Meta-Radiology

Quantum Artificial Intelligence: A Comprehensive Survey

--Manuscript Draft--

Manuscript Number:	METRAD-D-25-00117R1
Full Title:	Quantum Artificial Intelligence: A Comprehensive Survey
Short Title:	Quantum AI Review
Article Type:	Review
Keywords:	Quantum artificial intelligence; Hybrid quantum--classical computing; Variational quantum algorithms; Quantum transformers; NISQ devices
Corresponding Author:	Tianming Liu University of Georgia UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Georgia
Corresponding Author's Secondary Institution:	
First Author:	Hanqi Jiang
First Author Secondary Information:	
Order of Authors:	Hanqi Jiang
	Yi Pan
	Junhao Chen
	Zhengliang Liu
	Lichao Sun
	Quanzheng Li
	Lu Zhang
	Dajiang Zhu
	Xianqiao Wang
	Wei Liu
	Xiang Li
	Gang Li
	Wei Zhang
	Lin Zhao
	Xiaowei Yu
	Yingfeng Wang
	Tianming Liu
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED STATES OF AMERICA
Abstract:	Quantum Artificial Intelligence (QAI) has emerged at the nexus of quantum computing and AI, promising to redefine computational frontiers. This survey critically synthesizes the state-of-the-art through 2024, elucidating the profound bidirectional synergy between these fields. We analyze how classical machine learning is accelerating quantum hardware control, circuit optimization, and error correction. Conversely, we

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	assess the potential quantum advantage of algorithms, including variational and kernel-based methods, across domains such as drug discovery, financial modeling, and cybersecurity. Our analysis reveals a critical trade-off between the utility of near-term Noisy Intermediate-Scale Quantum (NISQ) devices and the long-term promise of fault-tolerant architectures. We identify fundamental obstacles to QAI's advancement, including hardware decoherence, algorithmic barren plateaus, and data-encoding bottlenecks. While QAI's potential is transformative, achieving practical quantum advantage requires a concerted effort to overcome these core challenges at the hardware-software interface. This work provides a roadmap for navigating the current landscape and prioritizing future research in this rapidly evolving discipline.
Additional Information:	
Question	**Response**

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Dear Editor and Reviewers,

We sincerely thank you for your constructive and encouraging comments. We have carefully revised the manuscript to strengthen its rigor, interpretability, and relevance for an imaging AI readership while preserving a balanced and non-hyped perspective on Quantum AI (QAI). Major revisions include: (i) more explicit and consistently placed caveats near performance summaries and any “advantage” statements; (ii) a new paragraph positioning the manuscript relative to recent surveys; (iii) correction and verification of author affiliations and addresses; (iv) improved table traceability by adding per-row citations and clarifying the context/assumptions behind comparisons; and (v) a dedicated section consolidating key challenges and future research directions.

Updates to the front matter and references are not marked due to constraints of the Elsevier CAS LaTeX template. Therefore, tracked changes are shown for the manuscript body only. These include adding an author (Zhengliang Liu), correcting affiliation mappings, and updating a reference to a 2024 systematic literature review. All scientific content changes are fully reflected in the marked-up manuscript.

Here is an important note on table numbering. Because we removed the former Table 1 (see Reviewer 2, Comment 2), subsequent tables have been renumbered in the revised manuscript. In this response letter, references to tables follow the revised manuscript numbering.

We respond to each reviewer comment point-by-point below. All changes are incorporated in the revised manuscript.

Responses to Reviewer 1

Comment 1 (Caveats near “advantage” statements and performance summaries)

Reviewer comment: The manuscript already discusses end-to-end costs and concentration/barren-plateau issues; please highlight these caveats wherever “advantage” appears in headlines/tables for healthcare readers.

Response: We agree and have revised the manuscript to ensure that any performance-oriented summary is accompanied by clear caveats about (i) NISQ noise and decoherence constraints, (ii) end-to-end costs (including data encoding/state preparation, sampling/shot overhead, and readout), and (iii) optimization/trainability limitations such as barren plateaus. In particular, we revised the surrounding narrative and table notes to avoid over-interpretation of isolated results and to explicitly separate NISQ practicality from fault-tolerant assumptions where relevant. We also refined wording to emphasize conditionality, for example “potential advantage under specified assumptions”, rather than implying universal or scalable advantage.

Comment 2 (What is new relative to recent surveys)

Reviewer comment: The manuscript reads as a comprehensive 2024–2025 update; an explicit "What's new vs. recent surveys" paragraph would help editors and readers.

