QC-BENCH: WHAT DO LANGUAGE MODELS KNOW ABOUT QUANTUM COMPUTING?

Anonymous authors
Paper under double-blind review

ABSTRACT

Language models have become practical tools for quantum computing education and research, from summarizing technical papers to explaining theoretical concepts. While existing benchmarks evaluate quantum code generation and circuit design, their understanding of quantum computing concepts has not been systematically measured. QC-Bench addresses this gap with over 6,000 expert-level questions on quantum algorithms, error correction, and security protocols. Evaluating 31 models from OpenAI, Anthropic, Google, and Meta reveals strong performance on established theory but systematic failures on advanced topics like quantum security and recent attack vectors. Human participants scored between 23% and 86%, with experts averaging 74% and all participants averaging 57%. Top-performing models exceeded the expert average, with Claude Sonnet 4 and GPT-5 reaching 88% overall, yet dropping to 76% on security questions. Additional evaluation across question formats and languages reveals variation in model performance, demonstrating that QC-Bench provides a necessary framework for measuring language model reliability in quantum computing contexts.

1 INTRODUCTION

Quantum computing has progressed significantly from theoretical research to experimental implementations with practical applications. Current quantum systems have rapidly evolved through successive technological breakthroughs from operating with just a few qubits to recently surpassing the 1000-qubit barrier AbuGhanem (2025), enabling exploration of quantum algorithms and protocols that were previously confined to theoretical analysis. This technical advancement drives progress in quantum simulation King et al. (2025); Halimeh et al. (2025); Puig et al. (2025), optimization problems Quinton et al. (2025); Phillipson (2024), and cryptographic applications Sahu & Mazumdar (2024); Ralegankar et al. (2021); Kalaivani et al. (2021). Beyond traditional quantum applications such as quantum simulation and cryptography, recent research explores its potential in finance Innan et al. (2024); Grossi et al. (2022), healthcare Ur Rasool et al. (2023); Flöther (2023), computer vision Li et al. (2020); Afane et al. (2025); ALRikabi et al. (2022), and wireless communication Narottama & Shin (2021); Narottama et al. (2023), among other promising real-world applications.

In parallel, Large Language Models (LLMs) have become sophisticated tools that address complex challenges across many disciplines. These AI systems now approach or exceed human expert performance in areas such as cybersecurity Tihanyi et al. (2024); Afane et al. (2024), medical diagnosis Subedi (2025), and legal reasoning Guha et al. (2023); Kant et al. (2025). As these two fields continue to evolve, their intersection becomes increasingly important for scientific communication, education, and research productivity. Despite significant advances in both domains, we face a critical knowledge gap in evaluating LLMs' understanding of specialized quantum concepts. While extensive benchmarking exists across numerous related domains, including mathematics Gao et al. (2024); Fang et al. (2024), physics Chung et al. (2025), and computer science Song et al. (2024), no standardized frameworks comprehensively assess quantum computing knowledge in these models. This absence is particularly concerning given the field's counterintuitive principles, and rapidly evolving terminology that challenge even domain experts. The complexity of quantum computing concepts, combined with their inherent mathematical abstraction, creates a particularly demanding test case for evaluating the depth of LLMs' specialized knowledge. Without reliable evaluation metrics, LLMs risk spreading plausible but incorrect quantum information to educational and research communities, as hallucinations, reasoning errors, and factual inaccuracies have been widely documented in similarly complex and technically demanding specialized domains. Orgad et al. (2024); Perković et al. (2024).

This creates an urgent need for robust quantum computing benchmarks as researchers, students, and industry professionals increasingly rely on these models for information and assistance with quantum tasks. The growing adoption of LLMs across academic institutions and quantum technology companies further amplifies the importance of ensuring these systems provide accurate information on this emerging field. To address these challenges, we present the following key contributions:

- We assemble **6,237 questions**: 5,400 multiple-choice questions comprising QC1000 (1,000 entirely human-authored from peer-reviewed literature, with QC500 translated into Spanish and French) and 4,400 human-validated questions filtered from 8,686 candidates, plus 837 format variants (416 true/false, 421 open-ended) for evaluating model performance across question formats.
- We conduct extensive evaluation across **31 models** from leading AI research organizations including OpenAI, Anthropic, Google, Meta, IBM, Microsoft, and DeepSeek, among others. We compare their performance against 43 quantum computing experts and practitioners to establish human baselines and assess how LLMs perform relative to human capabilities.
- We analyze model performance across different question formats and via Spanish and French translations of QC500, revealing significant accuracy declines in the translated sets and consistent sensitivity to question type, with larger drops in Spanish than in French.
- We explore the potential of our dataset for fine-tuning by using a subset of 4,000 questions to enhance the quantum knowledge of five smaller models, demonstrating performance improvements and establishing the benchmark's value beyond evaluation.

2 Related Work

Despite significant advancements in both quantum computing and LLMs, their intersection remains surprisingly underexplored. Recent research has begun addressing this gap from different angles. Kashani Kashani (2024) introduced QuantumLLMInstruct (QLMMI), a dataset of over 500,000 instruction-problem pairs covering quantum cryptography, spin chain models, and Trotter-Suzuki decompositions. However, QLMMI's primary purpose is to enable instruction fine-tuning rather than comprehensive evaluation of quantum knowledge. While extensive in size, QLMMI relies entirely on synthetically generated content through a four-stage LLM pipeline. In contrast, QC-Bench offers 1,200 human-authored evaluation questions extracted directly from research literature published over four decades, prioritizing authentic scientific content over synthetic generation. Wang et al. Wang et al. (2024) introduced GroverGPT, an approach to simulating quantum algorithms using LLMs. Their 8-billion-parameter model is fine-tuned to approximate Grover's quantum search algorithm without explicitly representing quantum states. While GroverGPT demonstrates impressive capabilities in predicting specific quantum circuit outputs, it focuses exclusively on a single quantum algorithm rather than evaluating comprehensive knowledge across the quantum computing domain.

Complementary efforts have emerged focusing on quantum code generation and circuit implementation capabilities. Vishwakarma et al. Vishwakarma et al. (2024) developed Qiskit HumanEval, a hand-curated benchmark of over 100 tasks designed to evaluate LLM performance in generating executable quantum code using the Qiskit SDK, complete with canonical solutions and comprehensive test cases. Guo et al. Guo et al. (2025) introduced QuanBench, which evaluates quantum code generation across 44 programming tasks using both functional correctness (Pass@K) and quantum semantic equivalence (Process Fidelity) metrics, finding that current LLMs achieve below 40% overall accuracy with frequent semantic errors including outdated API usage and incorrect algorithm logic. Yang et al. Yang et al. (2024) presented QCircuitNet, a large-scale hierarchical dataset for quantum algorithm design containing 120,290 data points with automatic syntax and semantic verification functions. At a lower abstraction level, Li et al. Li et al. (2023) developed QASMBench, a benchmark suite of low-level OpenQASM programs for evaluating NISQ devices and simulators. While these works provide valuable resources for assessing programming proficiency and implementation capabilities at various levels of quantum software development, they primarily target coding skills rather than evaluating deep conceptual understanding of quantum computing principles, algorithmic theory, or the ability to reason about quantum phenomena. QC-Bench addresses this by evaluating theoretical knowledge and conceptual understanding across quantum computing topics, from foundational algorithmic principles to advanced security protocols and attack vectors.

3 QC-BENCH DATASET

We constructed the QC-Bench dataset to evaluate quantum computing knowledge in LLMs across a wide range of topics and difficulty levels. To ensure comprehensive coverage and relevance, our team reviewed over 200 peer-reviewed research papers, preprints, and academic resources. From these sources, questions were directly selected to reflect both foundational knowledge and current advancements in the field. The dataset comprises QC1000, containing 1000 questions manually extracted from quantum computing literature, with QC500 as a 500-question subset selected for multilingual evaluation. To address concerns about model memorization, none of these questions are reproduced verbatim from source materials; instead, we extracted core concepts and reformulated them into original questions. This approach ensures that performance reflects genuine understanding rather than memorization of published text. After refining and validating this content, the benchmarks were finalized. The QC500 subset was translated into Spanish and French to evaluate LLM performance in languages other than English.

To expand our benchmark, Gemini 2.0 Flash, Gemini 1.5 Pro, GPT-4.0, and Claude 3.7 Sonnet were employed to extract additional relevant questions from the selected papers. Different prompt engineering techniques were tested to optimize question generation quality. While zero-shot prompting produced acceptable results, few-shot prompting with five carefully selected examples from the existing subsets significantly improved the relevance and technical accuracy of generated questions. This approach generated 8,686 candidate questions, subsequently filtered to remove low-quality or redundant items. The final selection included an additional 4,400 high-quality questions, bringing the total benchmark size to 5,400 multiple-choice questions. To evaluate model performance across different question formats, the benchmark was supplemented with 416 true/false questions and 421 open-ended questions. Figure 1 illustrates the distribution of these question types across different topics. Given the interconnected nature of quantum computing, some concepts naturally appear across multiple categories; for example, noise characterization relates to both error correction and hardware-level circuit design. The multiple-choice format enables precise evaluation of factual recall and conceptual understanding, while true/false questions assess binary comprehension, and open-ended questions evaluate explanatory capabilities.

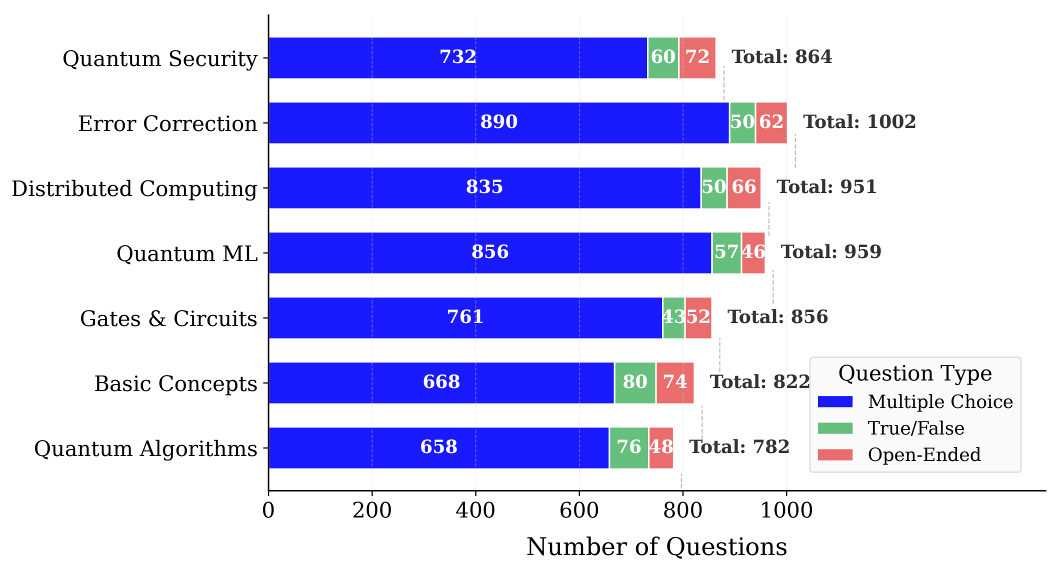

Figure 1: Breakdown of benchmark question topics and their internal composition by question type. Each horizontal bar shows the total number of questions per topic. We intentionally included a larger share of multiple choice items to enable standardized automated evaluation, whereas true/false items offer limited challenge and open-ended questions require manual scoring.

4 EXPERIMENTS

We evaluated 31 LLMs using a consistent benchmarking pipeline. Closed-source models, including GPT-5, GPT-4o (standard, mini), Claude (Sonnet 4, Sonnet 3.7, Haiku 3.5), and Gemini (1.5 Pro, 2.0 Flash), were accessed through their official APIs in Google Colab environments. Open-access models such as LLaMA3 (1B, 8B, 70B), LLaMA2 (13B), Phi (2.7B, 3.8B, 14.7B), Mistral (7B, 24B), Qwen1.5 (2.7B, MoE-A2.7B), Zephyr, DeepSeek, Gemma, Granite, and GPT-J were deployed using Hugging Face's Transformers library on a cluster equipped with two Tesla V100 GPUs (32GB each) using FP16 inference. For several larger models, including llama-3 (70b, 70b-versatile), and Gemma-9B, we used Groq's API instead of Hugging Face's Transformers library for a faster and more efficient evaluation. All models were configured with a temperature setting of 1 to balance deterministic responses with reasonable diversity in answer generation.

