Research Statement

Shuyue Jia, 4th year Ph.D. in Computer Engineering, Boston University brucejia@bu.edu

My research is dedicated to advancing fundamental and frontier autonomous AI toward the development of safe, reliable, extensible, and cost-efficient artificial general intelligence (AGI) systems for science and medicine. My research centers on three interconnected topics. The first focuses on developing trustworthy multimodal, multilingual, and multidomain foundation models, together with human-AI collaboration frameworks, to bridge the gap between frontier AI research and real-world scientific and clinical impact. The second investigates autonomous AI systems that improve accuracy and reliability while reducing computational cost through grounding intelligence in real-world evidence. The third explores accurate and responsive electroencephalography (EEG)-based brain-computer interface (BCI) systems by leveraging graph representation learning and network analysis to model brain dynamics. This statement summarizes these topics and outlines my future research agenda. Collectively, these efforts represent an important step toward my long-term vision of building next-generation intelligent systems that translate frontier AI advances into real-world impact across science, medicine, healthcare, education, robotics, and human-machine interaction.

Topic 1: Trustworthy multimodal, multilingual, and multidomain foundation model
Large foundation models have the potential to transform science, medicine, and education by enabling human-AI collaboration within real-world workflows. However, current multimodal AI systems remain limited by hallucinations, unreliable reasoning, and insufficient domain specialization, restricting their deployment in high-stakes scientific and clinical settings. Frontier models, in particular, struggle to interpret complex medical imaging data, such as magnetic resonance imaging (MRI) and computed tomography (CT), because they are primarily optimized for general-purpose tasks. To address these challenges, my research advances scalable learning frameworks for large language models (LLMs) and large vision-language models (LVLMs) that integrate real-world evidence, domain-specific knowledge, and human-AI collaboration frameworks to enable reliable, specialized, and deployable AI systems. In [1, 2], we introduced PodGPT, a continual pre-trained LLM enhanced with informative, up-to-date multilingual podcasts spanning science, technology, engineering, mathematics, and medicine (STEMM), together with a retrieval-augmented generation (RAG) framework that grounds model responses in emerging scientific literature. We further evaluated these methods on the nephrology self-assessment program to systematically characterize their strengths and limitations in a specialized clinical domain [3]. In [4], we developed ReMIND (Radiology-encoded Multimodal Interpretation for Neurological Disorders), a vision-language learning framework for comprehensive multi-sequence and multi-volumetric brain MRI analysis. ReMIND combines large-scale clinically grounded video instruction tuning with targeted supervised fine-tuning for radiology report generation, enabling clinically validated specialized intelligence for high-stakes medical imaging. The platform has been adopted by hundreds of physicians, radiologists, neurologists, and researchers across more than 50 cities worldwide. We further extended this line of research by introducing an innovative methodology to generate high-resolution 3D lung CT images from textual prompts and anatomical priors, demonstrating the versatility of multimodal understanding across imaging tasks [5]. Beyond image and video understanding and generation, we also developed vision models for image quality assessment, further expanding the applicability of AI [6, 7]. These studies represent a significant step toward translating large multimodal, multilingual, and multidomain foundation models into trustworthy AI systems that support precision science and medicine, through evidence-grounded text, image, and video understanding, generation, and quality assessment.