Response: We agree and added a dedicated positioning paragraph/section that explicitly states how this survey differs from and extends recent QAI surveys. This new content summarizes our distinct contributions, including (i) a bidirectional "AI for Quantum" and "Quantum for AI" roadmap framed for an applied audience, (ii) explicit NISQ-aware cost/limitations placed next to performance summaries, and (iii) a translation-minded discussion that is aligned with radiology/medical imaging workflows and reporting expectations.

Additional minor polishing

Reviewer comment: With tighter radiology anchoring and minor polishing, the paper will be strong for an imaging AI audience.

Response: We have strengthened the radiology anchor without narrowing the scope of the survey. We clarified medical-imaging-relevant workflow considerations and reporting expectations (e.g., reproducibility, end-to-end cost accounting, and evaluation transparency) while keeping the main survey domain-general. We also performed additional editorial cleanup for consistency and readability (terminology, abbreviations, and phrasing).

Responses to Reviewer 2

Comment 1 (Institution/address issues)

Reviewer comment: Several authors' institutions and addresses do not seem correct. Please check them.

Response: We have corrected and verified the affiliations and postal addresses for all authors. Institution names and addresses have been updated for consistency and accuracy across the title/affiliation page.

Comment 2 (Lacking clarity on platforms/tasks for table comparisons; missing per-method references in Tables)

Reviewer comment: Tables 1 and 2 present time/accuracy metrics without specifying platforms/tasks; add this context. Also add references for each method listed in Tables 2, 3, and 5.

Response: Thank you—this is an important point. We addressed it in two ways:

1. **Removal of the former Table 1 to avoid misleading cross-paper comparisons.**

   The previous Table 1 attempted to summarize time and accuracy-type metrics across methods. However, we found that the underlying studies are conducted under heterogeneous tasks, circuit families, hardware backends/simulators, and measurement protocols. Because these differences make a unified, fair, and interpretable comparison infeasible, we removed the former Table 1 in the revised manuscript to prevent readers from drawing unwarranted conclusions from non-controlled comparisons. The revised manuscript instead emphasizes how such metrics should be reported and compared, and why benchmark/task/platform specificity matters. As a result, subsequent tables were renumbered: original Tables 2–5 are now Tables 1–4, respectively.

2. **Strengthened traceability and interpretability for the remaining tables.**

   We added per-row citations and improved contextual clarity in the remaining tables so that readers can directly trace each entry to its source and interpret the setting/assumptions. Where "time" appears, we clarified the meaning, for example classical compilation time vs. decoding latency, and avoided implying that these values represent quantum execution time. We also improved the description of the corresponding tasks and evaluation settings in the associated text and/or table notes. As noted above, table numbering has been updated accordingly after removing the former Table 1.

These changes collectively improve scholarly rigor, traceability, and interpretability while avoiding potentially misleading "apples-to-oranges" comparisons.

Comment 3 (Challenges and future direction)

Reviewer comment: Please list key challenges (decoherence, barren plateaus, data-encoding bottlenecks) in a dedicated section/table, and outline future research directions.

Response: We agree and added a dedicated "Key Challenges and Future Directions" section that explicitly consolidates the major bottlenecks facing QAI, including hardware noise/decoherence, trainability issues, encoding and end-to-end I/O constraints, sampling overhead, compilation/routing effects, and benchmarking/reporting inconsistency. We also outline concrete near-term directions and longer-term directions. This dedicated section addresses the requested clarity and improves the manuscript's utility as a roadmap.

Closing

We appreciate the reviewers' thoughtful feedback and believe the revisions have significantly strengthened the manuscript's rigor, clarity, and value to a radiology/imaging AI audience while maintaining a balanced view of QAI. We hope the revised manuscript is now suitable for publication in *Meta-Radiology*.

Sincerely,

Tianming Liu (on behalf of all authors)

University of Georgia

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author *Click here to enter your name* is Choose an item for *Click here to enter the journal's name* and was not involved in the editorial review or the decision to publish this article.