For experiment preparation, all benchmark questions were structured in JSON format for efficient processing and consistent evaluation across different model architectures. We developed standardized prompting templates for each question type to ensure fair comparison between models. This data preparation approach facilitated automated evaluation pipelines and ensured comparable results despite the diversity of model implementations and access methods. The benchmark includes multiple-choice, true/false, and open-ended formats, with multilingual versions available for a subset of questions. Key findings from these experiments are presented in the following subsections, with complete results and detailed analyses available in the appendix.

LLM Model	Provider	Size	Access	Q500	Q1000	Q5400
Claude Sonnet 4	Anthropic	N/A	Anthropic API	91.80	89.90	88.55
GPT-5	OpenAI	N/A	OpenAI API	91.40	90.90	88.46
GPT-4o	OpenAI	N/A	OpenAI API	88.20	86.30	88.07
Claude Sonnet 3.7	Anthropic	N/A	Anthropic API	92.40	84.70	87.98
GPT-4.1 mini	OpenAI	N/A	OpenAI API	87.20	82.30	86.42
Gemini 2.0 Flash	Google	N/A	Google API	82.40	84.60	84.44
Gemini 1.5 Pro	Google	N/A	Google API	80.20	84.80	83.92
GPT-4o-mini	OpenAI	N/A	OpenAI API	80.00	81.90	83.85
llama-3.3-70b-versatile	Meta	70B	Groq API	81.40	82.00	82.07
Phi-4-reasoning-plus	Microsoft	14.7B	HuggingFace	87.00	89.30	81.74
Claude Haiku 3.5	Anthropic	N/A	Anthropic API	80.00	82.80	80.44
granite-3.3-8b-instruct	IBM	8.17B	HuggingFace	84.20	81.10	76.07
Llama-3.1-8B-Instruct	Meta	8.03B	HuggingFace	73.80	78.40	75.75
Phi-4-reasoning	Microsoft	14.7B	HuggingFace	81.00	80.20	75.59
GPT-4.1 nano	OpenAI	N/A	OpenAI API	86.00	86.20	74.58
zephyr-7b-beta	Hugging Face	7.24B	HuggingFace	84.00	83.00	73.70
DeepSeek-R1-Dist-Llama-8B	DeepSeek	8.03B	HuggingFace	78.00	85.20	73.62
gemma2-9b-it	Google	9B	Groq API	84.60	86.40	73.55
DeepSeek-R1-Dist-Qwen-7B	DeepSeek	7.62B	HuggingFace	78.20	86.90	72.51
Llama-3.1-8B	Meta	8B	HuggingFace	81.00	79.50	72.51
Mistral-7B-Instruct-v0.3	Mistral AI	7.25B	HuggingFace	82.00	80.90	72.43
Phi-4-mini-reasoning	Microsoft	3.84B	HuggingFace	72.00	69.10	72.40
llama3-70b	Meta	70B	Groq API	84.20	82.30	71.85
Llama-2-13b-chat-hf	Meta	13B	HuggingFace	86.40	89.10	71.79
Llama-3.2-1B-Instruct	Meta	1.24B	HuggingFace	82.20	86.00	71.55
gemma-7b	Google	7B	HuggingFace	72.80	74.30	69.70
phi-2	Microsoft	2.7B	HuggingFace	81.20	78.50	67.85
gemma-2-2b-it	Google	2.61B	HuggingFace	74.20	60.30	62.29
Qwen1.5-MoE-A2.7B	Qwen	14.3B	HuggingFace	74.00	61.70	60.74
EleutherAI/gpt-j-6b	EleutherAI	6B	HuggingFace	72.00	60.90	50.14
dolly-v1-6b	Databricks	6B	HuggingFace	36.80	34.30	48.29

Table 1: Evaluated language models with provider, size, access method, and accuracy on QC500, QC1000, and QC5400. Rows shaded in green mark the highest performing models overall, and rows shaded in light blue mark the best performing open-source models.

4.1 COMPREHENSIVE MODEL EVALUATION ON CORE BENCHMARK AND ACROSS TOPICS

Table 1 details the characteristics of each evaluated model and summarizes performance across the three benchmark subsets. Results from these experiments demonstrate that increasing dataset size from 500 to 5,400 questions does not substantially impact relative model performance.

Model	Error Correction	Quantum Algorithms	Quantum Security
Claude Sonnet 4	**92.81**	81.76	**76.09**
GPT-5	92.13	**82.30**	75.82
GPT-4o	92.02	79.18	75.68
Claude Sonnet 3.7	91.12	79.03	75.00
GPT-4.1 mini	90.67	77.51	74.73
Gemini 1.5 Pro	88.99	77.05	73.36
Gemini 2.0 Flash	89.66	76.14	73.09
GPT-4o-mini	92.02	84.27	72.95
Claude Haiku 3.5	83.71	74.16	71.17
llama3-70b	82.13	74.01	70.63
Phi-4-reasoning-plus	81.01	82.08	69.95
llama-3.3-70b-versatile	79.89	79.39	69.95
GPT-4.1 nano	79.10	69.89	68.99
granite-3.3-8b-instruct	77.64	70.20	65.16
Llama-3.1-8B-Instruct	77.19	67.92	64.62
zephyr-7b-beta	75.39	68.28	64.07
gemma2-9b-it	73.15	79.75	61.61
DeepSeek-R1-Distill-Llama-8B	73.15	65.23	60.38
DeepSeek-R1-Distill-Qwen-7B	72.25	73.66	58.33
Llama-3.1-8B	68.88	60.75	55.87
Phi-4-reasoning	67.30	75.63	56.46
Mistral-7B-Instruct-v0.3	66.85	74.55	51.91
Llama-2-13b-chat-hf	65.96	52.33	51.78
Llama-3.2-1B-Instruct	64.38	41.58	51.09
Phi-4-mini-reasoning	63.93	59.4	50.41
gemma-7b	62.70	53.76	48.22
Qwen1.5-MoE-A2.7B	47.30	38.53	46.45
phi-2	58.20	37.63	43.72
gemma-2-2b-it	53.93	27.24	40.16
EleutherAI/gpt-j-6b	36.63	24.55	38.52
dolly-v1-6b	25.84	22.58	30.87

Table 2: Model accuracy on selected quantum topics. Accuracy above 90% are shaded green and those below 50% are shaded red.

Models performing well on QC500 and QC1000 maintained comparable performance levels on larger benchmarks, suggesting that a carefully selected sample of a few hundred questions provides sufficient evaluation of quantum computing knowledge. Among the evaluated models, Claude 4 Sonnet achieved the highest overall performance, closely followed by GPT-5, GPT-4o, and Claude Sonnet 3.7. Notably, among open-source models, Phi-4-reasoning-plus, IBM Granite-3.3-8b-instruct, and Llama-3.1-8B-Instruct demonstrated reasonable performance on quantum computing tasks despite their smaller parameter counts. While these models still trail behind the larger proprietary systems, their relative competence suggests they could serve as practical starting points for domain-specific fine-tuning where computational resources are limited.

Table 2 shows a clear pattern: models handle basic concepts but decline sharply on advanced material, with the largest drop on quantum algorithms and security. Security questions were especially difficult, including recent work on phase mismatch attacks, crosstalk exploitation, QubitHammer, and quantum backdoors. These gaps highlight the challenge of fast moving areas that demand specialized knowledge, and the examples that follow illustrate the kinds of questions where even top models failed.

- What specific attack technique can manipulate the error rates of specific quantum gates ?
- What specific vulnerability does a quantum reorder attack exploit?
- What makes dynamical decoupling ineffective against QubitHammer attacks ?

This performance gap between foundational and advanced topics is particularly revealing. The disparity suggests that models have absorbed well-documented principles from extensive training data but struggle with recent developments where literature is sparser and terminology less standardized. The consistency of this pattern across model families, regardless of size or provider, indicates that the challenge lies in the nature of the material rather than individual model limitations. Notably, even the expanded evaluation with agentic and deep research modes (Section 4.5) showed only marginal improvements on these advanced topics, confirming that web access alone cannot compensate for gaps in specialized reasoning. This finding has practical implications: users relying on LLMs for quantum computing assistance should exercise particular caution in rapidly evolving subfields where model knowledge may lag behind current research.

4.2 Human Performance Baseline Study

To establish a human baseline for comparison with language model performance, we conducted a survey study with quantum computing researchers and practitioners. We carefully selected 30 questions from QC-Bench spanning different topic areas and complexity levels to assess human expertise across the quantum computing domain. The survey included questions from all seven categories. Participants were recruited from academic institutions and quantum computing research groups. Each respondent provided background information including their highest education level, years of experience in quantum computing, and age group. Further details on each participant's background and individual score are provided in the appendix, offering context for the distribution shown here. The sample questions below illustrate the style and difficulty of the survey items used in this comparison.

Sample Survey Questions

- Why is Shor's algorithm considered a threat to modern cryptographic security?
- How does quantum transpilation optimize quantum circuits for real hardware?
- Which quantum algorithm is specifically designed to process structured graph data?

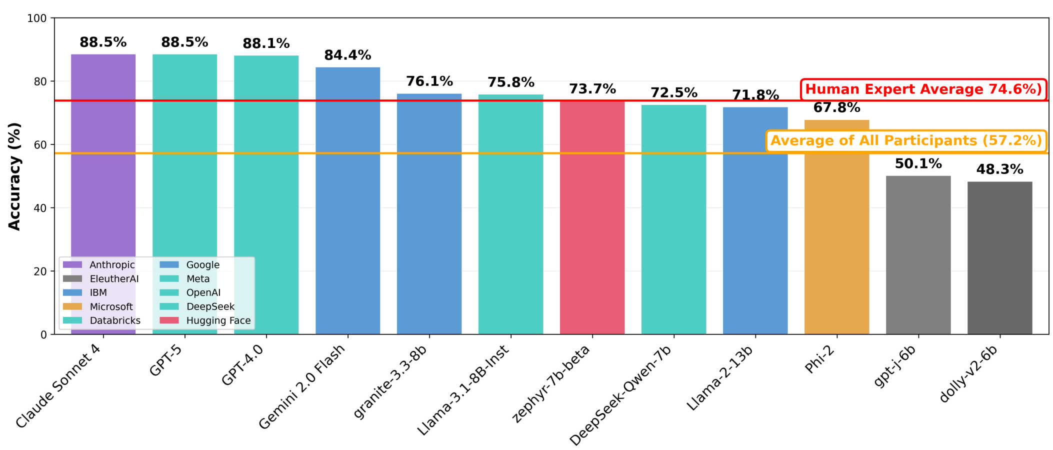

Figure 2: Performance comparison of selected LLMs across different capability tiers on the QC500 benchmark against human baselines. The visualization includes 12 representative models ranging from top performers to those scoring below novice human levels. Bars are colored by model provider.

Figure 2 presents a representative sample of LLM performance on the QC500 benchmark, showcasing models across the full performance spectrum. The majority of models shown exceed the all-participants average of 57.2%, while several surpass the expert average of 74.6%. The visualization highlights the dramatic performance range in quantum computing capabilities, from leading models like Claude 4 Sonnet (88.55%) and GPT-5 (88.07%) to models performing well below novice human levels, such as gpt-j-6b (50.14%) and dolly-v1-6b (48.29%). This selection demonstrates that quantum computing proficiency varies significantly across model families, sizes, and providers.

4.3 Performance Across Different Question Formats

Our evaluation extended beyond multiple-choice questions to assess model capabilities in diverse testing scenarios. For true/false questions, we modified the standard prompts to request binary verification of quantum computing statements. In open-ended questions we evaluated models' ability to generate explanations independently without options. Open-ended responses were graded based on factual correctness and conceptual completeness, with multiple-choice questions serving as the primary evaluation method while open-ended questions function as a supplementary diagnostic tool. Most models maintained strong performance on the true/false questions while showing clear degradation on open-ended assessments. Accuracy on the true/false set was tightly clustered, with smaller models often matching the large models once the task was reduced to a simple binary choice.

