Yichen (Eason) Lu

6105 Spirit St., Pittsburgh, PA| yichenl5@andrew.cmu.edu | +1 (217) 979-6718 | yichen14.github.io

Education

CARNEGIE MELLON UNIVERSITY
Master of Science in Artificial Intelligence   **Aug. 2023 – Dec. 2024**
Cumulative GPA: 3.84/4.00
Coursework: Large Language Models, LLM System, System Tool Chains for AI; Speech Recognition and Understanding

UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN
Bachelor of Science Statistics & Computer Science   **Aug. 2019 – May. 2023**
Cumulative GPA: 3.93/4.00 (Highest Honor; Dean's List)
Coursework: Computer Vision; Machine Perception; Autonomous Vehicle; Applied Machine Learning; Abstract Linear Algebra; Data Structures; Operation System; Statistical Modeling; Database Design; Web Development

Research Interests: Multimodal Large Language Models, Speech Language Models, Agent-based Translation System, Conversational AI

Publications

[1] ***FastAdaSP: Multitask-Adapted Efficient Inference for Large Speech Language Model***
Yichen Lu*, J Song*, CH Yang, Shinji Watanabe
EMNLP'2024 [Preprint]

[2] ***SynesLM: A Unified Approach for Audio-visual Speech Recognition and Translation via Language Model and Synthetic Data***
Yichen Lu*, J Song*, X Chang, H Bian, S Maiti, Shinji Watanabe
Interspeech'2024 Syndata4genai Workshop [Arxiv]

[3] ***Enhancing Audiovisual Speech Recognition by Bifocal Preference Optimization***
Yihan Wu, **Yichen Lu**, Y Peng, X Wang, R Song, Shinji Watanabe
Submitted to AAAI'2024

[4] ***Robust Audiovisual Speech Recognition Model with Mixture-of-Experts***
Yihan Wu, Yifan Peng, **Yichen Lu**, X Chang, R Song, Shinji Watanabe
SLT'2024

[5] ***Exploring Speech Recognition, Translation, And Understanding with Discrete Input: A Comparative Study***
Xuankai Chang, B Yan, K Choi, J Jung, **Yichen Lu,** S Maiti, R Sharma, J Shi, J Tian, Shinji Watanabe, Y Fujita, T Maekaku, P Guo, Y Cheng, P Denisov, K Saijo, H Wang
ICASSP'2024 [Paper]

[6] ***Noisy Positive-Unlabeled Learning with Self-Training for Knowledge Graph Completion***
Ruijie Wang, B Li, **Yichen Lu**, D Sun, J Li, Yuchen Yan, Shengzhong Liu, Hanghang Tong and Tarek F. Abdelzaher
ACL'2023 Findings[Paper]

Research Experiences

WAV LAB (CMU) | *Research Project | Advisor: Prof. Shinji Watanabe*   **July. 2023 – Present**
- Contribute to open source E2E Speech Processing toolkit **ESPNet** [GitHub], which is the most popular open-source speech processing repository.
- Research on **Speech Language Models efficient inference** via token merging and pruning. [[1](1)]
- Conduct research on **a unified approach for audio-visual language understanding** tasks. [[2](2), [3](3), [4](4)]
- Conduct research **on multimodal decoder-only language models** on speech related tasks. [[2](2)]
- Conduct research on utilizing SSL discrete token on ASR and ST. [[5](5)]
- Published paper on top computer science conferences.

CYPHY GROUP (UIUC) | *Research Assistant | Advisor: Prof. Tarek Abdelzaher*   **Jun. 2022 – May. 2023**
- Built graph neural network for defending global attack on dynamic link prediction task based on graph structure learning; reimplemented Models (**EvolveGCN, GAE/VGAE, Euler, VGRNN**) and several attack methods (Meta attack). [[5](5)] [GitHub]
- Designed a standard "config + trainer" deep learning infrastructure for GNN and conducted several experiments.
- Implemented a **Gradient Boost Regressor** kernel and **Random Forest Regressor** kernel for INCAS project, which is funded and managed by DARPA.

IM LAB (UIUC) | *Research Intern | Advisor: Prof. Kris Hauser, Prof. Yuxiong Wang*   **Jan. 2022 – Aug. 2022**
- Designed and developed a continual few-shot learning system's user interface, **interactive segmentation** component (RITM), context-aware copy-and-paste technique (CAPS) with a prototype-based **incremental few-shot semantic segmentation** model (PIFS), improving the robots' continuous learning abilities. [[Doc](Doc)]
- Designed and implemented a continual learning system benchmark in realistic setting.

Leadership/Project Experiences

ViDove (Website) | *Founder & Machine Learning Engineer*   **Feb. 2023 – Present**
- Developed an **Open-sourced End-to-end Video Translation Pipeline** for video subtitle generation in specific field.
- Managed open-source repository of ViDove[Github] **86 Stars**.
- Designed a specific testing metric for **video subtitle translation task** under realistic setting based on **COMET score**, **SacreBLEU**, and our own time stamp evaluation.
- Directed a technical team with 10 members and a marketing team with about 3 members.
- **IMPACT: Pigeon AI has already officially cooperated with StarCraft 2 World Team League (the greatest SC2 E-Sports tournament in the world) for helping them translate the tournament content.**

SUGAR MASSES CREATIVE | *Founder, Developer, & Director*   **Feb. 2014 - Dec. 2018**
- Launched the **largest official Minecraft tournament** in China (China Minecraft Construction Summit), **attracting over 1,000 contestants** and over **1 million online viewers** annually by leading a team of 30 members.
- Achieved long-term cooperation on Minecraft competitions with major internet companies in China including NetEase, Qihoo360, Tencent, and Youku; secured project investment from NetEase and JoyMe.com.
- Presented at the largest digital entertainment expo in China (ChinaJoy2016, with more than 325,000 entries by business partners, media, consumers, and exhibitors in the four-day event).

Work Experience

TrovaAI, Inc. | *Machine Learning Engineer Intern*   **May. 2023 – Aug. 2023**
- Designed an **RAG information retrieval pipeline for the AI agent** platform which ingests all types of documents, websites, tables, and live data, and create searchable vector and keyword indices.
- Designed and developed deep link function which will trigger the relevant application based on user's queries for AI agent.
- Designed and developed semantic search and AI assistant platform using LangChain and Pinecone Vector Database.
- Designed and developed the evaluation methods for whole AI Agent system via LLM, BLEU, COMET, etc.

VMware, Inc. | *Software Engineering Intern*   **Jul. 2021 - Jan. 2022**
- Individually designed and build up an automatic data analysis system for large-scale machine translation system(BERT-based).
- The system will automatically analyze the machine translation quality and its error rate and generate report monthly.
- Optimized data querying process speed by applying Elastic Stack to large dataset (2 tigabyte) management.

NetEase, Inc. (NASDAQ：NTES) | *Software Engineering Intern*   **Oct. 2020 - Feb. 2021**
- Developed the search, sorting, comic download, and comic sharing for "Marvel Unlimited" android version (Chinese official Marvel comic reading app; daily active user: 5000) independently using Java and the **android Jetpack tool library**. [Website]
- Finished comic download function of the app in a team of 3 following the development logic and architecture of android software of large commercial project; self-learned Android development techniques such as **DataBinding**, **DataViewModel** and Fragment.

Skills
- **Languages:** Python(Advanced), Java, C/C++, Rust, R, JavaScript, OCaml, Dart, SQL, CSS, TypeScript
- **Tools/Framework:** PyTorch, Git, LaTex, PySpark, React.js, Node.js, Flask, PostgreSQL, MongoDB, Linux, Slurm