Yanshu Li

Providence, RI, 02906 | yanshu_li1@brown.edu | (401)8377546 | Google Scholar

Education

Brown University, MS in Computer Science — Sep 2024 – Present

- GPA: 4.0/4.0
- **Core Modules:** Machine Learning, Natural Language Processing, Large Language Models, Multimodal Learning

Soochow University, China, BS in Artificial Intelligence — Sep 2020 – Jun 2024

- GPA: 3.8/4.0 (top 10%), TOEFL:107, GRE:332
- **Core Modules:** Machine Learning, Natural Language Processing, Computer Vision, Large Language Model, Pytorch/TensorFlow Programming, Applications of Visual Learning

Research Interests

Reasoning of LLM and LVLM, Vision-language Alignment, Multimodal Learning, Efficient AI, Multimodal Agents

Selected Publications

CAMA: Enhancing Multimodal In-Context Learning with Context-Aware Modulated Attention [Arxiv]

Yanshu Li, Jianjiang Yang, Ziteng Yang, Bozheng Li, Hongyang He, Zhengtao Yao, Ligong Han, Yingjie Victor Chen, Songlin Fei, Dongfang Liu, Ruixiang Tang

Accepted at AAAI 2026 (Oral)

CATP: Contextually Adaptive Token Pruning for Efficient and Enhanced Multimodal In-Context Learning [Arxiv]

Yanshu Li*, Jianjiang Yang*, Zhennan Shen, Ligong Han, Haoyan Xu, Ruixiang Tang

Accepted at AAAI 2026 (Oral)

TACO: Enhancing Multimodal In-context Learning via Task Mapping-Guided Sequence Configuration [Arxiv]

Yanshu Li, JianJiang Yang, Tian Yun, Pinyuan Feng, Jinfa Huang, Ruixiang Tang

Accepted at EMNLP 2025

M²IV: Towards Efficient and Fine-grained Multimodal In-Context Learning via Representation Engineering [Arxiv]

Yanshu Li*, Yi Cao*, Hongyang He, Qisen Cheng, Xiang Fu, Xi Xiao, Tianyang Wang, Ruixiang Tang

Accepted at COLM 2025 & Multimodal Augmented Generation via MultimodAl Retrieval @ ACL 2025 (Oral)

ReLoop: "Seeing Twice and Thinking Backwards" via Closed-loop Training to Mitigate Hallucinations in Multimodal Understanding

Jianjiang Yang*, ***Yanshu Li****, Ziyan Huang*

Accepted at EMNLP 2025 Findings

TRiCo: Triadic Game-Theoretic Co-Training for Robust Semi-Supervised Learning

Hongyang He, Xinyuan Song, Yangfan He, Zeyu Zhang, ***Yanshu Li***, Haochen You, Wenqiao Zhang

Accepted at NeurIPS 2025

M-ABSA: A Multilingual Dataset for Aspect-Based Sentiment Analysis [Arxiv]

ChengYan Wu, Bolei Ma, Yihong Liu, Zheyu Zhang, Ningyuan Deng, ***Yanshu Li***, Baolan Chen, Yi Zhang, Barbara Plank, Yun Xue

Accepted at EMNLP 2025

Draw with Thought: Unleashing Multimodal Reasoning for Scientific Diagram Generation [Arxiv]

Zhiqing Cui, Jiahao Yuan, Hanqing Wang, ***Yanshu Li***, Chenxu Du, Zhenglong Ding

Accepted at ACM Multimedia 2025

ConflictBank: A Benchmark for Evaluating Knowledge Conflicts in Large Language Models [Arxiv]

Zhaochen Su, Jun Zhang, Xiaoye Qu, Tong Zhu, ***Yanshu Li***, Jiashuo Sun, Juntao Li, Min Zhang, Yu Cheng

Accepted at NeurIPS 2024
Unveiling and Mitigating Short-Cut Learning in Multimodal In-Context Learning [Openreview]
Yanshu Li

Accepted at Spurious Correlation and Shortcut Learning @ ICLR 2025 (Oral)
Towards Generalizable Implicit In-Context Learning with Attention Routing [Arxiv]
Jiaqian Li*, ***Yanshu Li****, Ligong Han, Ruixiang Tang, Wenya Wang

Submitted to ICLR 2026
When Multimodal Models "Burn Out": Diagnosing and Healing Modality Fatigue via MAD + MAC
Jianjiang Yang*, ***Yanshu Li****, Ziyan Huang, Da Peng, Huaiyuan Yao

Submitted to ICLR 2026
DeCo-DETR: Decoupled Cognition DETR for efficient Open-Vocabulary Object Detection
Siheng Wang*, ***Yanshu Li****, Bohan Hu, Zhengtao Yao, Linshan Li, Yiran Wang, Zhengdao Li, Xi Xiao, HaiboZhan, Haoyan Xu, Zhikang Dong, Tianyang Wang, Qiang Sun, Jifeng Shen