The limited number of options in the true/false evaluation leaves less room to distinguish stronger reasoning ability, so the gap between the very top systems and the weakest models nearly disappears in this format. By contrast, multiple-choice questions with four options revealed a more visible separation among high-end models, highlighting that true/false items are not an effective way to validate deeper research questions. Open-ended questions told a different story. GPT-5 not only produced the highest scoring answers when evaluated for correctness but also consistently provided richer, more contextually grounded explanations than its peers, and those detailed responses were closely aligned with the correct conclusions in most cases. This pattern underscores that open-ended evaluation exposes real differences in reasoning quality that are obscured when models face only binary decisions. Table 3 presents the complete results.

4.4 Fine-tuning Potential for Quantum Knowledge

We explored QC-Bench's utility for enhancing quantum computing capabilities through targeted fine-tuning. Using a subset of 4,400 questions for training and 1,000 questions as a test set, we fine-tuned five smaller language models using LoRA (Low-Rank Adaptation).

Our fine-tuning implementation used PyTorch with the Transformers library, applying LoRA with rank=8 and alpha=16 targeting attention projection matrices. We used a learning rate of 1e-4 with AdamW optimizer, batch size of 4 with gradient accumulation over 4 steps, and trained for a single epoch with warmup steps to ensure stable adaptation without overfitting.

Table 4 demonstrates the results across our selected models. Llama-3.1-8B-Instruct showed the strongest adaptation with a 5% improvement, while Gemma 2B achieved a modest 3.7% gain. Qwen1.5-MoE-A2.7B showed minimal improvement despite its Mixture-of-Experts architecture. Surprisingly, Phi-4-mini-reasoning experienced a slight performance decline, and EleutherAI/gpt-j-6b demonstrated a substantial 7% drop in accuracy. These mixed results highlight how model architecture significantly influences fine-tuning outcomes, with instruction-tuned models generally showing better adaptation to specialized quantum computing knowledge than their general-purpose counterparts.

Model	**T/F (%)**	**O-E (%)**
GPT-5	93.27	**89.07**
Claude Sonnet 4	**93.99**	88.84
GPT-4o	93.75	86.22
Gemini 2.0 Flash	92.31	84.09
GPT-4.1 mini	93.03	79.81
llama-3.3-70b-versatile	91.35	74.58
Claude Haiku 3.5	93.75	78.15

Table 3: Accuracy on other question formats

Model	**Size**	**Before**	**After**
Llama-3.1-8B-Instruc	8B	74.75	79.80 ↑
Gemma 2B	7B	62.29	65.70 ↑
Qwen1.5-MoE-A2.7B	14.3B	58.50	58.90
Phi-4-mini-reasoning	3.84B	74.00	73.60
EleutherAI/gpt-j-6b	6B	31.80	24.80 ↓

Table 4: Performance change after fine-tuning

Since effective fine-tuning remains feasible only for smaller models that consistently underperform relative to frontier systems, we shift focus to inference-time augmentation strategies. Retrieval-augmented generation and emerging agentic reasoning capabilities present promising alternatives, particularly for the largest models where parameter updates prove impractical. These approaches sidestep the weight-adaptation bottleneck by empowering models to dynamically query external sources and execute multi-step reasoning during inference, offering a more scalable pathway for specialized domain applications.

4.5 Performance with Agentic and Deep Research Modes

Given the limited improvements observed from fine-tuning and the practical constraints of adapting larger models, we explored an alternative approach through agentic reasoning capabilities and deep research modes. These capabilities enable models to perform multi-step reasoning, search external sources, and synthesize information across multiple queries without requiring model adaptation or training data. This paradigm augments models with tool use and extended reasoning processes rather than modifying model weights to encode domain-specific knowledge. We evaluated frontier models equipped with their respective advanced capabilities: Claude Sonnet 4.5 Research Mode, Claude Sonnet 4.5 Extended Thinking, GPT-5.1 Deep Research, GPT-5.1 Agent Mode, and Gemini 3 Deep Research. These modes allow models to break down complex questions, search for relevant information, and synthesize answers through multi-step reasoning processes. We tested these capabilities on the QC500 subset, which provides a balanced evaluation across all quantum computing topics in our benchmark.

Model	**Before (%)**	**After (%)**	**Improvement**
Claude Sonnet 4.5 (Research Mode)	91.80	92.20	+0.40
Claude Sonnet 4.5 (Extended Thinking)	91.80	92.60	+0.80
GPT-5.1 (Deep Research)	91.40	92.80	+1.40
GPT-5.1 (Agent Mode)	91.40	91.20	-0.20
Gemini 3 (Deep Research)	87.40	89.20	+1.80

Table 5: Performance of frontier models with advanced reasoning capabilities on QC500. Average improvement is 0.84 percentage points.

Results in Table 5 show the performance of these advanced reasoning modes. GPT-5.1 with Deep Research achieved the highest score at 92.80%, while both Claude Sonnet 4.5 variants performed above 92%. Gemini 3 Deep Research showed the largest improvement, gaining 1.8 percentage points to reach 89.20%. Notably, GPT-5.1 Agent Mode showed a slight decline of 0.2 percentage points compared to the base model. The average improvement across all models was 0.84 percentage points, demonstrating modest gains from these advanced capabilities.

These results indicate that while agentic and deep research modes provide measurable benefits, the improvements remain relatively modest on our benchmark. This suggests that the fundamental challenge in quantum computing knowledge assessment lies in the breadth and depth of knowledge encoded during pretraining rather than in reasoning capabilities alone. Questions in QC-Bench are designed to test factual knowledge and conceptual understanding rather than multi-step reasoning, which may explain why reasoning-augmented modes show limited advantage. However, advanced reasoning modes offer key advantages: (1) they require no training data or computational resources for model adaptation, (2) they can access current information beyond the model's knowledge cutoff, and (3) they can be applied to the largest and most capable models where fine-tuning is often impractical.

4.6 Multilingual Benchmark Performance

To investigate how quantum computing knowledge transfers across languages, we evaluated all models on Spanish and French translations of QC500. This experiment provides quantitative insights into linguistic generalization of specialized technical knowledge. Figure 3 shows Spanish versus French accuracy for selected models. While most models fall along a diagonal cluster indicating correlated cross-lingual performance, the distribution reveals systematic language-dependent performance gaps. Across our full benchmark set, models lose on average 11.2 percentage points in French and 15.2 percentage points in Spanish relative to English baselines. This asymmetry is particularly notable, as Spanish exhibits approximately 55 percent greater performance degradation than French. Remarkably, only 34.5% of models maintain scores above 75% in Spanish, compared to 44.8% in French and 69.0% in English. The most linguistically consistent models (Claude 4 Sonnet, GPT-5, and Gemini 2.0 Flash) show standard deviations below 0.6 across languages, while the least consistent (Phi-4-reasoning) exhibits a standard deviation of 31.2.

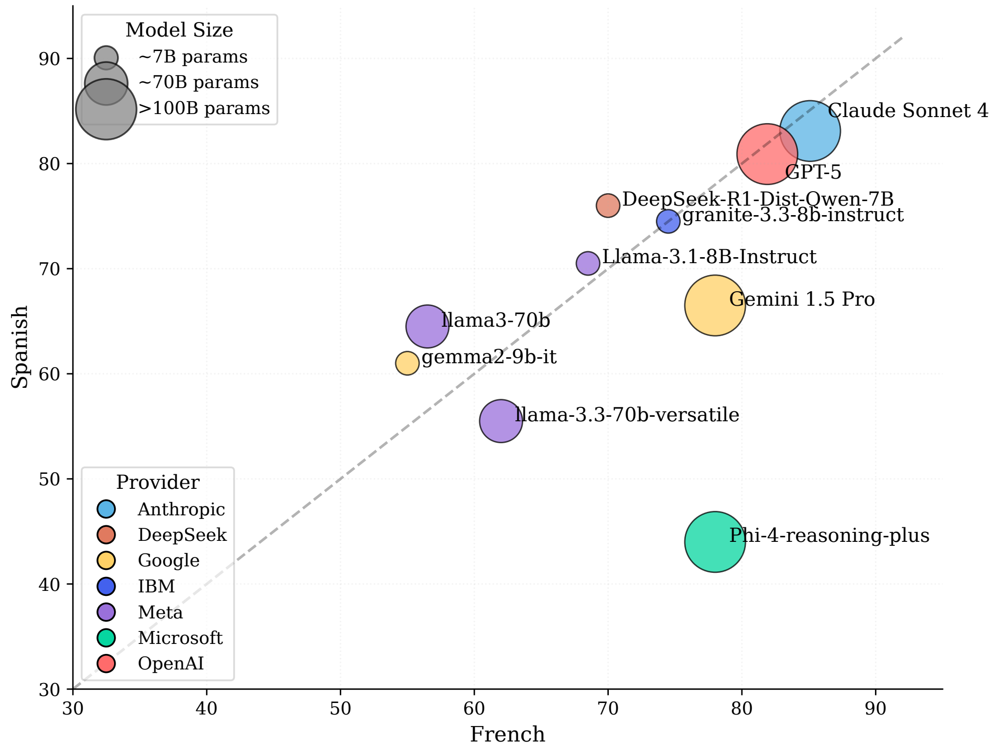

Figure 3: Bubble chart of Spanish (vertical) versus French (horizontal) accuracy on the QC500 benchmark. Each bubble's area is proportional to the model's parameter count; colors indicate providers. The diagonal dashed line marks equal performance across the two languages. Bubbles below the line signal larger accuracy loss in Spanish.

5 Discussion

Our evaluation reveals a clear performance pattern across all tested models: strong results on foundational topics with significant decline on advanced domains. Top models achieve over 92% accuracy on basic quantum concepts but drop below 77% for quantum security questions. This performance drop is particularly evident in questions about emerging attack vectors like phase mismatch attacks and QubitHammer, where even the most advanced models failed to provide accurate responses consistently. Notably, leading LLMs outperform many practitioners and experts in our human survey, where performance ranged from 26.6% to 86% depending on education level and experience (detailed results in the appendix). In addition, the results highlight a widening gap between recent state-of-the-art LLMs and smaller models, a trend that persists even after fine-tuning. These high-capacity systems show clear advantages not only on complex multiple-choice tasks but especially on open-ended questions, where they deliver more accurate and detailed explanations. Smaller models, by contrast, plateau despite fine-tuning, indicating that model scale and training pipelines remain critical for strong performance on demanding quantum computing assessments.

Question format comparison shows GPT-5 maintaining 89.07% accuracy on open-ended quantum explanations while most competitors show degradation without multiple-choice options. This suggests many models rely on recognizing answer patterns rather than constructing explanations from fundamental understanding. Our multilingual testing reveals concerning disparities, with average performance dropping 11.2 percentage points in French and 15.2 points in Spanish. Fine-tuning results demonstrate significant variation in how models adapt to quantum knowledge. Llama-3.1-8B-Instruct improved by 5.3% through fine-tuning, while EleutherAI/gpt-j-6b declined by 7%, suggesting that instruction-tuned models more readily incorporate specialized quantum knowledge.

As quantum computing advances toward practical implementation, retrieval-augmented generation can complement fine-tuning, particularly since practical fine-tuning is mainly feasible for smaller models. While targeted fine-tuning can modestly improve accuracy for compact systems, it remains costly and inflexible for the larger architectures that already set the performance frontier. Retrieval-augmented generation, by contrast, allows those high-capacity LLMs to access curated domain sources and continuously updated technical literature, avoiding the need for repeated full retraining.

6 Limitations and Future Work

QC-Bench offers a comprehensive evaluation of quantum computing knowledge, with English as the primary language and a large QC500 subset already available in Spanish and French. A next step is to expand coverage beyond QC500 by translating a larger portion of the benchmark into Spanish and French, and by adding more languages to better reflect global practice. Additional work includes increasing the diversity of non-English source materials and assessing cross-lingual consistency to provide a more complete view of multilingual performance.