Topic 2: Grounded and cost-efficient autonomous AI system

Autonomous AI systems are rapidly emerging as a foundational technology for manufacturing and digital infrastructure, with the potential to drive future economic growth and national security. Despite their remarkable benchmark performance, current frontier models remain limited in solving complex real-world tasks due to limited context window, insufficient multimodal reasoning, and increasing cost associated with extended chain-of-thought reasoning. My research addresses these challenges by designing grounded and cost-efficient autonomous AI systems that integrate memory, tool harness, and evidence grounding to improve accuracy, reliability, and computational efficiency. In [8], we introduced a multi-agent framework with a task-agnostic memory module inspired by human reading strategies for long-context document understanding. Compared with optical character recognition (OCR)-based approaches, our framework achieved significant performance improvements of 21.70% and 12.40% over `GPT-4o` and `Claude 3.5 Sonnet`, respectively, while reducing token consumption and inference cost. These results demonstrate the effectiveness of memory-augmented autonomous agents for scalable and efficient multimodal document understanding. Extending this research, in [9], we developed a unified open-source agentic RAG framework that integrates document retrieval, reranking, memory augmentation, evidence grounding, and diagnosis generation into a coherent multi-step reasoning pipeline for both open-ended and closed-form medical question-answering (QA). We further introduced a cache-and-prune memory bank mechanism that efficiently retains relevant evidence over long reasoning horizons, improving both diagnostic accuracy and computational efficiency on challenging medical QA tasks. Across five major medical QA benchmarks, our framework achieved superior or competitive performance compared with standalone LLMs, including `GPT-4`. In particular, it achieved 82.98% accuracy on the United States medical licensing examination (USMLE) Step 1 (*vs.* 80.67% for `GPT-4`) and 86.24% on USMLE Step 2 (*vs.* 81.67% for `GPT-4`), while closely matching `GPT-4` on Step 3. Collectively, these works demonstrate that combining agentic orchestration, tool harness, memory, and evidence grounding substantially improves the accuracy and reliability while reducing computational cost of autonomous AI systems for long-context multimodal understanding and high-stakes clinical reasoning.

Topic 3: Accurate and responsive EEG-based BCI system

EEG-based BCIs offer a non-invasive, portable, and high-temporal-resolution approach for direct human-machine interaction, restoring communication and assistive control (*e.g.*, wheelchair control) for individuals with paralysis, stroke, and other severe motor impairments. However, practical EEG-based BCIs remain challenged by substantial inter-subject variability and limited cross-subject generalization, and the need to simultaneously achieve high decoding accuracy and low response latency. Moreover, traditional methods often overlook the functional topology of brain networks, limiting their ability to capture complex neural interactions [10]. To address these challenges, we developed a pioneering graph-based deep learning method that explicitly models the functional relationships among EEG electrodes for robust neural decoding of motor imagery (MI) tasks [11]. Building on this foundation, we proposed a highly accurate and responsive MI recognition framework capable of decoding scalp EEG recordings as short as 0.4 seconds [12]. Our approach demonstrated strong robustness to inter-trial and inter-subject variability while maintaining low-latency inference, representing an important step toward practical, real-time EEG-based BCI systems. To accelerate research in this area, we also open-sourced a deep learning library `EEG-DL` for EEG signal classification that has become a *de facto* standard for EEG deep learning, receiving more than 1.2K GitHub stars and supporting both academic research and industrial applications. Together, these efforts advance the translation of EEG-based MI recognition from laboratory research to practical, real-world BCI systems.

Future work

Despite major advances in general-purpose AI, achieving systems that are safe, reliable, extensi-

ble, and computationally efficient remains a fundamental challenge. Although specialized models, such as those for MRI and CT, have achieved impressive performance across benchmarks, their ability to generalize across heterogeneous clinical modalities and support diverse downstream tasks remains understudied, particularly in real-world clinical settings. Moreover, most existing clinical foundation models are designed around data from a single patient encounter, limiting their capability to leverage longitudinal healthcare trajectories, reason over multiple patient visits, and remain robust to incomplete or evolving clinical information. Likewise, solving complex real-world problems increasingly requires large numbers of specialized agents that can collaborate to learn, reason, communicate, and act under dynamic and uncertain environments, yet existing agentic AI systems remain largely confined to limited-scale collaboration on digital tasks. Furthermore, translating EEG-based BCIs from research prototypes into practical products requires hardware/software co-design that jointly addresses hardware constraints, real-time signal acquisition, and system-level optimization, whereas current research primarily emphasizes algorithm development. Addressing these challenges is essential for transforming state-of-the-art AI research into trustworthy and deployable systems with measurable real-world impact.

