Tianhu Xiong

tianhux2@illinois.edu | (312)-292-7425 | Champaign, IL | linkedin.com/in/tianhu-xiong

EDUCATION

University of Illinois at Urbana-Champaign May 2028
B.S. in Mathematics & Computer Science (**James Scholar Honors**) GPA: 4.0/4.0

EXPERIENCE

Research Assistant, MONET Group – University of Illinois Urbana-Champaign October 2025 – Present
Advisor: Professor Klara Nahrstedt

Diffusion Language Models with Test-Time Training (Current Project) June 2026 – Present

- **Architected** a 1.3B-parameter Test-Time Training (TTT) model in **JAX/Flax** with end to end fast-weight updates and outer diffusion objectives, cutting pre-training loss by **4.8%** across multiple datasets with **+11.3%** beyond-window recurrence gain.
- **Optimized chunk-wise computation** and **multi-level nested rematerialization** for second-order bi-level graphs, balancing compute and memory to push cluster throughput to **59k tokens/s** on a **32-chip TPU v4 Pod**.

SlideForge: End-to-End Multi-Agent Editing Framework (EMNLP 2026) April 2026 – July 2026

- **Architected** an end-to-end multi-agent framework to programmatically decompose slides into a **Deck State Graph (DSG)**, achieving **0.950 SSIM** and **0.984 CLIP** reconstruction fidelity during **visual component extraction**.
- **Engineered** a robust document generation pipeline preserving on average **72.8** editable shapes per slide, achieving a **1.35×** editable-count ratio and **100%** structural preservation on text-panel restyling.
- **Developed** a closed-loop visual compiler with **headless rendering** to detect spatial anomalies and back-propagate errors for parametric self-repair, requiring an average of only **3.0** localized corrections per deck.

Circuit Tracing in Vision-Language Models (CVPR 2026 Finding, First Author) October 2025 – January 2026

- **Trained** usable **SAEs** and transcoders for **Gemma-3-4B** across an **8×H100 GPU** cluster (**60-hour** execution in **PyTorch/TransformerLens per model**), decomposing entangled visual-linguistic features into monosemantic concepts.
- **Engineered** a scalable circuit-tracing and attribution pipeline for **Gemma-3-4B**, pre-computing **~2TB of attention maps** across **28,000+** images in just **2 hours** on a single **H100 GPU**.
- **Implemented** targeted activation patching and feature steering modules to programmatically debug and mitigate model hallucinations at the tensor level.

PROJECTS

Multilingual Simultaneous Interpreting Software | [GitHub] May 2025 – August 2025
Independent Developer

- **Engineered** a **C++/Qt** desktop client integrating **Windows WASAPI Loopback** and **Media Foundation MFT** for real-time audio resampling, decoupling ingestion and rendering across worker threads.
- **Deployed** quantized streaming ASR (**Zipformer**) and offline **NMT** pipelines via **ONNX Runtime** and **CTranslate2**, achieving sub-**0.5s** latency for smooth real-time execution on offline **edge devices**.

High-Performance P2P NAT Traversal & Signaling System | [GitHub] January 2025 – April 2025
Independent Developer

- **Built** a custom reliable **UDP** protocol via **C++17** and **Standalone Asio**, implementing dynamic **RTO windowing** and **end-to-end encryption** to deliver **~90 MB/s** baseline and sustain **30+ MB/s** under **20%** simulated packet loss.
- **Architected** an asynchronous **Django REST** and **Channels WebSocket** signaling backend, coordinating automated **STUN** UDP hole-punching and real-time session negotiation secured by device-bound **JWT** authentication.

SKILLS

Programming Languages: Python, C++, Java, JavaScript, SQL, Bash
AI & ML Systems: JAX (XLA, Equinox), TPU & GPU Acceleration, PyTorch, Distributed Training, LLMs/VLMs, Agent Frameworks, Mechanistic Interpretability
Infrastructure & Databases: Vector Databases (FAISS/ChromaDB), Linux, Slurm, Redis, MySQL, Docker, Git

HONORS AND ACHIEVEMENTS

National Olympiad in Informatics (China NOIP 2022): First Prize; **USACO 2023 Open:** Platinum Division; **2026 CVPR Broadening Participation Scholarship, IEEE/CVF**; **2023 AMC 12:** Distinction; **2024 AIME:** Excellence (score: 10), Qualify for USAMO; **2024 Kaggle:** Chatbot Arena Human Preference Predictions, Silver; **2024 Kaggle:** Lumbar Spine Degenerative Classification, Silver; **Physics Bowl 2024:** Regional Top 10, National Gold