

Experience

Cofounder & CTO | Tilde Research

Jun 2024 – Jul 2025

Applied research for fundamentally understanding the limits of models so that we can unlock paradigm shifts in throughput, performance, and safety. Leading engineering to enable interpretability research across the entire stack, which includes infrastructure for pre-training, post-training, data curation + generation and more.

Raised \$8M from Khosla Ventures, BCV, and researchers at OpenAI and Anthropic.

Cofounder & CTO | Hazel

Jan 2023 – Jan 2024

Built Siri for realtors (I wanted to make my mom's work easier). Led engineering & product design. Automated various realtor back-office pipelines. At 17, raised \$2M from Pear VC, celebrity realtors, and top brokerage execs.

Researcher | Stanford Open Virtual Assistant Lab (OVAL)

Jun 2022 – Nov 2024

First high-schooler to present to Stanford NLP Group (and guest speak at a CS grad class). Advised directly by PI Monica Lam (CS) and Dr. Lynn Koegel (Psychiatry).

Researcher | UC Santa Barbara FourEyes Lab

Jun 2021 – Jan 2022

Primarily developed label-noise tolerant and modality-agnostic augmentation methods for multimodal learning.

Selected Research / Projects

Sparsity is Cool: Interpretability Insights into Sparse Attention

Dhruv Pai\*, Timor Averbuch\*, **Mason Wang\***, Ben Keigwin

Reverse engineered sparse attention models to understand how they train faster and beat full attention. Built harness for all our mechanistic analysis (i.e., the results shown in the post). Developed infra for pre-training models. Proposed and evaluated further inference-time improvements on NSA motivated by our insights, demonstrating minimal performance degradation.

Activault: Scalable, Efficient, and Fast Model Activation Storage

**Mason Wang\***, Dhruv Pai\*, Ben Keigwin

Built an entire S3-based engine for collecting, storing, and serving model activations. Aggressively optimized a pipeline to saturate both GPU utilization and network bandwidth. Reduced costs for interpretability research by an order of magnitude while maintaining efficiency and throughput.

Noora: Improving Empathetic Social Conversation in People with Autism

Built a full-stack AI coach that helps autistic adults practice social conversation. Ran user studies and adapted the UI for autism, where standard UX patterns often fail. Validated in a randomized clinical trial; published in Journal of Autism and Developmental Disorders.

Sparse Fusion for Multimodal Transformers

Yi Ding\*, Alex Rich\*, **Mason Wang**, Noah Stier, Matthew Turk, Pradeep Sen, Tobias Höllerer

By exploiting redundancies in representations across different modalities, we present a novel multimodal fusion method that matches SOTA while significantly reducing memory and computation.

Education

Stanford University

GPA: 0.0

Deferred freshman year to build Hazel and travel the world. Enrolled in (but couldn't attend) one quarter while cofounding Tilde Research to maintain eligibility to return.

Awards

- At 14, won **1st Place at UCLA's LAHacks** (Southern California's largest hackathon) for making 3D tours and virtual tours of houses easy (1,100+ competitors)
- At 15, **3rd Place at HackMIT** for minimal-latency video chat w/ RAE audio compression (1,100+ competitors)
- At 15, **3rd Place at COVIDathon** for a decentralized PWA to incentivize neighbors to help each other (1,300+ competitors)