Aim 1: Unified medical imaging foundation model with data harmonization
My previous work established a scalable vision-language learning framework for comprehensive multi-sequence and multi-volumetric brain MRI analysis. While this work demonstrated the potential of building clinically grounded LVLMs for neuroimaging, existing medical AI systems remain largely domain-, modality-, and task-specific, limiting their ability to jointly reason across heterogeneous clinical information, including patient demographics, family medical history, laboratory tests, medications, neuropsychological assessments, genetic and molecular biomarkers, electronic health records (EHRs), EEG, ECG, and medical imaging, such as MRI, CT, X-ray, whole-slide imaging (WSI), and positron emission tomography (PET). Building on this foundation, my future research will leverage multimodal imaging data from diverse research and clinical cohorts to develop a scalable omnimodal medical foundation model that learns a shared representation across heterogeneous clinical modalities while supporting a broad spectrum of downstream tasks, including diagnosis, QA, report generation, lesion localization, and tumor segmentation. Beyond representation learning, I will investigate medical data harmonization, continual learning, post-training methodologies, tool harness, and autonomous skill acquisition to enable end-to-end medical AI systems that continually expand their capabilities while remaining accurate, efficient, and clinically deployable.

Aim 2: Clinical foundation model for disease progression and personalized care
Most existing clinical foundation models are developed under isolated patient encounters, whereas real-world healthcare is inherently longitudinal, with patients receiving care across multiple visits over months or years. As a result, current models often fail to capture disease progression and treatment response from an evolving clinical context. Furthermore, longitudinal healthcare records naturally contain heterogeneous, incomplete, and long-context clinical information, presenting both a challenge and an opportunity to develop more robust clinical foundation models capable of temporal reasoning under long, incomplete, or evolving clinical evidence. My prior research established a foundation for comprehensive context understanding from individual encounters. My future research aims to develop a longitudinal clinical foundation model that learns from event-level healthcare trajectories spanning multiple clinical visits. By jointly modeling longitudinal patient history, imaging, temporal representation learning, cross-temporal reasoning, EHR integration, data harmonization, and incomplete clinical observations, this framework will enable accurate outcome prediction and personalized clinical decision-making, including diagnosis, prognosis, treatment planning, management, and medication recommendation, while remaining robust to missing and evolving patient information.

Aim 3: Agentic swarm technology with embodied intelligence
Current agentic AI is predominantly designed around single-agent reasoning or small-scale multi-agent collaboration, restricting their ability to solve complex real-world tasks that require large numbers of specialized agents to divide and conquer, while jointly perceiving, learning, reasoning, communicating, and acting within dynamic and uncertain environments. My previous work on single-agent RAG systems and collaborative multi-agent frameworks provides a strong foundation for addressing these challenges. Leveraging these advances, I will investigate grounded agentic swarm intelligence that enables large numbers of AI agents to coordinate, collaborate, and adapt across both digital and physical environments, with applications ranging from multi-robot collaboration to coordinated unmanned aerial vehicle (UAV) systems. By integrating AI grounding, embodied intelligence, tool use, multi-agent communication, and hierarchical task decomposition, I seek to build scalable autonomous systems capable of solving complex real-world tasks with greater reliability, efficiency, robustness, and adaptability.

Aim 4: Hardware/software co-design for EEG-based BCI system
My prior work has significantly improved the accuracy and responsiveness of EEG-based AI models. However, translating these algorithmic advances into practical BCI systems requires jointly addressing hardware constraints, real-time signal acquisition, and system-level optimization. My future research will investigate hardware/software co-design for EEG-based BCIs by jointly optimizing AI algorithms, sensing hardware, and edge computing platforms. I aim to bridge the gap between algorithm development and real-world deployment, enabling robust, low-latency, energy-efficient, and deployable neurotechnology.

Long-term vision
AI will become trustworthy not simply by scaling model size, but by grounding intelligence in real-world evidence, domain knowledge, memory, and interaction with humans and the environment. Over the next five years, I aspire to advance the foundations of safe, reliable, extensible, and cost-efficient AGI systems that can deliver meaningful real-world impact across science, medicine, healthcare, education, robotics, and human-machine interaction.