Submitted to ICLR 2026
Breaking Missing Modality Barrier: Bohl Concept-Driven Semi-Supervised Multimodal Learning for Fine-Grained Analysis
Hongyang He, ***Yanshu Li***, Xinyuan Song, Yangfan He, Zeyu Zhang, Yundi Hong, Haochen You, Yinda Chen, Wenqiao Zhang

Submitted to ICLR 2026
ViP^2-CLIP: Visual-Perception Prompting with Unified Alignment for Zero-Shot Anomaly Detection
Ziteng Yang*, Jingzehua Xu*, ***Yanshu Li****, Zepeng Li, Yeqiang Wang, Xinghui Li

Submitted to WACV 2026

Surveys

Thinking with Images for Multimodal Reasoning: Foundations, Methods, and Future Frontiers [Arxiv]
Zhaochen Su, Peng Xia, Hangyu Guo, Zhenhua Liu, Yan Ma, Xiaoye Qu, Jiaqi Liu, ***Yanshu Li***, Kaide Zeng, Zhengyuan Yang, Linjie Li, Yu Cheng, Heng Ji, Junxian He, Yi R. Fung

From Mind to Machine: The Rise of Manus AI as a Fully Autonomous Digital Agent [Arxiv]
Minjie Shen, ***Yanshu Li***, Lulu Chen, Qikai Yang

Professional Experience

NLP Engineer, Baidu, Inc. Shanghai, China, Apr 2024 - Jun 2024

- Build, evaluate, and maintain a reusable framework for fine-tuning Baidu's commercial LLM, ERNIE Bot, to support Retrieval-Augmented Generation (RAG) across diverse downstream applications. The framework is used across four subsidiary platforms.
- Fine-tune Retriever and Reranker models using a proprietary pipeline and implement a synthetic data augmentation tool using in-context learning with Llama 2 to boost model performance.
- Achieve 8% improvement over SOTA baselines in in-domain contextual retrieval and drive the launch of the RAG application on Baidu's open-source deep learning platform, PaddlePaddle.

NLP Engineer, iFlytek Suzhou, China, Oct 2023 - Feb 2024

- Collect, clean, and label high-quality multimodal data for iFlytek's commercial legal LLM and build a structured knowledge base to train the model to provide case-judgment assistance of business quality. The solution is deployed in three law firms.
- Build a domain-knowledge injection pipeline for iFlytek's in-house pre-trained LLM to extend capabilities and enable effective adaptation, accelerating the development of downstream industry-specific LLMs.
- Develop LLM-based software products for dialog generation and sentiment analysis, and ship them on-device in the iFlytek Smart Notebooks.

Research Experience

Research on cognitive science in LVLM, Serre Lab, Brown University May 2025 – Present

- Investigate the perceptual and cognitive mechanisms by which large vision-language models process cross-modal information, conducting in-depth analyses of neuron-level information flow.
- Optimize LVLM architectures and steer their attention dynamics to endow these models with human-like reasoning capabilities.
- Build multimodal datasets informed by insights from cognitive science for training or fine-tuning LVLMs, strengthening their cross-modal reasoning and closing existing research gaps.

Research on complex vision-language reasoning, Rutgers University Jul 2024 – Present

- Explore the in-context learning mechanisms of multimodal large models by investigating their internal workflows, focusing on the impact of external knowledge on transformers' internal attention and hidden layers.
- Enhance the in-context learning capabilities of LLMs and MLLMs by optimizing in-context demonstration selection, leveraging chain-of-thought (CoT) reasoning, and strengthening modality alignment.
- Investigate the theoretical foundations of multimodal in-context learning and leverage meta-learning techniques to enhance models' adaptability.

Research on equipping LLMs with human-level temporal reasoning and commonsense, Soochow University Jun 2023 – Jun 2024

- Focus on mechanisms underlying LLM performance degradation over time and explore strategies to improve temporal generalization and update their knowledge without catastrophic forgetting of old knowledge.

Research on sentiment analysis based on language models in human-centered NLP, Summer Intern, Nanyang Technological University, Singapore Jul 2022 – Nov 2022

- Design efficient methods to process large-scale social text data using LM and LLM and analyze the results to provide feedback, seeking high-quality solutions for social issues.

Projects

Kaggle & Google-Fast or Slow? Predict AI Model Runtime Silver Medal (top 4%)

- Develop a Graph Convolutional Network (GCN) model with embedding encoding for discrete attributes and ListMLE loss for training to address sorting-related challenges. Enhance model performance through weighted fusion of multiple models to reduce bias and improve robustness.

Kaggle-LLM Science Exam Bronze Medal (top 7%)

- Employ a sentence vector approach to extract the most similar original text based on prompt similarity for better question references. Develop a reading comprehension model with a novel linguistic approach.