From Quantum Computing to Large Language Models: Recent Advances and Results

Tuesday, 18 March 2025 14:30 (30 minutes)

Quantum Machine Learning (QML) faces significant challenges, particularly in encoding classical data and the reliance on quantum hardware for inference, limiting its practical applications. Meanwhile, classical large language models (LLMs) demand immense computational resources and exhibit low training efficiency, leading to substantial cost and scalability concerns. This talk will introduce Quantum Parameter Adaptation (QPA), a research work recently accepted at the top-tier AI conference ICLR 2025. QPA leverages quantum neural networks (QNNs) to generate parameters during training, while inference remains entirely classical. Applied to LLM fine-tuning, QPA significantly reduces the number of trainable parameters while maintaining or even improving performance, making fine-tuning large-scale models more efficient. By bridging quantum computing with large language models, this approach highlights how quantum technology can enhance modern AI, positioning it as a key enabler for the future of intelligent computing.

Presenter: LIU, Chen-Yu (NTU)

Session Classification: Hybrid Quantum Computing Workshop - I