

Yixiao Chen

(617) 372-0480 | chen.yixia@northeastern.edu | [LinkedIn](#)

Education

Northeastern University, Boston, MA

September 2021 - December 2023

Master of Science, Electrical and Computer Engineer

Beijing University of Technology, Beijing, China

September 2017 - May 2021

Bachelor of Engineering, Automation (Robotics Engineering)

Technical Skills

- **Programming Languages:** Python, C/C++, C#/.NET core, SQL.
- **Tools & Frameworks:** Linux Server, Pytorch, Git/Stash, Llamacpp, KTransformers, TensorRT, CoreML.
- **Models:** DeepSeek V2/V3, Qwen 2/2.5, LLaMA 2, VIT, VIM.

Experiences

Machine Learning Engineer, *Albao LLC & EmbodyX*

May 2024 - Now

- Prune and Quantize DeepSeek V3 model with MoE Pruning algorithm and Llamacpp. Deploy MoE LLM on AMD Windows platform with Vulkan and HIP/ROCm backend.
- Develop real-time chatting system integrating Large Language Models (LLMs) and Text-to-Speech (TTS), enabling seamless and natural human-computer interaction.
- Deploy Digital Avatar Projects with CoreML and TensorRT, accelerate inference for real-time performance.
- Develop trajectory prediction assistant with transformer-based models for license learning exams.
- Video data filtering with face centering and scene detection, audio filtering with multi expert mean score.

Computer Vision Engineering Co-op, *Cognex Corporation, Natick, MA*

January 2023 - August 2023

- Developed a .NET wrapper layer to port an Edge Learning Tool between two versions of VisionPro software.
- Feature enhancement and bug fixes for VisionPro APIs in C++ and QuickBuild Edit Control UI in C#.
- Utilized Doxygen and SandCastle in .NET programs to generate HTML documentation and CHM files.
- Expanded testing coverage in C++ and C# as part of the CI Infrastructure (e.g., API Unit Tests, BOM Tests, Color Benchmark Tests, GUI Sanity Tests).

Publication

- **Yixiao Chen**, Yue Yao, Ruining Yang, Md Zakir Hossain, Ashu Gupta, Tom Gedeon, "Unsupervised Search for Ethnic Minorities' Disc-Cup Segmentation Training Set". **Accepted by ICASSP 2025.**
- **Yixiao Chen**, Ruining Yang, Xin Chen, Jia He, Dongliang Xu, Yue Yao, "From Static to Dynamic: a Survey of Topology-Aware Perception in Autonomous Driving". **Accepted by ICCV WDFM-AD 2025.**
- **Yixiao Chen**, Yue Yao, Ruining Yang, Md Zakir Hossain, Ashu Gupta, Tom Gedeon, "Collaborative Compression for Large-Scale MoE Deployment on Edge". **Accepted by ICASSP 2026.**
- Xin Chen, Jia He, Maozheng Li, Dongliang Xu, Tianyu Wang, **Yixiao Chen**, Zhixin Lin, Yue Yao, "Are VLMs Ready for Lane Topology Awareness in Autonomous Driving?". **Accepted by ICASSP 2026.**
- Amir Taherin, **Yixiao Chen**, Sana Taghipour Anvari, Ruben Noroian, Charles Amante, Pu Zhao, Bin Ren, Yanzhi Wang and David Kaeli, "System-Level Analysis of Quantized Large Language Models: A Case Study on Edge Computing Platforms". Submitted to **DAC 2026.**
- Zhendong Mi, **Yixiao Chen**, Pu Zhao, Xiaodong Yu, Hao Wang, Yanzhi Wang, Shaoyi Huang, "Effective MoE-based LLM Compression by Exploiting Heterogeneous Inter-Group Experts Routing Frequency and Information Density". Submitted to **ICML 2026.**