ANDREW HANZHUO ZHANG

a663e-36z1120.github.io Google Scholar github.com/a663E-36z1120 linkedin.com/in/a663e-36z1120
andrewhz.1120@outlook.com +1 (647)-818-1672 Toronto, ON Canadian English & Mandarin

EDUCATION

University of Toronto Sep 2025 - May 2027 (Expected)
MSc in Computer Science supervised by **Prof. Anna Goldenberg** MSc-PhD Track
Research Area: AI and ML methods for biomedical and clinical applications.

University of Toronto Sep 2020 - May 2025 (Expected)
HBSc with 16 months full-time **ASIP** internship Graduation Grade: High Distinction (Expected)

Triple Major:

Computer Science	(12.0 Credits)
Cognitive Science	(8.0 Credits)
Physics	(8.0 Credits)

Awards:

University of Toronto Scholar	(Fall 2020)
Trinity College 6T5 Scholarship	(Fall 2021)
Dean's List Scholar	(Fall 2021-2024)

PUBLICATIONS & PRESS

[2] **Andrew Zhang**, Chunlin Li, Yuzhi Tang, Alex He-Mo, Nasim Montazeri Ghahjaverestan, Maged Goubran, and Andrew Lim. "*A Deep Learning Model for Inferring Sleep Stage from a Flexible Wireless Dual Sensor Wearable System without EEG*". In: *SLEEP* 47 (2024), A481–A482.

[4] Julie Choi, on behalf of the **Applied ML Team**. *Cerebras Selects Qualcomm to Deliver Unprecedented Performance in AI Inference.* Cerebras Systems Press Release. March 11, 2024.

MANUSCRIPTS

[1] **Andrew H. Zhang**†, Alex He-Mo†, Richard Fei Yin†, Chunlin Li, Yuzhi Tang, Dharmendra Gurve, Veronique van der Horst, Aron S. Buchman, Nasim Montazeri Ghahjaverestan, Maged Goubran, Bo Wang, and Andrew S. P. Lim. "*Mamba-based Deep Learning Approaches for Sleep Staging on a Wireless Multimodal Wearable System without Electroencephalography*". In: ☠ (2024).

[3] Chloe X. Wang†, Haotian Cui†, **Andrew H. Zhang**, Ronald Xie, Hani Goodarzi, and Bo Wang. "*scGPT-spatial: Continual Pretraining of Single-Cell Foundation Model for Spatial Transcriptomics*". In: ℞ (2025).

†These authors contributed equally.

RESEARCH HIGHLIGHTS

scGPT-Spatial – Single-cell Foundation Model for Spatial Transcriptomics [3] Sep 2023 - Feb 2025

Supervisor: Prof. Bo Wang University of Toronto & Vector Institute

- Member of research team investigating continually pretraining single-cell foundation model scGPT (Cui et al., 2024) on spatial transcriptomic modalities such as Visium, Xenium, and MERFISH to address the unique complexities of these data distributions.
- Designed and developed methods for embedding-based spatial cell type deconvolution and gene imputation downstream tasks.
- Developed and benchmarked auxiliary self-supervised training objective task heads to improve pretraining performance.

Speculative Decoding for LLMs with Unstructured Sparsity [4] May 2023 - May 2024

Supervisors: Mr. Abhay Gupta & Dr. Ganesh Venkatesh Cerebras Systems

- Developed experiments using LLaMa-based language models with unstructured sparsity for Speculative Decoding (Leviathan et al., 2023) as a part of the applied ML team's collaboration with Qualcomm [4] to deliver high throughput inference solutions.
- Investigated methods for improving token acceptance rate of speculative decoding such as sparse-dense KV cache sharing.
- Further explored single-model speculative decoding methods such as Medusa (Cai et al., 2024) and Hydra (Ankner et al., 2024) more suitable for the Cerebras CS-X inference stack.

