AI for Quantum at the Quantum Science Center: Towards Building a Connective Layer Across the Quantum Computing Stack

Agentic Software Team, Quantum Science Center (QSC)

Tirthankar Ghosal[1], Sharmin Afrose[1], Ethan Seefried[1], Nick Stapleton[2], Nirajan Banjade[3], Romit Ghosh[4], Sangram Deshpande[5], Vicente Leyton-Ortega[1], Travis Humble[1]

[1]Oak Ridge National Laboratory, [2]University of Tennessee, Knoxville, [3]Texas State University, San Marcos, [4]Rutgers University, New Brunswick, [5]North Carolina State University

Unifying thesis: artificial intelligence can serve as the connective and accelerating substrate of the quantum stack; spanning knowledge curation, scientific reasoning, and HPC-integrated execution.

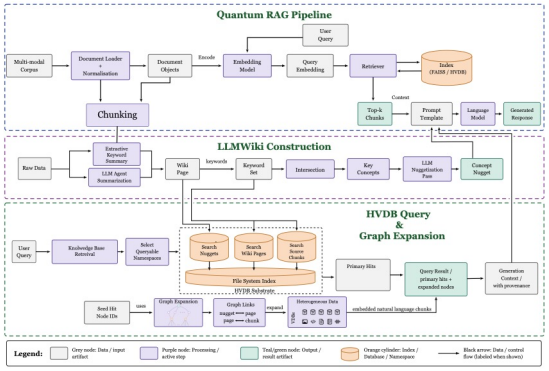
Knowledge Infrastructure

Quantum RAG

- Modular retrieval-augmented generation framework for quantum computing literature
- Configurable ingestion pipeline: text extraction, figure captioning, embedding, and vector indexing
- Models and providers are swappable at every stage

Quantum Computing LLMWiki

- Extends Quantum RAG with a curated, machine-readable knowledge base
- Indexes wiki pages and structured concept "nuggets" alongside the literature corpus
- Adds semantic context that strengthens AI-assisted search and reasoning



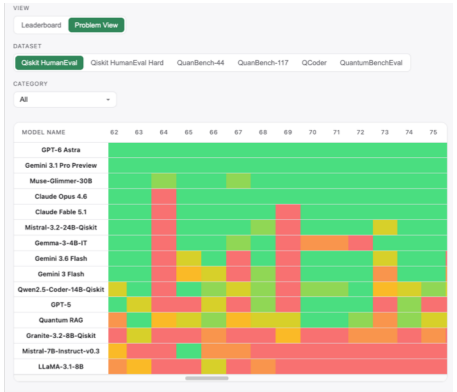
Evaluation & Benchmarking

Quantum Computing Leaderboard

- Standardized benchmark for evaluating frontier models on quantum code generation
- Automated pipelines measure correctness, execution, and model performance
- Enables reproducible comparisons and longitudinal tracking

IEEE QCE'26 shared-task benchmarks

- Open benchmarks that evaluate complete AI-assisted quantum workflows, not isolated models
- Representative tasks in optimization, simulation, and graph problems
- Scored on solution quality, quantum resource use, reproducibility, and transparent AI use



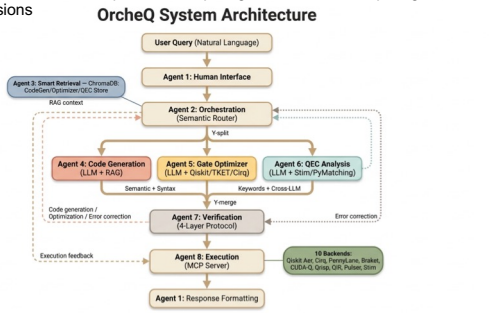
Domain Reasoning & Quantum Systems

Multi-agent framework

Contribution
- First end-to-end multi-agent system spanning quantum code generation, optimization, and error correction
- Built-in verification at each stage

Workflow
- Intent-based routing to specialist crews (CodeGen, Optimizer, QEC)
- Cross-LLM critique followed by real execution

Impact
- Lowers the expertise barrier to quantum programming
- Accelerates DOE's quantum computing and scientific computing missions

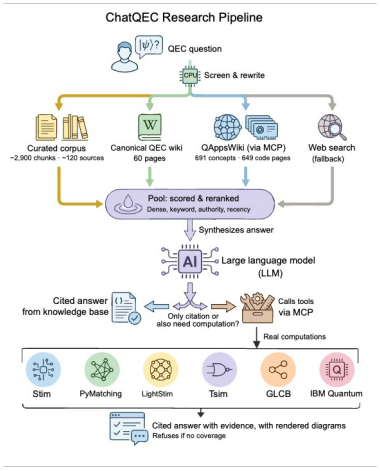


ChatQEC

Contribution
- RAG assistant for quantum error correction
- Answers grounded in a full-text corpus
- Runs tools such as Stim and PyMatching
- Maps circuits onto real IBM and IQM qubits

Workflow
- Sources ingested and embedded offline
- Each question is classified, rewritten, retrieved, and synthesized
- Tool calls, web-search fallback, and live layout calibration

Impact
- Cited answers with concrete circuit placement
- Better qubit selection on current hardware



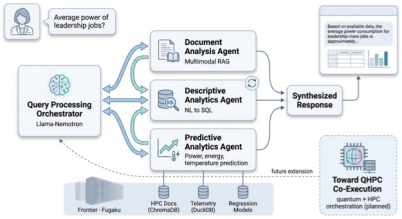
HPC Orchestration & QHPC Co-Execution

EPIC

Contribution
- LLM-orchestrated multi-agent framework for HPC operational analytics
- Replaces rigid, batch-oriented pipelines with on-demand natural-language queries

Workflow
- Orchestrator (Llama-Nemotron) routes queries to three fine-tuned agents: document RAG, NL-to-SQL analytics, and predictive regression
- Agent outputs are synthesized into a single response

Impact
- Fine-tuned agents outperform GPT-4o on SQL generation (87.4% vs. 69.3%) at 19x lower cost
- Generalizes across Frontier and Fugaku with no architecture changes
- Now being extended toward QHPC co-execution



This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE)
This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Science Center