022)**: top 5% of high school students.

## Publications

**Albert Tam**, Josh Vendrow, & Aleksander Madry. "Data Attribution for Image Segmentation Models." NeurIPS 2023 Workshop on Attributing Model Behavior at Scale (2023)

## Skills & Activities

- **Languages/frameworks**: Python, PyTorch, NumPy, Pandas, JavaScript, React, MongoDB, Express.js, Flask
- **Activities**: HackMIT, Harvard-MIT Math Tournament, MoveMENtality dance crew, MIT Asian Dance Team