Our evaluation relies primarily on accuracy as the central performance metric, which effectively captures models' factual knowledge but may not fully represent their conceptual understanding or reasoning capabilities. We chose accuracy for its interpretability, directness, and alignment with our goal of measuring factual correctness in quantum computing knowledge. Future research could explore alternative metrics such as calibration scores for confidence assessment, partial credit scoring for near-correct responses, or semantic similarity measures for evaluating open-ended explanations beyond binary correctness judgments.

Additionally, incorporating statistical frameworks such as error bars and confidence intervals would enhance the interpretability of results, as discussed by Miller (2024). Given the extensive nature of our evaluation across 31 models, multiple question formats, three languages, fine-tuning experiments, and agentic evaluation modes, incorporating this level of statistical rigor was beyond the current scope. We leave this as a direction for future work.

7 Conclusion

As Large Language Models (LLMs) are increasingly tasked with reading, explaining, and answering questions about quantum computing literature, rigorous domain evaluation is essential. QC-Bench provides a comprehensive assessment with 5,400 multiple-choice items plus 416 true/false and 421 open-ended questions across seven core domains. Across 31 systems, we find a consistent pattern: strong results on foundational material but marked drops on advanced topics. Top systems clear 92% on basic concepts yet fall below 77% on security questions, including items on recent developments in quantum security. Format matters: many models score well on multiple choice but degrade on open-ended responses; GPT-5 maintains the strongest open-ended performance among evaluated systems (89.07%) and produces more detailed, context-grounded explanations. Relative to human baselines (23.3%–86.7%), eight models exceed the all-participants average of 74.6% and surpass the expert average of 80.0%. Multilingual testing shows asymmetry, with average accuracy declines of 11.2 points in French and 16.2 in Spanish relative to English, indicating that quantum knowledge does not transfer uniformly across languages.

Methodologically, the results indicate a widening gap between state-of-the-art, high-capacity systems and smaller models, a difference that persists even after fine-tuning. Gains from fine-tuning are modest, typically only a few percentage points, and can sometimes reduce accuracy, making the computational cost difficult to justify for larger architectures. Evaluation of frontier models with agentic and deep research capabilities showed an average improvement of only 0.84 percentage points even with full internet access, suggesting that web search alone cannot compensate for gaps in specialized technical reasoning. Alternative approaches for enhancing performance on demanding technical domains remain an open research question. Quantum computing remains one of the most demanding areas for language models, and continued evaluation of LLM capabilities in this domain is essential for tracking progress and ensuring reliable performance as the field evolves.

REFERENCES

Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pp. 333–342, 2011.

Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.

Muhammad AbuGhanem. Ibm quantum computers: Evolution, performance, and future directions. *The Journal of Supercomputing*, 81(5):687, 2025.

Rajeev Acharya, Dmitry A Abanin, Laleh Aghababaie-Beni, Igor Aleiner, Trond I Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, et al. Quantum error correction below the surface code threshold. *Nature*, 2024.

Khalifa Afane, Wenqi Wei, Ying Mao, Junaid Farooq, and Juntao Chen. Next-generation phishing: How llm agents empower cyber attackers. In *2024 IEEE International Conference on Big Data (BigData)*, pp. 2558–2567. IEEE, 2024.

Mohamed Afane, Gabrielle Ebbrecht, Ying Wang, Juntao Chen, and Junaid Farooq. Atp: Adaptive threshold pruning for efficient data encoding in quantum neural networks. *arXiv preprint arXiv:2503.21815*, 2025.

Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 38(4):1207–1282, 2008.

Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 20–30, 1998.

Dorit Aharonov, Wim Van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Review*, 50(4): 755–787, 2008.

Victor V Albert, Kyungjoo Noh, Kasper Duivenvoorden, Dylan J Young, R T Brierley, Philip Reinhold, Christophe Vuillot, Linshu Li, Chao Shen, S M Girvin, et al. Performance and structure of single-mode bosonic codes. *Physical Review A*, 97(3):032346, 2018.

Panos Aliferis, Daniel Gottesman, and John Preskill. Quantum accuracy threshold for concatenated distance-3 codes. *Quantum Information & Computation*, 6(2):97–165, 2006.

HTS ALRikabi, Ibtisam A Aljazaery, Jaafar Sadiq Qateef, Abdul Hadi M Alaidi, and M Roa'a. Face patterns analysis and recognition system based on quantum neural network qnn. *iJIM*, 16(08):35, 2022.

Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1):210–239, 2007.

Mohammad H Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum boltzmann machine. *Physical Review X*, 8(2):021050, 2018.

Juan Miguel Arrazola, Alain Delgado, Bhaskar Roy Bardhan, and Seth Lloyd. Quantum-inspired algorithms in practice. *Quantum*, 4:307, 2020.

Juan Miguel Arrazola, Olivia Di Matteo, Nicolás Quesada, Soran Jahangiri, Alain Delgado, and Nathan Killoran. Universal quantum circuits for quantum chemistry. *Quantum*, 6:742, 2022.

Panagiotis Askaridis, Thomas Peijs, and Frank C Langbein. Quantum information processing in optical lattices and magnetic microtraps. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 54(10):104002, 2021.

David Avis, Charles Jordan, Jun Imoto, Yuki Sasaki, Sven Thomassen, Taiga Tsuda, and Seiya Yamanaka. Comparing small-and large-scale quantum computers using circuit simulation. *arXiv preprint arXiv:1904.11502*, 2019.

Koji Azuma, Kiyoshi Tamaki, and Hoi-Kwong Lo. Quantum repeaters for quantum key distribution: Progress and challenges. *NPJ Quantum Information*, 8(1):41, 2022.

Dave Bacon. Operator quantum error-correcting subsystems for self-correcting quantum memories. *Physical Review A*, 73(1):012340, 2006.

Adriano Barenco, Charles H Bennett, Richard Cleve, David P DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52(5):3457, 1995.

Rami Barends, Julian Kelly, Anthony Megrant, Andrzej Veitia, Daniel Sank, Evan Jeffrey, Ted C White, Josh Mutus, Austin G Fowler, Brooks Campbell, et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503, 2014.

Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. *Nature Communications*, 11(1):808, 2020.

Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

Marcello Benedetti, Mattia Fiorentini, and Michael Lubasch. Hardware-efficient variational quantum algorithms for time evolution. *Physical Review Research*, 3(3):033083, 2021.

Paul Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 22(5): 563–591, 1980.

Charles H Bennett. Quantum cryptography without Bell's theorem. *Physical Review Letters*, 68(21): 3121, 1992.

Charles H Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 560:7–11, 2014.

Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895, 1993.

Charles H Bennett, David P DiVincenzo, John A Smolin, and William K Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824, 1996.

Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

Daniel J Bernstein and Tanja Lange. Post-quantum cryptography. *Nature*, 549(7671):188–194, 2017.

Kishor Bharti, Alba Cervera-Lierta, Ting Hui Kyaw, Tobias Haug, Sierra Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022.

Mihir K Bhaskar, Ralf Riedinger, Bartholomeus Machielse, David S Levonian, Christian T Nguyen, Erik N Knall, Hongkun Park, Dirk Englund, Marko Loönchen, Denis D Sukachev, et al. Experimental demonstration of memory-enhanced quantum communication. *Nature*, 580(7801):60–64, 2020.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

Hugo Biard, Eufemio Moreno-Pineda, Mario Ruben, Edgar Bonet, Wolfgang Wernsdorfer, and Franck Balestro. Increasing the hilbert space dimension using a single coupled molecular spin. *Nature communications*, 12(1):4443, 2021.

Lukas Bischof, Stefan Teodoropol, Rudolf M Füchslin, and Kurt Stockinger. Hybrid quantum neural networks show strongly reduced need for free parameters in entity matching. *Scientific Reports*, 15 (1):4318, 2025.

Hillol Biswas. Data encoding for vqc in qiskit, a comparison with novel hybrid encoding. *arXiv preprint arXiv:2503.14062*, 2025.

Héctor Bombín and Miguel A Martin-Delgado. Topological quantum distillation. *Physical Review Letters*, 97(18):180501, 2006.

Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models. *arXiv preprint arXiv:2403.07059*, 2024.

Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum cryptanalysis of hash and claw-free functions. *ACM SIGACT News*, 28(2):14–19, 1997.

Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005.

Angela Sara Cacciapuoti, Marcello Caleffi, Rodney Van Meter, and Lajos Hanzo. Quantum internet: Networking challenges in distributed quantum computing. *IEEE Network*, 34(1):137–143, 2020.

A Robert Calderbank and Peter W Shor. Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098, 1996.

Marcello Caleffi and Angela Sara Cacciapuoti. Quantum switch for the quantum internet: Noiseless communications through noisy channels. *IEEE Journal on Selected Areas in Communications*, 38 (3):575–588, 2020.

Yudong Cao, Anmer Daskin, Steven Frankel, and Sabre Kais. Quantum circuit design for solving linear systems of equations. *Molecular Physics*, 110(15-16):1675–1680, 2012.

Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Harry Chi, Zohreh Davoudi, John Debes, Michael Geller, James Glick, Jacob Hauser, Jeremy Laroco, Norbert M Linke, Marco Pistoia, Takashi Shinozaki, et al. Programmable quantum processor with scalable connectivity and native multi-qubit gates. *Nature Communications*, 13(1):4449, 2022.

John Chiaverini, Dietrich Leibfried, Tobias Schaetz, MD Barrett, RB Blakestad, JW Britton, Wayne M Itano, John D Jost, Emanuel Knill, C Langer, et al. Realization of quantum error correction. *Nature*, 432(7017):602–605, 2004.

Andrew M Childs and Wim Van Dam. Quantum algorithms for algebraic problems. *Reviews of Modern Physics*, 82(1):1–52, 2010.

Andrew M Childs, Richard Cleve, Enrico Deotto, Edward Farhi, Sam Gutmann, and Daniel A Spielman. Exponential algorithmic speedup by a quantum walk. *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 59–68, 2003.

Koki Chinzei, Quoc Hoan Tran, Yasuhiro Endo, and Hirotaka Oshima. Resource-efficient equivariant quantum convolutional neural networks. *arXiv preprint arXiv:2410.01252*, 2024.

Frederic T Chong, Diana Franklin, and Margaret Martonosi. Programming quantum computers using design automation. *Design Automation Conference (DAC)*, pp. 1–6, 2017.

Cheng Chu, Fan Chen, Philip Richerme, and Lei Jiang. Qdoor: Exploiting approximate synthesis for backdoor attacks in quantum neural networks. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 1098–1106. IEEE, 2023.

Daniel JH Chung, Zhiqi Gao, Yurii Kvasiuk, Tianyi Li, Moritz Münchmeyer, Maja Rudolph, Frederic Sala, and Sai Chaitanya Tadepalli. Theoretical physics benchmark (tpbench)–a dataset and study of ai reasoning capabilities in theoretical physics. *arXiv preprint arXiv:2502.15815*, 2025.

J Ignacio Cirac and Peter Zoller. Quantum computations with cold trapped ions. *Physical Review Letters*, 74(20):4091, 1995.

J Ignacio Cirac, Peter Zoller, H Jeff Kimble, and Hideo Mabuchi. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Physical Review Letters*, 78 (16):3221, 1997.

Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998.

Iris Cong, Soonwon Choi, and Mikhail D Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.

David G Cory, Mark D Price, Wojciech Maas, Emanuel Knill, Raymond Laflamme, Wojciech H Zurek, Timothy F Havel, and S S Somaroo. Experimental quantum error correction. *Physical Review Letters*, 81(10):2152, 1998.

David G Cory, Raymond Laflamme, Emanuel Knill, Lorenza Viola, Timothy F Havel, Nicolas Boulant, Gregory Boutis, Evan Fortunato, Seth Lloyd, Rudy Martinez, et al. NMR based quantum information processing: Achievements and prospects. *Fortschritte der Physik: Progress of Physics*, 48(9-11):875–907, 2000.

Andrew W Cross, Lev S Bishop, Sarah Sheldon, Paul D Nation, and Jay M Gambetta. Validating quantum computers using randomized model circuits. *Physical Review A*, 100(3):032328, 2019.

