Tian Zhang

Philadelphia, PA
tianzh@seas.upenn.edu

ABOUT

I am a fifth-year Ph.D. student in Computer and Information Science at the University of Pennsylvania, advised by Erik Waingarten. My research develops algorithms for massive datasets, with a particular interest in similarity search as well as its applications in information retrieval, Retrieval Augmented Generation (RAG) and recommendation systems. I am also interested in VLM/LLM post-training and applications, most recently as an Applied Scientist Intern at Amazon. Before Penn, I received my B.S. in Computer Science and Statistics from the University of Michigan–Ann Arbor, where I did undergraduate research with Thatchaphol Saranurak.

EDUCATION

University of Pennsylvania — Philadelphia, PA
Ph.D. in Computer and Information Science — *Expected Jan 2027*

- GPA: 3.98/4.00. Relevant coursework: Machine Learning; Algorithms for Big Data; Randomized Algorithms; Advanced Database Systems.

University of Michigan — Ann Arbor, MI
B.S. in Computer Science and Statistics (double major), with High Distinction — *2022*

- GPA: 3.94/4.00. Relevant coursework: Data Mining; Intro to AI; Conversational AI; Web Systems.

PUBLICATIONS

Authors in alphabetical order, as is the convention in theoretical CS.

- Learn Where to Store: Query-Aware IVF. Zack Ives, Jiaming Liang, Ashwin Padaki, Erik Waingarten, Tian Zhang. *In Submission.*
- Fast Geometric Spanners via Approximate Nearest Neighbor Search. Alexandr Andoni, Manuel Paez, Krish Singal, Tian Zhang. *In Submission.* [arXiv]
- Prune, Don't Rebuild: Efficiently Tuning α-Reachable Graphs for Nearest Neighbor Search. Tian Zhang, Ashwin Padaki, Jiaming Liang, Zack Ives, Erik Waingarten. *NeurIPS 2026.* [arXiv]
- Approximate Nearest Neighbor in Ultra-High Dimensional ℓ_∞. Nathan White, Tian Zhang. *In Submission.* [arXiv]
- Near-Optimal Directed Euclidean Spanners in High Dimensions. Rajesh Jayaram, Shyamal Patel, Cliff Stein, Erik Waingarten, Tian Zhang. *STOC 2026.* [DOI]
- Average-Distortion Sketching. Yiqiao Bao, Anubhav Baweja, Nicolas Menand, Erik Waingarten, Nathan White, Tian Zhang. *FOCS 2025.* [arXiv]
- Data-Dependent LSH for the Earth Mover's Distance. Rajesh Jayaram, Erik Waingarten, Tian Zhang. *STOC 2024.* [arXiv]
- Vertex Sparsifiers for Hyperedge Connectivity. Han Jiang, Shang-En Huang, Thatchaphol Saranurak, Tian Zhang. *ESA 2022.* [arXiv]

EXPERIENCE

Amazon — Seattle, WA
Applied Scientist Intern, IP Protection — *May 2026 – Aug 2026*

- Designed and evaluated a two-step infringement-detection pipeline that first identifies candidate brands, then judges infringement per brand — outperforming the single-pass end-to-end production model by up to 6.6 points on accuracy while making errors attributable to each stage.
- Built the end-to-end training and evaluation infrastructure for teaching a vision–language model to recognize brands from a one-million-brand catalog, and delivered the team's best-performing model on brand recognition — improving brand-recognition accuracy from 78.3% to 90.4% — through a novel post-training recipe validated by a 20+ model ablation study.
- Discovered systematic flaws in the team's primary benchmark (59% duplicate records, incomplete answer labels that understated model accuracy) and built an LLM-based audit

pipeline over 53K products to correct it — the corrected benchmark was adopted as the team standard.

Kuaishou Technology Beijing, China
Data Scientist Intern, Recommendation System Group *Jun 2021 – Aug 2021*

- Built video-retrieval pipelines feeding candidate sets to downstream re-ranking on a large-scale short-video platform serving hundreds of millions of users.
- Trained Word2Vec-variant embeddings of videos and apps on random-walk sequences over a user–app–video relation graph, enabling personalized retrieval from users' installed-app lists.

VISITING POSITIONS

Simons Institute for the Theory of Computing Berkeley, CA
Visiting Graduate Student, Sublinear Algorithms Program *May 2024 – Jun 2024*

University of Wisconsin–Madison Madison, WI
Visiting International Student (GPA: 4.00/4.00) *Jan 2019 – Aug 2019*

TEACHING

- **University of Pennsylvania**: Course development, Algorithms for Big Data (CIS 5030, Spring 2026); Teaching Assistant, Algorithms for Big Data (CIS 6770, Spring 2025) and Introduction to Algorithms (CIS 3200, Fall 2023).
- **University of Michigan**: Grader, Introduction to Algorithms (EECS 477, Fall 2021).

SKILLS

Programming: Python, C++, R.

AI/ML: LLM/VLM post-training and applications, basic machine learning algorithms (linear regressions and variants, random forest), statistical analysis, algorithm design for similarity search, large recommendation systems, and RAG systems.

Tools: SQL, Git, Linux.

AWARDS & HONORS

- **Abel Lukens Stout Fellow**, University of Pennsylvania *2022*
- **James B. Angell Scholar**, University of Michigan *2021*
- **Excellent Visiting International Student Award**, University of Wisconsin–Madison *2019*