Yingjue Bian

Carnegie Mellon University, Pittsburgh, PA 15213, USA

(+1) 412-689-8040 • yingjueb@andrew.cmu.edu • Personal Academic Page

EDUCATION

Carnegie Mellon University (CMU)   **Pittsburgh, U.S.**
Master of Science in Biomedical Engineering – Research (Early Graduation)   2024 - 2026
GPA: 4.0/4.0

Southeast University (SEU)   **Nanjing, China**
Bachelor of Engineering in Biomedical Engineering   2020 - 2024
GPA: 3.56/4.0 | **Average Score**: 87.86/100

PUBLICATIONS & MANUSCRIPTS

- D. Jiang, T. Wang, **Y. Bian**, S. Tang, T.-S. Lee, "Self-organization of Shape and Texture Functional Domains in Macaque V4," submitted to *Cerebral Cortex*, 2025. *(Under Review)*
- **Y. Bian**, Y.-H. Chen, Y. Xu, W. Zheng, M. Sawan, "Mirror Neuron System in Hands Motor Imagery via Action Observation: An EEG–fNIRS Study," manuscript in preparation, 2025.

TECHNICAL SKILLS

- **Computational**: Python (PyTorch, MNE, transformer, TensorFlow, vllm), MATLAB (EEGLAB), C++
- **Biomedical/Cognitive**: EEG/fNIRS Acquisition (BrainVision Recorder), Inertial Motion Analysis
- **Languages**: Mandarin (Native), English (IELTS 7, GRE 320)

RESEARCH EXPERIENCE

Neural Coding of Macaque V4 via Digital-Twin Modeling   **Pittsburgh, USA**
Research Assistant, School of Computer Science & Neuroscience Institute, Carnegie Mellon University,
Supervisor: Prof. Tai–Sing Lee   Sep 2024 – Present

- Integrated a deep-learning V4 digital twin with natural-scene and shape–texture batteries; proposed *Unit Preference Index* and *Dispersity* to quantify selectivity; observed dispersion–tuning correlation ($r = 0.59$, $p < 0.01$).
- Designed 3D–2D object stimuli in Blender; identified neuron subtypes with modality-specific selectivity; indices (Dispersion, Shape–Texture) showed consistent negative correlations with preference (e.g., $r = -0.376, -0.623$).
- Used VLMs (Qwen-2.5-VL) to caption maximally driving stimuli, linking semantic descriptors to neural tuning profiles.

Hands Motor Imagery via Actions Observation: An EEG-fNIRS Study   **Hangzhou, China**
Visiting Student, CenBRAIN, Westlake University, Supervisor: Prof. Mohamad Sawan   Jun 2023 - Nov 2023

- Proposed a brain-computer interface protocol including motor execution, motor imagery, picture-guided motor imagery, and video-guided motor imagery tasks and collected EEG & fNIRS data from 42 volunteers.
- Conducted time-frequency analysis of EEG data, applying advanced signal processing techniques to investigate neural activity across different frequency bands.
- Applied statistical parametric mapping to analyze fNIRS data with NIRSlab, investigating brain activity and extracting meaningful neural correlates.

Diagnosis of Early-stage Parkinson's Disease based on Inertial Sensors   **Nanjing, China**
Research Assistant, Medical Imaging and AI Lab, Southeat University,
Supervisor: Dr. Ping Zhou   Jan 2022 - Apr 2022

- Developed an LSTM-FCN classifier with Model-Agnostic Meta-Learning (MAML), boosting diagnostic accuracy from 75% to 80%.
- Collected and augmented gait data from 110 participants; performed denoising, axis calibration, period segmentation, and zero-padding alignment.

- Built baseline 1-D CNN and LSTM models and executed few-shot transfer learning to ensure robustness on limited clinical samples.

SELECTED PROJECTS

Large-Scale Multilingual S2ST Corpus Construction and Benchmarking

Jul 2025 – Present

- **Developed a 100K+ hour multilingual S2ST corpus** from YouTube auto-dubbed multi-track videos, incorporating domain-aware sampling, audio alignment, and quality filtering to address temporal asynchrony and ensure high-quality parallel speech pairs.
- **Implemented and evaluated a transformer-based end-to-end S2ST baseline**, using the open-source StreamSpeech (ACL'24) framework, benchmarking offline and simultaneous translation with BLEU, TER, and MCD metrics.

Exploring Unsupervised Pretraining Paradigms for Speech Representation Learning

Jul 2025 – Present

- **Contrastive and Masked Modeling at Scale** Investigated self-supervised objectives including contrastive learning (e.g., CLIP/BLIP-inspired audio-text alignment) and masked autoencoding (AudioMAE), analyzing their scaling behavior, invariance properties, and downstream transferability.
- **Autoregressive Pretraining vs. Causal Decoding**, Compared autoregressive pretraining approaches (Apple AIM series) with non-causal masked autoencoders, evaluating the potential of causal decoding for temporally coherent speech modeling and its trade-offs against bidirectional masked objectives.

End-to-End Large Language Model Training and Alignment System

Jul 2025 – Aug 2025

- Built an end-to-end language model training and alignment pipeline from scratch, including a custom byte-level BPE tokenizer, Transformer implementation, and AdamW-based training loop; pre-trained and fine-tuned a LLaMA-style model on TinyStories and OpenWebText.
- Developed a scalable data processing pipeline for Common Crawl, incorporating deduplication, filtering, and contamination detection, producing a high-quality training corpus with unified loading and splitting workflows.
- Applied supervised fine-tuning and RLHF (PPO/DPO) for mathematical reasoning and instruction-following tasks, optimizing KV cache and batching strategies to reduce inference latency and memory usage; achieved stable PPL convergence and significant task performance gains over unaligned baselines.

Wearable Human Technology - Robust Joint-Moment Estimation via IMU   **Pittsburgh, USA**
Feb 2025 – May 2025

- Implemented a Late Fusion framework (per-sensor LSTM encoders + MLP) with stochastic modality dropout, preserving baseline accuracy when one or two sensors failed (RMSE $\leq$ 0.152 Nm/kg, $R^2 \geq 0.88$).
- Designed a two-stage **Self-Supervised Transformer**: pre-training on AMASS + MoVi, followed by fine-tuning; achieved **RMSE = 0.102 Nm/kg, R^2 = 0.779** on an external dataset.
- Ran leave-one-subject-out and transfer-learning analyses; showed that fine-tuning on data from **three** subjects suffices for generalization and identified the pelvis IMU as the most informative channel.

HONORS & AWARDS

- 7th Challenge Cup National College Student Business Plan Competition (School-level **1st Prize**)
- 8th National Biomedical Engineering Innovation Design Competition for College Students (Provincial **3rd Prize**, National **3rd Prize**)
- Outstanding cadres of Student Association of School of Information Science and Technology
- Summer 2025 BME Research Excellence Award of Carnegie Mellon University