Davide Cuomo, Marcello Caleffi, and Angela Sara Cacciapuoti. Towards a distributed quantum computing ecosystem. *IET Quantum Communication*, 1(1):3–8, 2020.

Giuseppe D'Adamo, Matteo Schiavon, and Thomas Haner. Distributed quantum computing and network control for accelerated vqe. *IEEE Transactions on Quantum Engineering*, 3:1–12, 2022.

Axel Dahlberg, Matthew Skrzypczyk, Tim Coopmans, Leon Wubben, Filip Rozpedek, Matteo Pompili, Arian Stolk, Przemyslaw Pawelczak, Rob Knegjens, Julio de Oliveira Filho, et al. A link layer protocol for quantum networks. *Proceedings of the ACM Special Interest Group on Data Communication*, pp. 159–173, 2019.

Flavio Del Santo and Nicolas Gisin. Which features of quantum physics are not fundamentally quantum but are due to indeterminism? *Quantum*, 9:1686, 2025.

Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. Topological quantum memory. *Journal of Mathematical Physics*, 43(9):4452–4505, 2002.

David Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 400 (1818):97–117, 1985.

David Deutsch and Artur Ekert. Quantum computation. *Physics World*, 11(3):47, 1998.

David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

Shalini Dhar, Ashish Khare, Ashutosh Dhar Dwivedi, and Rajani Singh. Securing iot devices: A novel approach using blockchain and quantum cryptography. *Internet of things*, 25:101019, 2024.

Eleni Diamanti, Hoi-Kwong Lo, Bing Qi, and Zhiliang Yuan. Practical challenges in quantum key distribution. *npj Quantum Information*, 2(1):1–12, 2016.

David P DiVincenzo. Quantum gates and circuits. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):261–276, 1998.

David P DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik: Progress of Physics*, 48(9-11):771–783, 2000.

David P DiVincenzo and Peter W Shor. Fault-tolerant quantum computation. *Physical Review Letters*, 77(15):3260, 1996.

Yuxuan Du, Tao Huang, Shan You, Min-Hsiu Hsieh, and Dacheng Tao. Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*, 8(1):62, 2022.

Vedran Dunjko and Hans J Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, 2018.

Artur K Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661, 1991.

Meng Fang, Xiangpeng Wan, Fei Lu, Fei Xing, and Kai Zou. Mathodyssey: Benchmarking mathematical problem-solving skills in large language models using odyssey math data. *arXiv preprint arXiv:2406.18321*, 2024.

Edward Farhi and Hartmut Neven. Classification with quantum neural networks on near term processors. *arXiv preprint arXiv:1802.06002*, 2018.

Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.

Richard P Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.

Frederik F Flöther. The state of quantum computing applications in health and medicine. *Research Directions: Quantum Technologies*, 1:e10, 2023.

Austin G Fowler, Matteo Mariantoni, John M Martinis, and Andrew N Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3):032324, 2012.

Jay M Gambetta, Jerry M Chow, and Matthias Steffen. Building logical qubits in a superconducting quantum computing system. *npj Quantum Information*, 3(1):1–7, 2017.

Bofei Gao, Feifan Song, Zhe Yang, Zefan Cai, Yibo Miao, Qingxiu Dong, Lei Li, Chenghao Ma, Liang Chen, Runxin Xu, et al. Omni-math: A universal olympiad level mathematic benchmark for large language models. *arXiv preprint arXiv:2410.07985*, 2024.

Vlad Gheorghiu and Michele Mosca. Quantum resource estimation for large scale quantum algorithms. *Future Generation Computer Systems*, 162:107480, 2025.

Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical Review Letters*, 100(16):160501, 2008.

Daniel Gottesman. Stabilizer codes and quantum error correction. *California Institute of Technology*, 1997.

Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vladimir Stojevic, Andrew G Green, and Simone Severini. Hierarchical quantum classifiers. *npj Quantum Information*, 4(1):65, 2018.

Michele Grossi, Noelle Ibrahim, Voica Radescu, Robert Loredo, Kirsten Voigt, Constantin Von Altrock, and Andreas Rudnik. Mixed quantum–classical method for fraud detection with quantum feature selection. *IEEE Transactions on Quantum Engineering*, 3:1–12, 2022.

Lov K Grover. A fast quantum mechanical algorithm for database search. *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pp. 212–219, 1996.

Stanley P Gudder. The hilbert space axiom in quantum mechanics. In *Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology: Essays in Honor of Wolfgang Yourgrau*, pp. 109–127. Springer, 1983.

Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, et al. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in Neural Information Processing Systems*, 36:44123–44279, 2023.

Xiaoyu Guo, Minggu Wang, and Jianjun Zhao. Quanbench: Benchmarking quantum code generation with large language models. *arXiv preprint arXiv:2510.16779*, 2025.

Jad C Halimeh, Monika Aidelsburger, Fabian Grusdt, Philipp Hauke, and Bing Yang. Cold-atom quantum simulators of gauge theories. *Nature Physics*, pp. 1–12, 2025.

Aram W Harrow and Debbie W Leung. Coherent communication of classical messages. *Physical Review Letters*, 92(9):097902, 2004.

Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009.

Akel Hashim, Andras Gyenis, Jie Zhang, Youngsun Nam, Tanay Mundada, Deanna M Lee, Melissa Niedzielski, Antonio Mezzacapo, Stefan Istvan, Matthew D Reed, et al. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *Physical Review X*, 11(4):041039, 2021.

Matthew B Hastings, Dave Wecker, Bela Bauer, and Matthias Troyer. Improving quantum algorithms for quantum chemistry. *arXiv preprint arXiv:1403.1539*, 2014.

Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

Matthew Hayward. Quantum computing and shor's algorithm. *Sydney: Macquarie University Mathematics Department*, 1, 2008.

Fei Hua, Meng Wang, Gushu Li, Bo Peng, Chenxu Liu, Muqing Zheng, Samuel Stein, Yufei Ding, Eddy Z Zhang, Travis Humble, et al. Qasmtrans: A qasm quantum transpiler framework for nisq devices. In *Proceedings of the SC'23 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis*, pp. 1468–1477, 2023.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Power of data in quantum machine learning. *Nature Communications*, 12(1):2631, 2021.

Nouhaila Innan, Abhishek Sawaika, Ashim Dhor, Siddhant Dutta, Sairupa Thota, Husayn Gokal, Nandan Patel, Muhammad Al-Zafar Khan, Ioannis Theodonis, and Mohamed Bennai. Financial fraud detection using quantum graph neural networks. *Quantum Machine Intelligence*, 6(1):7, 2024.

Satoshi Ishizaka and Tohya Hiroshima. Asymmetric quantum teleportation with complete bell-state measurement. *Physical Review Letters*, 101(24):240501, 2008.

Shi Jin, Nana Liu, and Yue Yu. Time complexity analysis of quantum algorithms via linear representations for nonlinear ordinary and partial differential equations. *Journal of Computational Physics*, 487:112149, 2023.

Stephen P Jordan. Fast quantum algorithm for numerical gradient estimation. *Physical Review Letters*, 95(5):050501, 2005.

Siddarth K Joshi, Björn Hessmo, Alex Ciurana, Javier A. Martín-Sánchez, Mohsen Razavi, and Remigiusz Augusiak. Trusted node quantum repeaters: A modular approach to the quantum internet. *New Journal of Physics*, 22(10):103005, 2020.

Siddarth K Joshi, Vittorio Giovannetti, Seth Lloyd, and Stefano Pirandola. Distributed quantum sensing networks: Challenges and opportunities. *Nature Communications*, 15(1):201, 2024.

V Kalaivani et al. Enhanced bb84 quantum cryptography protocol for secure communication in wireless body sensor networks for medical applications. *Personal and ubiquitous computing*, 27 (3):875, 2021.

Justin Kalloor, Mathias Weiden, Ed Younis, John Kubiatowicz, Bert De Jong, and Costin Iancu. Quantum hardware roofline: Evaluating the impact of gate expressivity on quantum processor design. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 805–816. IEEE, 2024.

Manuj Kant, Sareh Nabi, Manav Kant, Roland Scharrer, Megan Ma, and Marzieh Nabi. Towards robust legal reasoning: Harnessing logical llms in law. *arXiv preprint arXiv:2502.17638*, 2025.

Shlomo Kashani. Quantumllminstruct: A 500k llm instruction-tuning dataset with problem-solution pairs for quantum computing. *arXiv preprint arXiv:2412.20956*, 2024.

Julian Kelly, Rami Barends, Austin G Fowler, Anthony Megrant, Evan Jeffrey, Theodore C White, Daniel Sank, Josh Y Mutus, Brooks Campbell, Yu Chen, et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519(7541):66–69, 2015.

Sumeet Khatri and Mark M Wilde. Principles of quantum communication theory: A modern approach. *Communications of the ACM*, 64(1):76–85, 2021.

H Jeff Kimble. The quantum internet. *Nature*, 453(7198):1023–1030, 2008.

Andrew D King, Alberto Nocera, Marek M Rams, Jacek Dziarmaga, Roeland Wiersema, William Bernoudy, Jack Raymond, Nitin Kaushal, Niclas Heinsdorf, Richard Harris, et al. Beyond-classical computation in quantum simulation. *Science*, 388(6743):199–204, 2025.

Alexei Yu Kitaev. Quantum measurements and the abelian stabilizer problem. *Electronic Colloquium on Computational Complexity*, 3(3), 1995.

Alexei Yu Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191, 1997.

Morten Kjaergaard, Mollie E Schwartz, Johannes Braumüller, Philip Krantz, Joel I-Jan Wang, Simon Gustavsson, and William D Oliver. Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11:369–395, 2020.

Emanuel Knill, Raymond Laflamme, and Gerard J Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52, 2001.

Karol Kowalski and Nicholas P Bauman. Quantum flow algorithms for simulating many-body systems on quantum computers. *Physical Review Letters*, 131(20):200601, 2023.

Wojciech Kozlowski, Axel Dahlberg, and Stephanie Wehner. Designing a quantum network protocol. *Proceedings of the 16th International Conference on emerging Networking EXperiments and Technologies*, pp. 1–16, 2020.

Walter O Krawec, Bing Wang, and Ryan Brown. Finite key security of simplified trusted node networks. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 1777–1787. IEEE, 2024.

David A Kreplin and Marco Roth. Reduction of finite sampling noise in quantum neural networks. *Quantum*, 8:1385, 2024.

Hari Krovi. Improved quantum algorithms for linear and nonlinear differential equations. *Quantum*, 7:913, 2023.

Janusz Kusyk, Samah M Saeed, and Muharrem Umit Uyar. Survey on quantum circuit compilation for noisy intermediate-scale quantum computers: Artificial intelligence to heuristics. *IEEE Transactions on Quantum Engineering*, 2:1–16, 2021.

Raymond Laflamme, Emanuel Knill, Cécile Negrevergne, David G Cory, Timothy F Havel, Rudy Martinez, and Lorenza Viola. NMR quantum information processing and entanglement. *Progress of Physics*, 50:703–713, 2002.

Julien Laurat, Gaelle Keller, JoséAugusto Oliveira-Huguenin, Claude Fabre, Thomas Coudreau, Alessio Serafini, Gerardo Adesso, and Fabrizio Illuminati. Entanglement of two-mode gaussian states: characterization and experimental production and manipulation. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(12):S577, 2005.

Gabriela Barreto Lemos, Victoria Borish, Garrett D Cole, Sven Ramelow, Radek Lapkiewicz, and Anton Zeilinger. Quantum imaging with undetected photons. *Nature*, 512(7515):409–412, 2014.

Ang Li, Samuel Stein, Sriram Krishnamoorthy, and James Ang. Qasmbench: A low-level quantum benchmark suite for nisq evaluation and simulation. *ACM Transactions on Quantum Computing*, 4 (2):1–26, 2023.

Chun-Lin Li, He-Liang Chen, and Man-Hong Wang. Quantum reinforcement learning. *arXiv preprint arXiv:1810.00481*, 2018.

