Wangshu Zhu

(+1) 646-240-9849 GoogleScholar Homepage GitHub wz2708@columbia.edu

Education

Columbia University **Sep 2024 – May 2026**

Master of Science in Electrical Engineering GPA: 4.04/4.00 *New York, US*

- Awards: 2025 Spring MS Honors Students, Advanced Master's Research Specialization
- Courses: Deep Learning & Neural Network, Reinforcement Learning, Generative AI, Advanced Big Data & AI

Wuhan University **Sep 2020 – May 2024**

Bachelor of Science in Information Engineering GPA: 3.71/4.00 *Wuhan, China*

- Courses: Data Structure, Python Programing, Java Programing, Applied Machine Learning, Artificial Intelligence

Selected Publications and Manuscripts

- **A. Jaber, W. Zhu**, T. Zheng, et al. *AutoClimDS: Climate Data Science Agentic AI—A Knowledge Graph is All You Need. arXiv preprint arXiv:2509.21553, 2025.* Under review at *Nature Machine Intelligence.*

Research Experience

AutoClimDS: Knowledge-Grounded Multi-Agent System **May 2025 – Present**

LEAP Lab, Columbia University — Advised by Prof.Tian Zheng *Research Assistant*

- Addressed the challenge of complex research reasoning by developing a modular multi-agent framework capable of executing end-to-end workflows, from natural-language intent parsing to data acquisition and model-driven analysis.
- Developed a KG-grounded, ReAct-orchestrated reasoning backbone for adaptive tool use; fine-tuned a domain-adapted transformer (ClimateBERT) achieving 99% semantic linking in the metadata-to-variable mapping.
- Implemented specialized agents (Discovery, Retrieval, Modeling, Verification) collaborating via shared KG memory and persistent state tracking, demonstrating reproducible, intent-aware multi-agent coordination ability.
- Validated autonomous reasoning by reproducing complex data-driven analyses from natural-language prompts, establishing the system as a generalizable blueprint for agentic research automation.

SegmentFusion: Unified AR–MDM Fusion Training for LLM **Sep 2025 – Present**

CS Department, Rutgers University — Advised by Prof.Chengzhi Mao *Research Assistant*

- Proposed a unified training framework bridging autoregressive (AR) and masked diffusion modeling (MDM) paradigms within a single sequence, enabling controllable transition between causal and bidirectional generation.
- Devised a segment-level dynamic masking scheme partitioning sequences into AR and MDM regions with causal masks, allowing the model to learn both forward drafting and backward revision within one optimization loop.
- Unified a joint AR–MDM optimization objective that combines teacher-forced NLL and diffusion-based reconstruction losses, ensuring stable convergence and interpretable cross-task signal separation.
- Demonstrated a novel "write-then-revise" reasoning behavior, integrating AR fluency with MDM coherence toward self-reflective, reasoning-capable LLMs—a step toward unified cognitive modeling.

BioTrack: Self-Supervised Behavioral State Tracking Framework **Oct 2025 – Present**

IBL Lab, Columbia University — Advised by Prof.Liam Paninski *Research Assistant*

- Developed an uncertainty-aware tracking framework for animal behavior using SAM2 and YOLO-based detection, integrated uncertainty-aware segmentation to enable self-corrective tracking restarts under low-confidence.
- Designed a data-driven augmentation scheme generating occluded-frame reconstructions to fine-tune vision backbones, improving pose stability and identity preservation across behavioral states.
- Extending BEAST's spatiotemporal embedding framework toward integration with tracking-based behavioral segmentation, enabling closed-loop and interpretable neural-behavioral modeling (BEAST: arXiv:2509.21553).

FairTraj: Trajectory Prediction under Missingness Protocols **May 2025 – Aug 2025**

AIDL Lab, Columbia University — Advised by Prof.Zoran Kostić *Research Assistant*

- Identified a systemic bias in trajectory prediction benchmarks, where model performance is confounded by how missing tracks are reconstructed rather than by intrinsic predictive capability.
- Developed the COSMOS dataset pipeline, extracting long-horizon NYC traffic trajectories, converting them into the nuScenes-compatible schema, and implementing four principled missingness protocols for controlled benchmarking.
- Standardized experimental design by employing Unitraj as a unified training spine, enabling controlled comparisons of AutoBot, Wayformer, and MTR models, with Brier-FDE as a calibration-aware complement to FDE.
- Delivered quantitative evidence that Protocol of Train-Rich & Eval-Pure consistently yields the best accuracy–calibration trade-off, revealing bias in standard trajectory prediction benchmarks.

Interpretable Multimodal Framework for Collision Prediction **Jan 2025 – May 2025**

DitecT Lab, Columbia University — Advised by Prof.Sharon Di *Research Assistant*

- Formulated an interpretable multimodal reasoning framework for early collision prediction, repurposing Qwen2.5-VL as a unified vision–language model for temporal risk anticipation.
- Constructed a structured visual grounding pipeline integrating Grounding DINO, SAM2, VideoDepthAnything, and RAFT, enabling explicit geometric and motion decomposition for human-auditable reasoning.
- Implemented LoRA-based fine-tuning under 4-bit quantization, coupling cross-entropy and regression objectives to jointly optimize collision prediction, achieving 94% precision, 46% recall while providing rationales for prediction.
- Demonstrated that structured multimodal prompting enables large vision–language models to generalize from perception to temporal causal reasoning, advancing explainable visual reasoning for autonomous systems.

Internship Experience

Clinical Multi-Agent: Toward Trustworthy Clinical Decision Support **May 2025 – Aug 2025**

Graphen Inc. New York *Data Science Research Intern*

- Tackled real-world diagnostic uncertainty from patient narratives, where single-agent LLMs often misclassify due to reasoning shortcuts and lack of evidence grounding, motivating a multi-agent, knowledge-aware framework.
- Designed a dual-path retrieval strategy mirroring physician workflows—integrating UMLS-backed structured guidelines and historical case analogies to enhance evidence grounding and diagnostic realism.
- Introduced a Reception module for interpretable symptom extraction via trainee-style decomposition, and a Critic–Expert debate mechanism to iteratively validate lever symptoms, reducing diagnostic ambiguity
- Achieved 96.5% Top-1 accuracy and 0.98 MRR on held-out test sets, surpassing single-agent (73.9%) and CoT baselines (77.1%) while maintaining interpretability; submitted to AMIA 2025 Late-Breaking Paper track.

Teaching Assistant Experience

ECBM E4040: Neural Networks and Deep Learning **Sep 2025 — Dec 2025**

Department of Electrical Engineering, Columbia University - Taught by Prof.Zoran Kostić

- Assisted in a graduate-level deep learning course covering various architectures with over 90 enrolled students.
- Refined and graded assignments and exams on optimization, forward and backward propagation mechanisms.
- Provided mentorship for course assignments and final projects, guiding students to reproduce and extend papers.

Technical Skills

Python, PyTorch, TensorFlow; LangChain, HuggingFace, Gradio; SQL; Java; Overleaf