☐The authors declare the following financial interests (e.g., any funding for the research project)/personal relationships (e.g., the author is an employee of a profitable company) which may be considered as potential competing interests:

Click here to enter your full declaration

Highlights

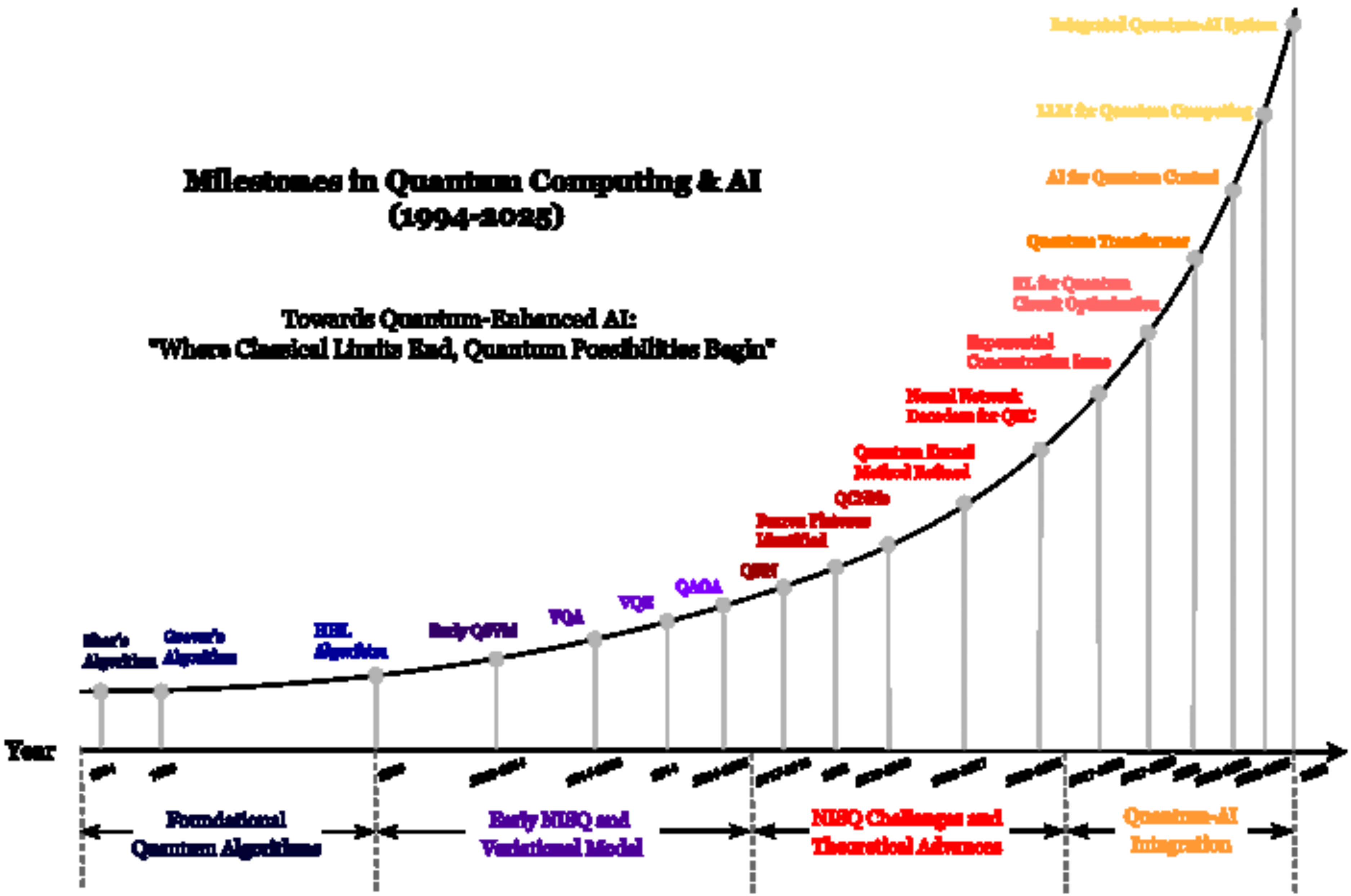
- We introduce a bi-directional Quantum-AI framework, AI-for-QC and QC-for-AI, and provide a practical roadmap for integrating both paradigms across the full stack.

- We quantify AI-enabled gains: calibration time drops from 8 hours to 2 hours, circuit depth falls by up to 42%, and ML decoders cut logical error rates by about half.

- We synthesize comparative guidance on VQAs, quantum kernels, and hybrid/quantum Transformers, clarifying trainability limits and criteria for credible quantum advantage.

- We deliver a clinically grounded playbook for radiology, covering small-data use cases, plug-in hybrid modules, OOD validation, fairness assessments, and TRIPOD/CONSORT-AI reporting.