YaoChong Li, Ri-Gui Zhou, RuQing Xu, Jia Luo, and WenWen Hu. A quantum deep convolutional neural network for image recognition. *Quantum Science and Technology*, 5(4):044003, 2020.

Daniel A Lidar and Todd A Brun. Quantum error correction. *Cambridge University Press*, 27:1–10, 2013.

Norbert M Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A Landsman, Kenneth Wright, and Christopher Monroe. Experimental comparison of two quantum computing architectures. *Proceedings of the National Academy of Sciences*, 114(13):3305–3310, 2017.

Seth Lloyd. Hybrid quantum computing. *arXiv preprint quant-ph/9304003*, 1993.

Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Physical Review Letters*, 121(4):040502, 2018.

Hoi-Kwong Lo and Hoi Fung Chau. Unconditional security of quantum key distribution over arbitrarily long distances. *Science*, 283(5410):2050–2056, 1999.

Hoi-Kwong Lo, Marcos Curty, and Kiyoshi Tamaki. Secure quantum key distribution. *Nature Photonics*, 8(8):595–604, 2014.

Dmitri Maslov, Gerhard W Dueck, D Michael Miller, and Cecile Negrevergne. Quantum circuit simplification and level compaction. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 27(3):436–444, 2008.

Dominic Mayers. Unconditional security in quantum cryptography. *Journal of the ACM (JACM)*, 48 (3):351–406, 2001.

Matt McEwen, Lara Faoro, Kaveh Arya, Andrew Dunsworth, Trent Huang, Shirin Kim, Bruno Morvan, Simon Nigg, Naoki Kanazawa, Kevin Satzinger, et al. Removing leakage-induced correlated errors in superconducting quantum error correction. *Nature Communications*, 14(1): 3408, 2023.

David C McKay, Thomas Alexander, Luciano Bello, Michael J Biercuk, Lev Bishop, Jungsang Chen, Jerry M Chow, Antonio D Córcoles, Daniel Egger, Stefan Filipp, et al. Qiskit backend specifications for OpenQASM and OpenPulse experiments. *arXiv preprint arXiv:1809.03452*, 2018.

Miralem Mehic, Libor Michalek, Emir Dervisevic, Patrik Burdiak, Matej Plakalovic, Jan Rozhon, Nerman Mahovac, Filip Richter, Enio Kaljic, Filip Lauterbach, et al. Quantum cryptography in 5g networks: A comprehensive overview. *IEEE Communications Surveys & Tutorials*, 26(1):302–346, 2023.

Rodney Van Meter, Takahiko Satoh, Thaddeus D Ladd, William J Munro, and Kae Nemoto. Path selection for quantum repeater networks. *Networking Science*, 3(1):82–95, 2013.

Evan Miller. Adding error bars to evals: A statistical approach to language model evaluations. *arXiv preprint arXiv:2411.00640*, 2024.

Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.

Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3 (3):030503, 2018.

Ashley Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(1):1–8, 2016.

Michele Mosca. Quantum algorithms. *arXiv preprint arXiv:0808.0369*, 2008.

Mario Motta, Chong Sun, Adrian TK Tan, Matthew J O'Rourke, Erika Ye, Austin J Minnich, Fernando GSL Brandao, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16(2):205–210, 2020.

William J Munro, Kae Nemoto, Rebecca G Beausoleil, and Timothy P Spiller. High-efficiency quantum-nondemolition single-photon-number-resolving detector. *Physical Review A*, 71(3): 033819, 2005.

William J Munro, Koji Azuma, Kiyoshi Tamaki, and Kae Nemoto. Inside quantum repeaters. *IEEE Journal of Selected Topics in Quantum Electronics*, 21(3):78–90, 2015.

Prakash Murali, Jonathan M Baker, Ali Javadi Abhari, Frederic T Chong, and Margaret Martonosi. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 1015–1029, 2019.

Sreraman Muralidharan, Linshu Li, Jungsang Kim, Norbert Lütkenhaus, Mikhail D Lukin, and Liang Jiang. Optimal architectures for long distance quantum communication. *Scientific Reports*, 6(1): 20463, 2016.

Bhaskara Narottama and Soo Young Shin. Quantum neural networks for resource allocation in wireless communications. *IEEE transactions on wireless communications*, 21(2):1103–1116, 2021.

Bhaskara Narottama, Triwidyastuti Jamaluddin, and Soo Young Shin. Quantum neural network with parallel training for wireless resource optimization. *IEEE Transactions on Mobile Computing*, 23 (5):5835–5847, 2023.

Ashwin Nayak and Felix Wu. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 53(5):659–663, 1999.

Michael A Nielsen and Isaac L Chuang. Quantum computation by measurement and quantum memory. *Physical Review Letters*, 79(2):321, 1997.

Michael A Nielsen and Isaac L Chuang. Quantum computation and quantum information. *Physics Today*, 55(2):60, 2002.

Michael A Nielsen and Isaac L Chuang. Quantum computation and quantum information: 10th anniversary edition. *Cambridge University Press*, 2010.

Morten IP Nielsen and Karsten Flensberg. Fault-tolerant quantum computation with majorana fermions. *Physical Review Letters*, 126(7):076401, 2021.

Hadas Orgad, Michael Toker, Zorik Gekhman, Roi Reichart, Idan Szpektor, Hadas Kotek, and Yonatan Belinkov. Llms know more than they show: On the intrinsic representation of llm hallucinations. *arXiv preprint arXiv:2410.02707*, 2024.

Mateusz Ostaszewski, Lea M Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko. Reinforcement learning for optimization of variational quantum circuit architectures. *Advances in Neural Information Processing Systems*, 34:18182–18194, 2021.

Mihir Pant, Hari Krovi, Don Towsley, Leandros Tassiulas, Liang Jiang, Prithwish Basu, Dirk Englund, and Saikat Guha. Routing entanglement in the quantum internet. *npj Quantum Information*, 5(1): 25, 2019.

Tom Peham, Lukas Burgholzer, and Robert Wille. Equivalence checking of quantum circuits with the zx-calculus. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 12(3): 662–675, 2022.

Alejandro Perdomo-Ortiz, Marcello Benedetti, John Realpe-Gómez, and Rupak Biswas. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Science and Technology*, 3(3):030502, 2018.

Gabrijela Perković, Antun Drobnjak, and Ivica Botički. Hallucinations in llms: Understanding and addressing challenges. In *2024 47th MIPRO ICT and electronics convention (MIPRO)*, pp. 2084–2088. IEEE, 2024.

Sébastien Perseguers. Quantum networks: from a physics experiment to a quantum device. *Reports on Progress in Physics*, 76(9):096001, 2013.

Michael J Peterer, Samuel J Bader, Xiaoyue Jin, Fei Yan, Archana Kamal, Theodore J Gudmundsen, Peter J Leek, Terry P Orlando, William D Oliver, and Simon Gustavsson. Coherence and decay of higher energy levels of a superconducting transmon qubit. *Physical review letters*, 114(1):010501, 2015.

Frank Phillipson. Quantum computing in logistics and supply chain management an overview. *arXiv preprint arXiv:2402.17520*, 2024.

Stefano Pirandola, Bhaskar Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724–733, 2018.

Stefano Pirandola, Ulrik L Andersen, Leonardo Banchi, Mario Berta, Darius Bunandar, Roger Colbeck, Dirk Englund, Tobias Gehring, Cosmo Lupo, Carlo Ottaviani, et al. Advances in quantum cryptography. *Advances in Optics and Photonics*, 12(4):1012–1236, 2020.

Matteo Pompili, Sophie LE Hermans, Simon Baier, Hans KC Beukers, Peter C Humphreys, Raymond N Schouten, Raymond FL Vermeulen, Marijn J Tiggelman, Laura dos Santos Martins, Bas Dirkse, et al. Realization of a multinode quantum network of remote solid-state qubits. *Science*, 372(6539):259–264, 2021.

John Preskill. Reliable quantum computers. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):385–410, 1998.

John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.

Ricard Puig, Marc Drudis, Supanut Thanasilp, and Zoë Holmes. Variational quantum simulation: a case study for understanding warm starts. *PRX Quantum*, 6(1):010317, 2025.

Xiaogang Qiang, Yizhi Wang, Shichuan Xue, Renyou Ge, Lifeng Chen, Yingwen Liu, Anqi Huang, Xiang Fu, Ping Xu, Teng Yi, et al. Implementing graph-theoretic quantum algorithms on a silicon photonic quantum walk processor. *Science Advances*, 7(9):eabb8375, 2021.

Finley Alexander Quinton, Per Arne Sevle Myhr, Mostafa Barani, Pedro Crespo del Granado, and Hongyu Zhang. Quantum annealing applications, challenges and limitations for optimisation problems compared to classical solvers. *Scientific Reports*, 15(1):12733, 2025.

Vishakha K Ralegankar, Jagruti Bagul, Bhaumikkumar Thakkar, Rajesh Gupta, Sudeep Tanwar, Gulshan Sharma, and Innocent E Davidson. Quantum cryptography-as-a-service for secure uav communication: applications, challenges, and case study. *Ieee Access*, 10:1475–1492, 2021.

Deepak Ranga, Aryan Rana, Sunil Prajapat, Pankaj Kumar, Kranti Kumar, and Athanasios V Vasilakos. Quantum machine learning: Exploring the role of data encoding techniques, challenges, and future directions. *Mathematics*, 12(21):3318, 2024.

Minati Rath and Hema Date. Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy. *EPJ Quantum Technology*, 11 (1):72, 2024.

Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113(13):130503, 2014.

Matthew D Reed, Leonardo DiCarlo, Simon E Nigg, Luyan Sun, Luigi Frunzio, Steven M Girvin, and Robert J Schoelkopf. Realization of three-qubit quantum error correction with superconducting circuits. *Nature*, 482(7385):382–385, 2012.

Ben W Reichardt. Span programs and quantum query complexity: The general adversary bound is nearly tight for every boolean function. *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pp. 544–551, 2009.

Xiangyu Ren, Mengyu Zhang, and Antonio Barbalace. A hardware-aware gate cutting framework for practical quantum circuit knitting. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, pp. 1–9, 2024.

Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017.

Evandro CR Rosa, Eduardo I Duzzioni, and Rafael De Santiago. Optimizing gate decomposition for high-level quantum programming. *Quantum*, 9:1659, 2025.

Swastik Kumar Sahu and Kaushik Mazumdar. State-of-the-art analysis of quantum cryptography: applications and future prospects. *Frontiers in Physics*, 12:1456491, 2024.

Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Reviews of Modern Physics*, 81(3):1301, 2009.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.

Maria Schuld and Francesco Petruccione. *Supervised learning with quantum computers*. Springer, 2018.

Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. *An introduction to quantum machine learning*, volume 56. Taylor & Francis, 2015.

Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.

Priyabrata Senapati, Samuel Yen-Chi Chen, Bo Fang, Tushar M Athawale, Ang Li, Weiwen Jiang, Cheng Chang Lu, and Qiang Guan. Pqml: Enabling the predictive reproducibility on nisq machines for quantum ml applications. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 1413–1424. IEEE, 2024.

Peter W Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):R2493, 1995.

Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332, 1999.

Peter W Shor and John Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Physical Review Letters*, 85(2):441, 2000.

Scott Sikorski. Gpu-based quantum circuit simulation transpilation optimizations. *IEEE Transactions on Quantum Engineering*, 4:1–12, 2023.

Christoph Simon. Towards quantum repeater networks using multiplexed atomic memories. *Physical Review A*, 95(1):012308, 2017.

Jérémy Simon. Theoretical investigation of solid-state quantum memories in rare-earth doped oxides. *Quantum Science and Technology*, 1(1):012002, 2015.

Thomas R Smith, Mark R Drummond, and Michael D Reid. Quantum computational advantage using photons. *Physical Review A*, 99(5):052335, 2019.

Xiaoshuai Song, Muxi Diao, Guanting Dong, Zhengyang Wang, Yujia Fu, Runqi Qiao, Zhexu Wang, Dayuan Fu, Huangxuan Wu, Bin Liang, et al. Cs-bench: A comprehensive benchmark for large language models towards computer science mastery. *arXiv preprint arXiv:2406.08587*, 2024.