References

[1] **Jia, Shuyue**, Subhrangshu Bit, Edward Searls, Lindsey A. Claus, Pengrui Fan, Varuna H. Jasodanand, Meagan V. Lauber, Divya Veerapaneni, William M. Wang, Rhoda Au, et al. MedPodGPT: A multilingual audio-augmented large language model for medical research and education. *medRxiv*, 2024.

[2] **Jia, Shuyue**, Subhrangshu Bit, Edward Searls, Meagan V. Lauber, Pengrui Fan, William M. Wang, Lindsey A. Claus, Varuna H. Jasodanand, Divya Veerapaneni, Rhoda Au, et al. PodGPT: An audio-augmented large language model for research and education. *Nature npj Biomedical Innovations*, 2(1):26, 2025.

[3] Meysam Ahangaran, **Jia, Shuyue**, Shlok Chitalia, Ambarish Athavale, Jean M. Francis, Michael William O'Donnell, Santhoshi Rupa Bavi, Uma Datta Gupta, and Vijaya B. Kolachalama. Performance of open-source large language models on nephrology self-assessment program. *medRxiv*, 2026.

[4] Diala Lteif, **Jia, Shuyue**, Subhrangshu Bit, Artem Kaliaev, Asim Z. Mian, Juan E. Small, Balamurugan Mangaleswaran, Bryan A. Plummer, Sarah A. Bargal, Rhoda Au, et al. Vision-language framework for multi-sequence brain magnetic resonance imaging. *medRxiv*, 2026.

[5] Yanwu Xu, Li Sun, Wei Peng, **Jia, Shuyue**, Katelyn Morrison, Adam Perer, Afrooz Zandifar, Shyam Visweswaran, Motahhare Eslami, and Kayhan Batmanghelich. MedSyn: Text-guided anatomy-aware synthesis of high-fidelity 3-D CT images. *IEEE Transactions on Medical Imaging*, 43(10):3648–3660, 2024.

[6] **Jia, Shuyue**, Baoliang Chen, Dingquan Li, and Shiqi Wang. No-reference image quality assessment via non-local dependency modeling. In *24th IEEE International Workshop on Multimedia Signal Processing, MMSP 2022, Shanghai, China, September 26-28, 2022*, pages 1–6. IEEE, 2022.

[7] Zhaopeng Feng, Keyang Zhang, **Jia, Shuyue**, Baoliang Chen, and Shiqi Wang. Learning from mixed datasets: A monotonic image quality assessment model. *Electronics Letters*, 59(3):e12698, 2023.

[8] Li Sun, Liu He, **Jia, Shuyue**, Yangfan He, and Chenyu You. DocAgent: An agentic framework for multi-modal long-context document understanding. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, EMNLP 2025, Suzhou, China, November 4-9, 2025*, pages 17701–17716. Association for Computational Linguistics, 2025.

[9] **Jia, Shuyue**, Subhrangshu Bit, Varuna H. Jasodanand, Yi Liu, and Vijaya B. Kolachalama. Agentic memory-augmented retrieval and evidence grounding for medical question-answering tasks. *International Journal of Medical Informatics*, page 106339, 2026.

[10] Yimin Hou, Lu Zhou, **Jia, Shuyue**, and Xiangmin Lun. A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. *Journal of Neural Engineering*, 17(1):016048, 2020.

[11] **Jia, Shuyue**, Yimin Hou, Xiangmin Lun, Ziqian Hao, Yan Shi, Yang Li, Rui Zeng, and Jinglei Lv. GCNs-Net: A graph convolutional neural network approach for decoding time-resolved EEG motor imagery signals. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7312–7323, 2022.

[12] **Jia, Shuyue**, Yimin Hou, Xiangmin Lun, Shu Zhang, Tao Chen, Fang Wang, and Jinglei Lv. Deep feature mining via the attention-based bidirectional long short term memory graph convolutional neural network for human motor imagery recognition. *Frontiers in Bioengineering and Biotechnology*, 9:706229, 2022.