Graphical Abstract (for review)
Click here to access/download;Graphical Abstract (for review);graphic abstract.png



Revised Manuscript with Changes Marked
Click here to view linked References

Quantum Artificial Intelligence: A Comprehensive Survey

Hanqi Jiang[a,1], Yi Pan[a,1], Junhao Chen[a,1], Zhengliang Liu[a], Lichao Sun[b], Quanzheng Li[c], Lu Zhang[d], Dajiang Zhu[e], Xianqiao Wang[a], Wei Liu[f], Xiang Li[c], Gang Li[h], Wei Zhang[i], Lin Zhao[j], Xiaowei Yu[k], Yingfeng Wang[l] and Tianming Liu[a,*]

[a]*School of Computing, University of Georgia, 415 Boyd Graduate Studies Research Center, Athens, 30602, Georgia, United States*
[b]*Department of Computer Science and Engineering, Lehigh University, 19 Memorial Drive West, Bethlehem, 18015, Pennsylvania, United States*
[c]*Department of Radiology, Harvard University, 55 Fruit Street, Boston, 02114, Massachusetts, United States*
[d]*Department of Computer Science, Indiana University Indianapolis, 535 W. Michigan Street, Indianapolis, 46202, Indiana, United States*
[e]*Department of Computer Science and Engineering, University of Texas at Arlington, 500 UTA Boulevard, Arlington, 76019, Texas, United States*
[f]*Department of Radiation Oncology, Mayo Clinic, 5881 East Mayo Boulevard, Phoenix, 85054, Arizona, United States*
[g]*Department of Radiology, Massachusetts General Hospital and Harvard Kempner Institute, 55 Fruit Street, Boston, 02114, Massachusetts, United States*
[h]*Department of Radiology, University of North Carolina at Chapel Hill, 130 Mason Farm Road, Chapel Hill, 27599, North Carolina, United States*
[i]*School of Computer and Cyber Sciences, Augusta University, 1120 15th Street, Augusta, 30912, Georgia, United States*
[j]*Department of Biomedical Engineering, New Jersey Institute of Technology, University Heights, Newark, 07102, New Jersey, United States*
[k]*Department of Computer Science, Missouri University of Science and Technology, 500 West 15th Street, Rolla, 65409, Missouri, United States*
[l]*Department of Computer Science and Engineering, University of Tennessee at Chattanooga, 615 McCallie Ave, Chattanooga, 37403, Tennessee, United States*

ARTICLE INFO

Keywords:
Quantum artificial intelligence
Hybrid quantum–classical computing
Variational quantum algorithms
Quantum transformers
NISQ devices

Abstract

Quantum Artificial Intelligence (QAI) has emerged at the nexus of quantum computing and AI, promising to redefine computational frontiers. This survey critically synthesizes the state-of-the-art through 2024, elucidating the profound bidirectional synergy between these fields. We analyze how classical machine learning is accelerating quantum hardware control, circuit optimization, and error correction. Conversely, we assess the potential quantum advantage of algorithms, including variational and kernel-based methods, across domains such as drug discovery, financial modeling, and cybersecurity. Our analysis reveals a critical trade-off between the utility of near-term Noisy Intermediate-Scale Quantum (NISQ) devices and the long-term promise of fault-tolerant architectures. We identify fundamental obstacles to QAI's advancement, including hardware decoherence, algorithmic barren plateaus, and data-encoding bottlenecks. While QAI's potential is transformative, achieving practical quantum advantage requires a concerted effort to overcome these core challenges at the hardware-software interface. This work provides a roadmap for navigating the current landscape and prioritizing future research in this rapidly evolving discipline.

1. Introduction

1.1. The Convergence of Quantum Computing and Artificial Intelligence

The confluence of quantum computing (QC) and artificial intelligence (AI), particularly machine learning (ML) and large language models (LLMs), marks the emergence of a transformative field often termed Quantum Artificial Intelligence (QAI) [1]. This interdisciplinary domain investigates the synergistic potential arising from integrating quantum mechanical principles with sophisticated AI algorithms. The core premise of QAI rests on a bidirectional relationship: quantum systems' unique computational capabilities can address intractable problems within AI, while AI's power can overcome significant hurdles in the development and operation of quantum computers.

