

# Xingchen Zhao

4 years of experience in Machine Learning Engineering/Research, 7 ML papers published

☎ +1-571-525-4502   🔗 [tinyurl.com/bddsvxzh](https://tinyurl.com/bddsvxzh)   ✉ [zhao.xingc@northeastern.edu](mailto:zhao.xingc@northeastern.edu)   🔗 [linkedin.com/in/xingchenzhao/](https://linkedin.com/in/xingchenzhao/)

## Education

### Northeastern University

Boston, MA

Master of Science in Computer Engineering

August 2021 - April 2024

PhD in Computer Engineering

(Dropout) August 2021 - January 2023

### University of Pittsburgh

Pittsburgh, PA

Bachelor of Science in Computer Science, GPA:3.71/4.0

January 2018 – May 2021

UC San Diego, ML Visiting Scholar

November 2020 - May 2021

## Working Experience

### Machine Learning Applied Scientist Intern

August 2023 – Jan 2024

Amazon Robotics

Westborough, MA

- Led development of light-weight real-time object detector for AR-ID Localizer, enhancing performance by **5%** mAP and increasing speed by **120%** on edge devices, achieved by distribute training on 40k shipping label images.
- Designed unified model for localizing 1D and 2D barcodes on Amazon packages and vendor products, improving accuracy by 3%. Facilitated efficient package sorting for logistics staff, enhancing delivery effectiveness for billions of customers.
- Deployed models in **ONNX** format on edge device, convert to **TensorRT** for optimized FP16 and Int8 inferencing.
- Evaluated model generality across applications using **AWS SageMaker** for scalable model comparisons.

### Founding Machine Learning & Software Engineer

February 2023 – Now

Learnie AI — [develop.learnieai.com](https://develop.learnieai.com) — Partnered with the US's largest edu company, HMH

Boston, MA

- Developed real-time AI-driven K-12 education platform leveraging NLP, Speech Recognition, Text-to-Speech using ChatGPT, Azure TTS API, and fine-tuned **LLaMA** model using Hugging Face's PEFT library on NVIDIA A100 GPU.
- Implemented a **LangChain** text-to-talking-head pipeline, achieving **65%** enhanced speed through PyTorch optimizations and parallelized TTS API; containerized for full-stack integration as a Flask **microservice** using **Docker**.
- Enhanced visual quality of talking head by merging Wav2Lip and Real-ESRGAN for lip-syncing and super-resolution.

### Founding Machine Learning & Software Engineer

April 2023 – Now

Chat AI Zoo — [chataizoo.com](https://chataizoo.com)

Boston, MA

- Led team of 10, building AIGC social platform for user-generated LLM agents; achieved 1K daily users and 100K clicks.
- Architected and developed backend systems for agent-based and model-based LLM, integrated with Weaviate vector database and Azure Search API; supported custom agents and advanced features using Langchain and FastAPI.
- Improved UX and API design for chat display, management, and dynamic content, leading to 50% rise in daily users.

### Machine Learning Engineer Intern

May 2022 – August 2022

SRI International

Princeton, NJ

- Developed label-free segment model, saved 50% on costs; won 1st in DarkZurich with self-supervised algorithm.
- Used **DeepSpeed** and ZeRO for distribute training with model/data parallelism, achieving a **7%** accuracy boost.
- Enhanced model efficiency by **90%** using Knowledge Distillation; deployed on AWS using ONNX and TensorRT.

### CV&NLP Researcher

February 2022 – March 2023

NEC Labs America

Boston, MA

- Developed a single stream architecture enhancing **image-language** alignment and semantic grounding at multi-levels.
- Achieved superior results in **image-text retrieval** and **VQA** against larger models using BERT and ViT efficiencies.
- Designed a two-stream model merging DETR and BERT, elevating object-aware **multimodal** sentiment analysis.

### Machine Learning Researcher

January 2020 – January 2023

Northeastern University, University of Pittsburgh, UCSD

Boston, Pittsburgh, San Diego

- Authored 8 ML papers on CV, NLP, and Multimodality; reviewed 20+ papers in CV, NLP, and Data Mining domains.
- Developed a **multi-modal** technique with Transformers, boosting object detection accuracy by 4.5% over SOTA.
- Implemented Fourier-based style calibration, enhancing vision model generalization by **6%** on benchmarks.

## Selected Publications (Machine Learning/Artificial Intelligence)

- **Zhao, X.**, Sicilia, A., ..., "Test-time Fourier Style Calibration for Domain Generalization", IJCAI, 2022.
- **Zhao, X.**, Minhas, D., ..., "Robust White Matter Hyperintensity Segmentation on Unseen Domain", ISBI, 2021.
- **Zhao, X.**, Xuehai H., ..., "Learning by Ignoring, with Application to Domain Adaptation", arXiv preprint 2012.14288

## Technical Skills

**Machine Learning:** Neural Language Processing, Computer Vision, Multimodal, Object Detection, Segmentation

**Programming languages:** Python, C++, Java, C, R, JavaScript, HTML, CSS, Swift, SQL

**Software Frameworks:** PyTorch, TensorFlow, DeepSpeed, TensorRT, Huggingface, Spark, OpenCV, React, NodeJS

**DevOps and Cloud Technologies:** Docker, Git, Kubernetes, Google Cloud, AWS