Andrew Steane. Error correcting codes in quantum theory. *Physical Review Letters*, 77(5):793, 1996.

Andrew M Steane. Efficient fault-tolerant quantum computing. *Nature*, 399(6732):124–126, 1999.

Krishna Subedi. The reliability of llms for medical diagnosis: An examination of consistency, manipulation, and contextual awareness. *arXiv preprint arXiv:2503.10647*, 2025.

Francesco Tacchino, Chiara Macchiavello, Dario Gerace, and Daniele Bajoni. An artificial neuron implemented on an actual quantum processor. *npj Quantum Information*, 5(1):26, 2019.

Shuntaro Takeda and Akira Furusawa. Wavelength-multiplexed quantum networks with ultrafast frequency combs. *Nature Photonics*, 17(1):6–15, 2023.

Maika Takita, Andrew W Cross, Antonio D Córcoles, Jerry M Chow, and Jay M Gambetta. Experimental demonstration of fault-tolerant state preparation with superconducting qubits. *Physical Review Letters*, 119(18):180501, 2017.

Yizhuo Tan, Navnil Choudhury, Kanad Basu, and Jakub Szefer. Qubithammer attacks: Qubit flipping attacks in multi-tenant superconducting quantum computers. *arXiv preprint arXiv:2504.07875*, 2025.

Hiroaki Terashima and Masahito Ueda. Nonunitary quantum circuit. *International Journal of Quantum Information*, 3(04):633–647, 2005.

Barbara M Terhal. Quantum error correction for quantum memories. *Reviews of Modern Physics*, 87 (2):307, 2015.

Norbert Tihanyi, Mohamed Amine Ferrag, Ridhi Jain, Tamas Bisztray, and Merouane Debbah. Cybermetric: a benchmark dataset based on retrieval-augmented generation for evaluating llms in cybersecurity knowledge. In *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 296–302. IEEE, 2024.

Raihan Ur Rasool, Hafiz Farooq Ahmad, Wajid Rafique, Adnan Qayyum, Junaid Qadir, and Zahid Anwar. Quantum computing for healthcare: A review. *Future Internet*, 15(3):94, 2023.

Rodney Van Meter. Path to quantum repeaters. *IEEE Security & Privacy*, 14(4):58–66, 2016.

Rodney Van Meter, Joe Touch, and Clare Horsman. Quantum networking and internetworking. *IEEE Network*, 23(4):6–13, 2009.

Rodney Van Meter, Marcello Caleffi, and Lajos Hanzo. Quantum internetworking. *IEEE Journal on Selected Areas in Communications*, 38(3):421–424, 2020.

Lieven MK Vandersypen, Matthias Steffen, Gregory Breyta, Costantino S Yannoni, Mark H Sherwood, and Isaac L Chuang. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature*, 414(6866):883–887, 2001.

Davide Venturelli, Minh Do, Eleanor Rieffel, and Jeremy Frank. Compiling quantum circuits to realistic hardware architectures using temporal planners. *Quantum Science and Technology*, 3(2): 025004, 2018.

Sanjay Vishwakarma, Francis Harkins, Siddharth Golecha, Vishal Sharathchandra Bajpe, Nicolas Dupuis, Luca Buratti, David Kremer, Ismael Faro, Ruchir Puri, and Juan Cruz-Benito. Qiskit humaneval: An evaluation benchmark for quantum code generative models. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 1169–1176. IEEE, 2024.

Haoran Wang, Pingzhi Li, Min Chen, Jinglei Cheng, Junyu Liu, and Tianlong Chen. Grovergpt: A large language model with 8 billion parameters for quantum searching. *arXiv preprint arXiv:2501.00135*, 2024.

Stephanie Wehner, David Elkouss, and Ronald Hanson. Quantum internet: A vision for the road ahead. *Science*, 362(6412):eaam9288, 2018.

Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Physical Review Letters*, 109(5):050505, 2012.

Robert Wille, Lukas Burgholzer, and Alwin Zulehner. Mapping quantum circuits to IBM QX architectures using the minimal number of SWAP and H operations. *Design Automation Conference (DAC)*, pp. 1–6, 2019.

Colin P Williams and Alexander G Gray. Automated design of quantum circuits. In *NASA International Conference on Quantum Computing and Quantum Communications*, pp. 113–125. Springer, 1998.

Peter Wittek. *Quantum machine learning: what quantum computing means to data mining*. Academic Press, 2014.

William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886): 802–803, 1982.

Chuanqi Xu and Jakub Szefer. Security attacks abusing pulse-level quantum circuits. *arXiv preprint arXiv:2406.05941*, 2024.

Chuanqi Xu, Jessie Chen, Allen Mi, and Jakub Szefer. Securing nisq quantum computer reset operations against higher energy state attacks. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pp. 594–607, 2023.

Xiaoqing Xu, Quntao Li, Xilin Su, Xiaoqian Guo, Minghua Liu, and Myung-Joong Kim. Secure quantum machine learning over the cloud. *Physical Review A*, 101(5):052343, 2020.

Rui Yang, Yuntian Gu, Ziruo Wang, Yitao Liang, and Tongyang Li. Qcircuitnet: A large-scale hierarchical dataset for quantum algorithm design. *arXiv preprint arXiv:2410.07961*, 2024.

Ed Younis and Costin Iancu. Quantum circuit optimization and transpilation via parameterized circuit instantiation. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 465–475. IEEE, 2022.

Kai Yu, Song Lin, and Bin-Bin Cai. Quantum convolutional neural network with flexible stride. *arXiv preprint arXiv:2412.00645*, 2024.

Chi Zhang, Xiao-Long Hu, Cong Jiang, Jiu-Peng Chen, Yang Liu, Weijun Zhang, Zong-Wen Yu, Hao Li, Lixing You, Zhen Wang, et al. Experimental side-channel-secure quantum key distribution. *Physical Review Letters*, 128(19):190503, 2022a.

Kaining Zhang, Liu Liu, Min-Hsiu Hsieh, and Dacheng Tao. Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits. *Advances in Neural Information Processing Systems*, 35:18612–18627, 2022b.

Yusheng Zhao, Hui Zhong, Xinyue Zhang, Yuqing Li, Chi Zhang, and Miao Pan. Bridging quantum computing and differential privacy: Insights into quantum computing privacy. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 1, pp. 13–24. IEEE, 2024.

Alwin Zulehner, Alexandru Paler, and Robert Wille. An efficient methodology for mapping quantum circuits to the IBM QX architectures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 38(7):1226–1236, 2018.

Wojciech Hubert Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75(3):715, 2003.

A Appendix

A.1 Question Extraction and Generation

For the QC500 and QC1000 subsets, researchers systematically reviewed the selected quantum computing papers and identified key concepts, algorithms, and principles that capture both foundational and advanced material. Rather than copying sentences directly, each question was crafted by rephrasing important findings and definitions from the literature to create original items while preserving scientific accuracy. Draft questions underwent multiple rounds of verification for technical correctness and clarity to ensure they tested understanding rather than memorization of phrasing. In parallel, we leveraged language models in a controlled setting to suggest candidate questions from the same source papers, but every suggestion was filtered and rewritten by researchers to maintain consistency with the human-validated style and difficulty. This combined process produced a balanced set of high-quality questions that reflect authentic quantum computing research while supporting rigorous evaluation across diverse topics and difficulty levels. The question creation process involved developing multiple answer options for each extracted concept. Below we illustrate this process with an example:

Example: Question Development Process

Initial concept from paper: Quantum circuit synthesis involves decomposing unitary operations into implementable gate sequences.
Generated question: What is the primary purpose of quantum circuit synthesis?
Initial answer options (6 generated):

- A. To convert a quantum circuit into a classical circuit by removing superposition properties
- B. To merge multiple unitary matrices into a single high-dimensional operator without gate decomposition
- C. To decompose a unitary matrix representing the circuit into a sequence of gates from the native gate set
- D. To encode classical information into qubit states without performing any gate-level modifications
- E. To simulate quantum circuits on classical computers using tensor networks
- F. To optimize quantum algorithms for specific hardware architectures

Final selection (4 options): Options E and F were eliminated as they describe related but distinct processes. The final question includes the correct answer (C) and three plausible distractors that test understanding of quantum circuit concepts.

Quantum Circuit Synthesis

Question: What is the primary purpose of quantum circuit synthesis?

- A. To convert a quantum circuit into a classical circuit by removing superposition properties
- B. To merge multiple unitary matrices into a single high-dimensional operator without gate decomposition
- C. To decompose a unitary matrix representing the circuit into a sequence of gates from the native gate set
- D. To encode classical information into qubit states without performing any gate-level modifications

Answer: C

A.2 Automated Question Mining from Research Papers

To expand our benchmark beyond human-authored questions, we employed Large Language Models (LLMs), specifically Gemini 2.0 Flash, Gemini 1.5 Pro, GPT-4.0, and Claude 3.7 Sonnet, to extract additional questions from 212 carefully selected quantum computing papers. This automated extraction process generated an initial pool of over 8,000 candidate questions, each with six potential answer options.

The filtering process involved multiple stages. First, we identified and removed questions that strayed from quantum computing into adjacent domains. Examples of filtered questions include:

Example: Filtered Question - General Cybersecurity

Question: Which encryption protocol is most commonly used for securing HTTP connections?

A. TLS/SSL
B. SSH
C. IPSec
D. WPA2

Reason for filtering: While encryption is relevant to quantum cryptography, this question addresses classical network security without quantum computing connection.

Example: Filtered Question – Mathematical Modeling

Question: In the analysis of an ordinary differential equation system, what does a non-positive log-norm of the coefficient matrix imply?

A. The system is unstable for all inputs
B. The matrix has only imaginary eigenvalues
C. The solution decays or remains bounded over time
D. The matrix is diagonalizable over the complex field

Reason for filtering: While this concept appears in resource analyses for quantum-inspired algorithms, it tests classical stability theory in differential equations and does not assess quantum computing knowledge.

After removing duplicate questions, filtering irrelevant content, and conducting manual quality review, we retained 4,400 high-quality questions. For each retained question, we selected the four most relevant answer options from the initial six, ensuring each question had one correct answer and three well-crafted distractors that effectively test quantum computing knowledge.

A.3 Question Translation and Multilingual Validation

For the QC500 subset, we created Spanish and French translations using a multi-stage process. We employed the same four LLMs (Gemini 2.0 Flash, Gemini 1.5 Pro, GPT-4.0, and Claude 3.7 Sonnet) to generate initial translations. A typical translation prompt was structured as follows:

```
"Translate the following quantum computing question from English to French,
maintaining technical accuracy and appropriate scientific terminology:
[Question and answer options]"
```

For each question, we collected translations from all four models and selected the most accurate version. This selection was then reviewed by individuals proficient in both languages who verified technical terminology and ensured conceptual accuracy. The translation process preserved the semantic content while adapting to language-specific conventions for scientific terminology.

Examples of translated questions include:

French Translation Example 1

Question: Pourquoi les attaques par impulsion à grande échelle sont-elles difficiles à réaliser dans les systèmes partagés ?

A. Elles dépendent d'un accès chiffré aux qubits
B. Elles nécessitent un accès à la machine au niveau administrateur
C. Elles requièrent de nombreux qubits, auxquels les utilisateurs n'ont généralement pas accès
D. Elles échouent si la machine est calibrée

Answer: C

French Translation Example 2

Question: Pourquoi les algorithmes quantiques paramétriques sont-ils difficiles à vérifier sémantiquement ?

A. Ils utilisent des paramètres fixes définis dans le matériel
B. Ils reposent uniquement sur un post-traitement classique
C. Leurs paramètres entraînés manquent d'interprétabilité inhérente
D. Leur structure est identique pour tous les ensembles de données

Answer: C

Spanish Translation Example 1

Question: ¿Cuál es la principal diferencia entre la privacidad diferencial clásica y la privacidad diferencial cuántica?

A. La PD cuántica extiende las garantías de privacidad a estados cuánticos indistinguibles utilizando distancias de traza
B. La PD cuántica elimina la necesidad de análisis probabilístico
C. La PD cuántica se aplica solo a registros de qubits entrelazados
D. La PD cuántica se impone eliminando los resultados de medición de qubits

Answer: A

Spanish Translation Example 2

Question: ¿Qué algoritmo clásico se utiliza comúnmente después del paso cuántico del Algoritmo de Shor?

