Boyu Guan

Phone (WeChat) : 18640473542
email: guanboyu2022@ia.ac.cn
Personal Website: https://boyuguan.github.io · Location: Beijing
Objective: Algorithm Internship –NLP & Multimodal Learning

EDUCATION

Institute of Automation, CAS, Beijing, *PhD* Sep. 2022 –Present

- Supervised by Prof. Chengqing Zong; expected graduation: June 2027
- Research: NLP, Multimodal LLMs, Video-Guided Machine Translation
- GPA: 3.6 / 4; completed core CS courses and English CSAPP with labs

Northeastern University, B.S. in Mathematics, Shenyang, China Sep. 2018 –Jun. 2022

- GPA: 3.76 / 5.00, top 3/31; admitted to Ph.D. program via recommendation
- Received multiple honors including the Second-Class Scholarship for Outstanding Students and the Second-Class Scholarship from Northwest Institute for Nonferrous Metal Research

Academic Research

Key Technologies for Large-Scale Multilingual Multimodal Neural Machine Translation, National Natural Science Foundation Key Project Jan. 2024 – Present

- **Research Focus**: Specializing in video-guided multimodal machine translation, with an emphasis on addressing challenges such as data scarcity, high computational cost, and unstable performance.
- **Publications**:
  - First-author paper accepted to **COLING 2025** (CCF B, **Oral Presentation, 9.8%**):
    TriFine: A Large-Scale Dataset of Vision-Audio-Subtitle for Tri-Modal Machine Translation and Benchmark with Fine-Grained Annotated Tags.
    Proposed the first large-scale tri-modal dataset (vision, audio, subtitle) with seven fine-grained annotation types. Introduced FIAT, a model-agnostic method for fine-grained multimodal input enhancement, significantly improving translation quality and efficiency.
  - First-author paper submitted to **EMNLP 2025 Main Conference** (CCF B):
    SHIFT: Selected Helpful Informative Frame for Video-Guided Machine Translation.
    Proposed SHIFT, a lightweight and pluggable modality-adaptive framework that selects keyframes or falls back to text-only inputs per sample via a clustering-selection mechanism, achieving quality gains while drastically reducing multimodal inference cost.
- **Patent**:
  - Co-inventor (advisor as first inventor); invention patent under substantive examination:
    A Video Machine Translation Method and Device Integrating Fine-Grained Multimodal Information.

Internship Experience

Biren Technology, Software Engineering Intern, Beijing Feb. 2023 –Aug. 2023

- Optimized pre-training and inference pipelines for LLaMA, LLaMA2, and ChatGLM; responsible for operator extraction, migration, and adaptation
- Focused on activation checkpointing and 3D parallelism; customized frameworks like DeepSpeed, Megatron-DeepSpeed, and Transformers for performance tuning

Profile Summary

Proficient in Python and Linux; experienced in multimodal data processing and model training, alignment, and optimization. Familiar with PyTorch, HuggingFace Transformers, and DeepSpeed. Strong mathematical and algorithmic foundation with solid CS knowledge. Self-driven, resilient, and passionate about multimodal AI.
Enjoy photography, beginner-level fitness, and gaming for relaxation and work-life balance.