Deep Learning Approaches to Wearable Sensor Sleep Staging [2][1] Sep 2022 - Dec 2024

Supervisor: Prof. Andrew Lim Sunnybrook Research Institute

- Led research project at the Sleep and Brain Health Laboratory investigating deep learning approaches for accurate sleep staging using the Sibel Health ANNE One — a wireless wearable system that measure ECG, PPG, accelerometry, and temperatures.
- Poster presented at the SLEEP 2024 conference in Houston, Texas; Abstract published in the journal *SLEEP* [2].
- Further investigation [1] of approaches using Mamba (Gu & Dao, 2023) achieves state-of-the-art performance.

EMPLOYMENT HISTORY

Vector Institute — May 2024 - Sep 2024

Research Intern — Toronto, ON, Canada

- Full-time research internship at WangLab supervised by Prof. Bo Wang.
- Continuation of work from the CSC494/495 research course (Sep 2023 - May 2024) on scGPT-Spatial. (See Research Highlights)
- Further exploratory work on inference-time evolutionary muti-agent LLM reasoning with Monte-Carlo tree search.

Cerebras Systems — May 2023 - May 2024

Applied ML Research Engineer — Toronto, ON, Canada

- Full-time 12 months ASIP co-op internship term as a part of the applied ML team.
- Focused on speculative decoding for LLaMa-based models with unstructured sparsity. (See Research Highlights)

Sunnybrook Research Institute — Sep 2022 - Sep 2023

Student Researcher — Toronto, ON, Canada

- Part-time research position exploring deep learning approaches to wearable sensor sleep staging without EEG under the supervision of Prof. Andrew Lim at the Sleep and Brain Health Laboratory. (See Research Highlights)

Sunnybrook Research Institute — May 2022 - Sep 2022

Software Engineer — Toronto, ON, Canada

- Full-time 4 months ASIP co-op internship term as a full-stack engineer developing the medical time-series annotation platform CrowdEEG (Schaekermann et al., 2020) at the Sleep and Brain Health Laboratory. (See Engineering Experience)

ENGINEERING EXPERIENCE

brainblots – a Brain Signal Algorithmic Art Project — Personal Project

- Co-founded brainblots – a brain signal algorithmic art collective to provide human beings with additional dimensions of expressing ourselves beyond what evolution gave us by using the Muse EEG headband.
- Deployed our project at art events across Toronto, New York City, and Boston, collecting 'brainblots' of hundreds of individuals. Digital artworks exhibited at Time Square, New York City in June 2022, and curated as NFTs.

GPT-Neox - Open Source Contributions — Cerebras Systems

- Took initiative to upstream bug fixes and new features from Cerebras's internal LLM pretraining test-bench forked from EleutherAI's GPT-Neox project, such as integration of FlashAttention-2 (Dao, 2023).

CrowdEEG — Sunnybrook Research Institute

- A collaborative annotation tool for medical time series that was initially a demo platform developed by Schaekermann et al.
- My internship adapted it to become a fully functional open-source project to support clinical studies at the Sleep and Brain Health Laboratory, which was eventually deployed into production at the Augmented Intelligence Lab of the University of Waterloo.

Gesture Imitation Robotic Hand — Coursework & Personal Project

- A 3D-printed robotic hand that imitates hand gestures in real time with computer vision which began as coursework for MIE438.
- Designed and developed the computer vision pipeline and communication protocol between Raspberry Pi and Arduino Mega. Optimized PWM motor control loops.

TEACHING & MENTORING

COG402H1: Seminar in Cog. Sci. - Cognitive Scientific Theories of Consciousness — Fall 2024

- Taught seminar session on *Insights into the Functions and Nature of Consciousness through Generalizing Global Workspace Theory to Artificial Neural Networks*.

NeurotechUofT — Summer 2021 - Fall 2023

- Led the organization at the position of **signal processing team lead**. Led EEG signal processing workshops and tutorials using the OpenBCI Cyton board and Muse EEG headband with Python.

CSC165H1: Mathematical Expression and Reasoning for Computer Science — Winter 2021

- Leader of Recognized Study Group for the course at the University of Toronto.
- Held formal proof tutorials and course content office hours for participating students.