

Yiming Cheng

+86 17740867835 |eaminc0328@gmail.com |Beijing, China

EDUCATION

Tsinghua University	Department of Electronic Engineering	<i>Sep.2019-Jul.2024</i>
Bachelor of Engineering in <i>Electronic Engineering(Major)</i> <i>Minor in Statistics ,Minor in Laws</i>		
• <i>GPA: 3.28 GPA for Statistic minor:3.49 TOEFL: 106</i>		
University of Chicago	Department of Computer Science	<i>Expected:Dec.2025</i>
<i>Master in Computer Science(Pre-doc)</i>		
• <i>GPA: 3.90</i>		

PUBLICATIONS

- Yi Yang, **Yiming Cheng**, Hao Feng, Zhu Han “**Emotion-Aware Scene Adaptation: A Bandwidth-Efficient Approach for Generating Animated Short**”, accepted by MDPI-sensors
- **X Lan, Y Cheng, L Sheng, C Gao, Y Li** “**Depression detection on social media with large language models**”, submitted to The Web Conference 2024 - Web4Good
- Lan, x, Piao, ..., **Cheng, Y**, Gao, c, Li, Y, Niu, Y, Song, Y, Gai, K, & Li, Y (2024). Recommendation for Inclusivity of Underrepresented Producers in Usergenerated Content Platforms.submitted to NIPS2024
- Yi Yang, Hao Feng, **Yiming Cheng**, Yitong Ma, Zhu Han, “Minimizing Hallucinations and Communication Cost: Adversarial Debate and Voting Mechanisms in LLM-Based Multi-Agents,” submitted to MDPI
- **Yiming Cheng**, “Research on Recommendation System Technology Based on Large Language Models,” Graduation Design, Tsinghua University, 2024.
- **Patent**: Yi Yang, **Yiming Cheng**, Hao Feng, et al. “A Semantic Encoding and Decoding Framework for Converting Visual Content into Virtual Animated Visual Representations.”

RESEARCH & PROJECT EXPERIENCES

Future Intelligent Lab(FIBLAB),Tsinghua University *Jul.2022—present*
Research Assistant *Advisor: Assistant Prof. ChenGao*

Recommendation for Inclusivity of Underrepresented Producers in User-generated Content Platform

This project was aimed to provide fair recommendations for high-quality, low-attention creators on UGC with GNN.

- Take the pioneering step to thinking of the inclusivity issue of underrepresented producers in UGC(user-generated content) platform.
- Propose to construct a heterogeneous graph that can enrich the relations of vulnerable populations, and further propose graph neural networks to learn representations based on enriching features from multi-hop neighbors.

City Socioeconomic Simulator based on Large Language Models

This project was aimed to develop a city simulator while integrating a language model-based agent interaction system.

- Use UE to Build a visual model scene of Beijing (CBD district)
- Use python to write scripts for agents to interface with LLM and design the agents' memory mechanism to do POI recommendation.(POI means point of interest in the city)
- Implement and optimize the recommendation algorithm, with plans for integration into the laboratory's large simulation system and a future submission to SIGKDD.

Signal Processing Lab,Tsinghua University

Mar.2022—present

Research Assistant

Advisor: Aso Prof. YiYang

Emotion-Aware Scene Adaptation: A Bandwidth-Efficient Approach for Generating Animated Shorts

This project aims to recognize and restore emotions in semantic transmission scenarios under low bandwidth constraints

- Use the PyTorch framework, build an image element and emotion recognition model based on the CLIP model and InceptionV3, and use PAD (Pleasure-Arousal-Dominance) for emotion scoring.
- Enhance the generated semantics using the EmoCap model trained based on PAD scores for emotion style, ultimately achieving higher emotional coherence than the baseline on the received new video frames.

Wireless Networking, Signal Processing and Security Lab, University of Houston

April.2022—present

Research Assistant

Advisor: Prof. ZhuHan

Scalable AI Generative Content for Vehicular Network Semantic Communication

- This project aims to establish a large-model-based semantic communication channel and test its accuracy on a vehicular dataset
- Build and test a channel in PyTorch that uses CLIP to convert original images into semantics and then uses Stable Diffusion to restore semantics back into images.

Lmcache Team

Sep.2024-present

Open Source Contributor

Advisor: Prof. Junchen Jiang

Working on open-source project:

- **LMCache:** The first open-source Knowledge Delivery Network (KDN) that accelerates LLM applications up to 8x faster, at 8x lower cost.
- **VLLM/production stack:** Scale from single vLLM instance to distributed vLLM deployment without changing any application code.
- Contributed 458 lines of code, ranked No.6

Internship PROJECTS

Beijing Yuanqing Huihong Information Technology Co., Ltd. (Company)

Software Engineer

June.2023—Sep.2023

- Refactor the Sunflower library (the main functions include JSON parsing, MQTT, B-Stack device information parsing, and data transmission encryption) for the company's Internet of Things (IoT) data platform using Go-lang
- Perform functional and performance testing on the refactored Sunflower library.
- Collaborate with hardware interns to debug and ensure successful MQTT-based data transfer of bridge deflection, vibration frequency, and temperature data from LuZhou Bridge to the company's database.

Beijing Thunisoft Information Technology Co., Ltd.

Software Engineer

July.2022—Sep.2022

- Use Spring Batch to develop a batch job scheduling system supporting complex workflows and dependency management. Scheduled tasks are executed as planned using Cron expression triggers.
- Integrate Quartz scheduler for enhanced flexibility.
- Data integrity and stability are assured with Spring transaction management and JDBC operations.

OTHERS

Social works: Deputy Minister of Projects Department, Science Association, Department of Electronic Engineering, Tsinghua University

Computer Skills: C, C++, SQL server, Python, Java, R, Go, Linux, Matlab, Verilog, k8s etc.