A. Algoritmo de Dijkstra
B. Expansión de fracciones continuas
C. Integración de Monte Carlo
D. Búsqueda binaria

Answer: B

A.4 Question Format Diversification

To evaluate models' performance across different cognitive tasks, we expanded our benchmark with true/false and open-ended questions. For this expansion, we selected 40 additional research papers to ensure diverse content and avoid repetition. Both LLMs and human experts generated questions following similar protocols to the initial question creation phase.

This process yielded 416 true/false questions and 421 open-ended questions. True/false questions were created by converting factual statements into binary assessments:

True/False Question Examples

Question: The Bloch sphere is a geometrical representation of pure quantum states of a two-level quantum mechanical system.
Answer: True

Question: Dirac notation can only represent pure states, not mixed states.
Answer: False

Question: The E91 protocol is based on entangled particles and provides a method for secure quantum key distribution.
Answer: True

Question: Quantum error correction codes do not require any additional qubits beyond the physical qubits used to represent the logical qubit.
Answer: False

Open-ended questions were designed to assess deeper understanding and explanatory capabilities:

Open-Ended Question Examples

Question: What is the no-cloning theorem and its implication for quantum information?

Question: How does the Heisenberg uncertainty principle affect the measurement of quantum states?

Question: In the context of quantum states, what distinguishes a pure state from a mixed state?

Question: Explain the significance of the CNOT gate in quantum entanglement.
Sample Answer: The CNOT gate, or controlled-NOT gate, is crucial for creating entanglement between two qubits, as it flips the state of the target qubit only if the control qubit is in the state $|1\rangle$.

For each open-ended question, we developed sample answers to facilitate consistent evaluation across different models. These questions were assessed manually to determine whether model responses captured the essential concepts and technical accuracy required for each topic.

A.5 Human Performance Baseline Study

Table 6 reports accuracies for the first 20 respondents. Scores range from 23.3% to 86.7%, with an overall average of 57.2%. Participants with 5+ years of experience achieved an average of 79.4%, providing a reference point for expert-level performance. Education level shows a clear pattern: all PhD-trained participants scored at or above 73.3%, while no BS-level participant reached 60%. These results provide a concrete reference distribution for interpreting model-human comparisons in the main results.

References for Each Topic

Table 7 lists the literature sources and citation coverage for all seven benchmark topics. Rapidly developing areas such as quantum cybersecurity and quantum machine learning rely heavily on the most recent papers to capture ongoing advances, whereas foundational categories such as quantum theory and quantum error correction draw on a broader historical record to reflect the principles that remain central to the discipline. This distribution ensures that the benchmark balances up-to-date research with enduring theoretical foundations, giving a clear view of how source material supports each topic area.

Participant	Education	Experience	Age Group	Score	Accuracy
P1	MS	2–5 yrs	25–35	19/30	63.3%
P2	BS	<1 yr	18–25	14/30	46.7%
P3	PhD	5+ yrs	35–45	25/30	83.3%
P4	MS	1–2 yrs	25–35	21/30	70.0%
P5	PhD	2–5 yrs	35–45	23/30	76.7%
P6	BS	<1 yr	18–25	12/30	40.0%
P7	MS	1–2 yrs	25–35	17/30	56.7%
P8	MS	5+ yrs	35–45	24/30	80.0%
P9	PhD	2–5 yrs	25–35	22/30	73.3%
P10	BS	1–2 yrs	18–25	16/30	53.3%
P11	PhD	5+ yrs	45–55	26/30	86.7%
P12	BS	<1 yr	18–25	11/30	36.7%
P13	MS	2–5 yrs	25–35	20/30	66.7%
P14	PhD	1–2 yrs	25–35	22/30	73.3%
P15	BS	<1 yr	18–25	8/30	26.7%
P16	PhD	5+ yrs	35–45	25/30	83.3%
P17	MS	2–5 yrs	25–35	23/30	76.7%
P18	PhD	5+ yrs	45–55	24/30	80.0%
P19	BS	<1 yr	18–25	7/30	23.3%
P20	MS	1–2 yrs	25–35	18/30	60.0%
Expert Average (5+ years experience):					**79.4%**
All Participants Average:					**57.2%**

Table 6: Human participant survey results on 30-question quantum computing assessment

A.6 Fine-tuning Methodology

Our fine-tuning experiments employed Low-Rank Adaptation (LoRA) to efficiently adapt smaller language models to quantum computing knowledge while maintaining computational feasibility. The implementation utilized a carefully selected subset of 4,167 question-answer pairs from the QC-Bench dataset for training, with an additional 1,000 questions reserved for evaluation. The training data was formatted as concatenated prompt-completion pairs to maximize learning efficiency within context length constraints. We applied LoRA with rank 8 and alpha 16, specifically targeting the attention projection matrices (q_proj, k_proj, v_proj, o_proj) which are critical for knowledge representation. The training configuration employed a batch size of 4 with gradient accumulation over 4 steps, resulting in an effective batch size of 16, paired with a conservative learning rate of 1e-4 using the AdamW optimizer. To ensure stable convergence, we implemented 50 warmup steps followed by training for a single epoch, which empirical testing showed was sufficient to achieve knowledge transfer without overfitting. The models were loaded in FP16 precision to reduce memory requirements while maintaining numerical stability, with automatic device mapping to optimize GPU utilization. Early stopping was monitored through validation accuracy computed every 200 steps, though most models converged within the single epoch. This approach resulted in training only approximately 0.5-2% of total model parameters, demonstrating that quantum computing knowledge can be effectively incorporated through targeted parameter updates rather than full model retraining.

Topic	References	Years	#
Basic Concepts	Aharonov et al. (1998); Terashima & Ueda (2005); Arrazola et al. (2022); Williams & Gray (1998); Hayward (2008); Xu et al. (2023); Peterer et al. (2015); Schuld & Killoran (2019); Gudder (1983); Biard et al. (2021); Kowalski & Bauman (2023); Del Santo & Gisin (2025); Younis & Iancu (2022); Hua et al. (2023); Sikorski (2023); Nielsen & Chuang (2010); Preskill (2018); Deutsch (1985); Feynman (1982); Kjaergaard et al. (2020); Bharti et al. (2022); Zurek (2003); Deutsch & Ekert (1998); Bennett et al. (1993); Quinton et al. (2025); Phillipson (2024); Cirac & Zoller (1995); Benioff (1980); Giovannetti et al. (2008); AbuGhanem (2025); Farhi et al. (2000); DiVincenzo (2000); Lloyd (1996); Knill et al. (2001); King et al. (2025); Halimeh et al. (2025); Puig et al. (2025); Munro et al. (2005); Harrow & Leung (2004); Steane (1996)	1980–2025	39
Gates & Circuit Design	Zhang et al. (2022b); Ren et al. (2024); Peham et al. (2022); Kusyk et al. (2021); Ostaszewski et al. (2021); Rosa et al. (2025); DiVincenzo (1998); Kalloor et al. (2024); Senapati et al. (2024); Cao et al. (2012); Venturelli et al. (2018); Barenco et al. (1995); Vandersypen et al. (2001); Steane (1999); Laflamme et al. (2002); Cory et al. (2000); Cross et al. (2019); Linke et al. (2017); Smith et al. (2019); Maslov et al. (2008); McKay et al. (2018); Chong et al. (2017); Hashim et al. (2021); Zulehner et al. (2018); Wille et al. (2019); Murali et al. (2019)	1995–2025	26
Quantum Machine Learning	Wittek (2014); Bowles et al. (2024); Vishwakarma et al. (2024); Ranga et al. (2024); Rath & Date (2024); Biswas (2025); Bischof et al. (2025); Kreplin & Roth (2024); Chinzei et al. (2024); Afane et al. (2025); Yu et al. (2024); Schuld et al. (2015); Havlíček et al. (2019); Cerezo et al. (2021); Biamonte et al. (2017); Schuld & Petruccione (2018); Farhi & Neven (2018); Dunjko & Briegel (2018); Benedetti et al. (2019); Lloyd & Weedbrook (2018); Beer et al. (2020); Huang et al. (2021); Mitarai et al. (2018); Rebentrost et al. (2014); Grant et al. (2018); Cong et al. (2019); Schuld et al. (2020); Amin et al. (2018); Perdomo-Ortiz et al. (2018); Moll et al. (2018); Arrazola et al. (2020); Romero et al. (2017); Tacchino et al. (2019); Li et al. (2018); Abbas et al. (2021)	2014–2025	35
Distributed Computing	Cuomo et al. (2020); Cacciapuoti et al. (2020); Wehner et al. (2018); Kimble (2008); Simon (2017); D'Adamo et al. (2022); Dahlberg et al. (2019); Caleffi & Cacciapuoti (2020); Pompili et al. (2021); Simon (2017); Van Meter (2016); Munro et al. (2015); Meter et al. (2013); Van Meter et al. (2009); Lloyd (1993); Cirac et al. (1997); Perseguers (2013); Pant et al. (2019); Ishizaka & Hiroshima (2008); Simon (2015); Laurat et al. (2005); Avis et al. (2019); Van Meter et al. (2020); Joshi et al. (2020); Lemos et al. (2014); Pirandola et al. (2018); Azuma et al. (2022); Takeda & Furusawa (2023); Joshi et al. (2024); Khatri & Wilde (2021); Bhaskar et al. (2020); Askaridis et al. (2021); Chi et al. (2022); Kozlowski et al. (2020); Muralidharan et al. (2016)	1993–2024	35
Quantum Security	Dhar et al. (2024); Mehic et al. (2023); Chu et al. (2023); Zhao et al. (2024); Xu et al. (2023); Krawec et al. (2024); Zhang et al. (2022a); Xu & Szefer (2024); Tan et al. (2025); Sahu & Mazumdar (2024); Ralegankar et al. (2021); Kalaivani et al. (2021); Pirandola et al. (2020); Bernstein & Lange (2017); Lo et al. (2014); Bennett & Brassard (2014); Xu et al. (2020); Ekert (1991); Bennett (1992); Scarani et al. (2009); Lo & Chau (1999); Shor & Preskill (2000); Mayers (2001); Renner (2008); Wootters & Zurek (1982); Diamanti et al. (2016)	1982–2025	26
Error Correction	Fowler et al. (2012); Shor (1995); Lidar & Brun (2013); Terhal (2015); Aharonov & Ben-Or (2008); Chiaverini et al. (2004); Reed et al. (2012); Bombín & Martin-Delgado (2006); Gottesman (1997); Nielsen & Flensberg (2021); Steane (1996); Kitaev (1997); Preskill (1998); Bacon (2006); Aliferis et al. (2006); Calderbank & Shor (1996); Steane (1999); Dennis et al. (2002); Acharya et al. (2024); Barends et al. (2014); Kelly et al. (2015); Cory et al. (1998); DiVincenzo & Shor (1996); Bravyi & Kitaev (2005); Albert et al. (2018); Bennett et al. (1996); Gambetta et al. (2017); McEwen et al. (2023); Takita et al. (2017)	1995–2024	29
Quantum Algorithms	Montanaro (2016); Mosca (2008); Childs & Van Dam (2010); Hastings et al. (2014); Gheorghiu & Mosca (2025); Krovi (2023); Jin et al. (2023); Qiang et al. (2021); Benedetti et al. (2021); Du et al. (2022); Motta et al. (2020); Grover (1996); Shor (1999); Harrow et al. (2009); Ambainis (2007); Kitaev (1995); Nayak & Wu (1999); Childs et al. (2003); Cleve et al. (1998); Farhi et al. (2000); Nielsen & Chuang (1997); Aharonov et al. (2008); Bennett et al. (1997); Deutsch & Jozsa (1992); Nielsen & Chuang (2002); Brassard et al. (1997); Jordan (2005); Reichardt (2009); Wiebe et al. (2012); Aaronson & Arkhipov (2011)	1995–2025	30

Table 7: Topic Coverage and Source Papers