Quantum computing harnesses phenomena such as superposition and entanglement to offer fundamentally new paradigms for information processing, promising capabilities that surpass classical machines for specific computational tasks [2]. Concurrently, AI has demonstrated remarkable success in extracting patterns from vast datasets, automating complex decision-making processes, and optimizing intricate systems, particularly through recent advances in ML and the development of powerful LLMs [3].

*Corresponding author.
ORCID(s):
[1]These authors have contributed equally to this work.

The field of QAI explores how these complementary strengths can be synergistically combined through two primary research directions. The first, “Quantum for AI,” investigates quantum algorithms that could potentially accelerate or enhance ML tasks, optimization problems, and data analysis. The second, “AI for Quantum,” applies AI techniques to address challenges in quantum hardware design, control, error correction, and algorithm discovery ~~[?]~~[4]. This dual approach signifies QAI’s evolution from a purely theoretical concept into a field with tangible, albeit nascent, potential for transformative technological impact.

Two concurrent trends make this an auspicious moment for QAI. On the quantum side, prototype devices with tens to low hundreds of qubits are increasingly stable and programmable, enabling experimental studies of hybrid quantum–classical workflows in the so-called NISQ (noisy intermediate-scale quantum) regime [2]. On the AI side, LLMs and modern ML pipelines have scaled to billions of parameters and industrial deployments, but face escalating computational and energy costs [3]. QAI is motivated by the possibility that carefully chosen quantum subroutines can provide algorithmic or constant-factor advantages for certain ML primitives, while advanced AI methods can in turn accelerate the path toward useful quantum computation by improving calibration, control, and fault management ~~[? 1]~~ [4, 1].

1.2. Motivation and Potential Impact

The motivation for exploring QAI stems significantly from the recognized limitations of classical computation in the face of escalating demands from modern science and technology. The relentless growth in data volume and the complexity of contemporary AI models, particularly deep learning architectures and LLMs, strain the capabilities of current computing infrastructures. Moore’s Law, the historical driver of exponential growth in classical hardware performance, is encountering physical limits, making it increasingly difficult and expensive to achieve further significant gains in computational power through traditional means. Training state-of-the-art LLMs, for instance, requires massive investments in time and hardware resources, accessible only to a few entities. Quantum computing emerges as a potential pathway to transcend these limitations. By exploiting quantum mechanical effects, QC promises to efficiently solve certain classes of problems currently considered intractable for even the most powerful classical supercomputers. This includes complex optimization problems frequently encountered in logistics and finance, simulations of quantum systems crucial for materials science and drug discovery, and specific linear algebra tasks underlying some ML algorithms.

Conversely, the very nature of quantum computing presents formidable challenges arising from its operation based on counterintuitive quantum phenomena, its susceptibility to noise, and the complexity of controlling and scaling quantum hardware. The high-dimensional Hilbert spaces, the intricacies of quantum dynamics, and the need for precise control make designing, calibrating, and operating quantum computers exceptionally difficult. Here, AI, with its proficiency in pattern recognition, data-driven learning, and optimization in high-dimensional spaces, offers essential tools. AI algorithms can analyze experimental data to characterize quantum systems, optimize control pulses for quantum gates, design efficient quantum circuits, develop error correction strategies, and potentially even discover novel quantum algorithms. This interplay suggests that progress in quantum computing may itself depend significantly on advancements in AI.

The potential impact of successful QAI development spans numerous domains. Accelerated drug discovery and materials science through enhanced quantum simulations [5], improved financial modeling and optimization, breakthroughs in optimization problems relevant to logistics and supply chains, enhanced capabilities in cybersecurity through quantum-resistant cryptography and potentially quantum-enhanced attack analysis [6], and fundamentally new approaches to scientific discovery through the analysis of complex datasets [7] represent just some of the anticipated applications.

1.3. Survey Structure

This survey provides a comprehensive overview of the QAI landscape, synthesizing current research and highlighting key developments, challenges, and future trajectories. Section 2 establishes the theoretical groundwork, outlining the fundamental concepts of both quantum computing and relevant AI paradigms, along with the essential mathematical formalism. Section 3 delves into the application of AI and LLMs to advance quantum computing technologies, covering areas from hardware design and calibration to circuit optimization and error correction (AI for QC). Section 4 explores the converse direction: the potential for quantum computing to enhance AI and LLM capabilities, focusing on Quantum Machine Learning (QML) algorithms and their prospective advantages (QC for AI). Section 4.2 provides a comparative analysis of the various methodologies discussed in Sections 3 and 4, evaluating their strengths, weaknesses, and performance characteristics.

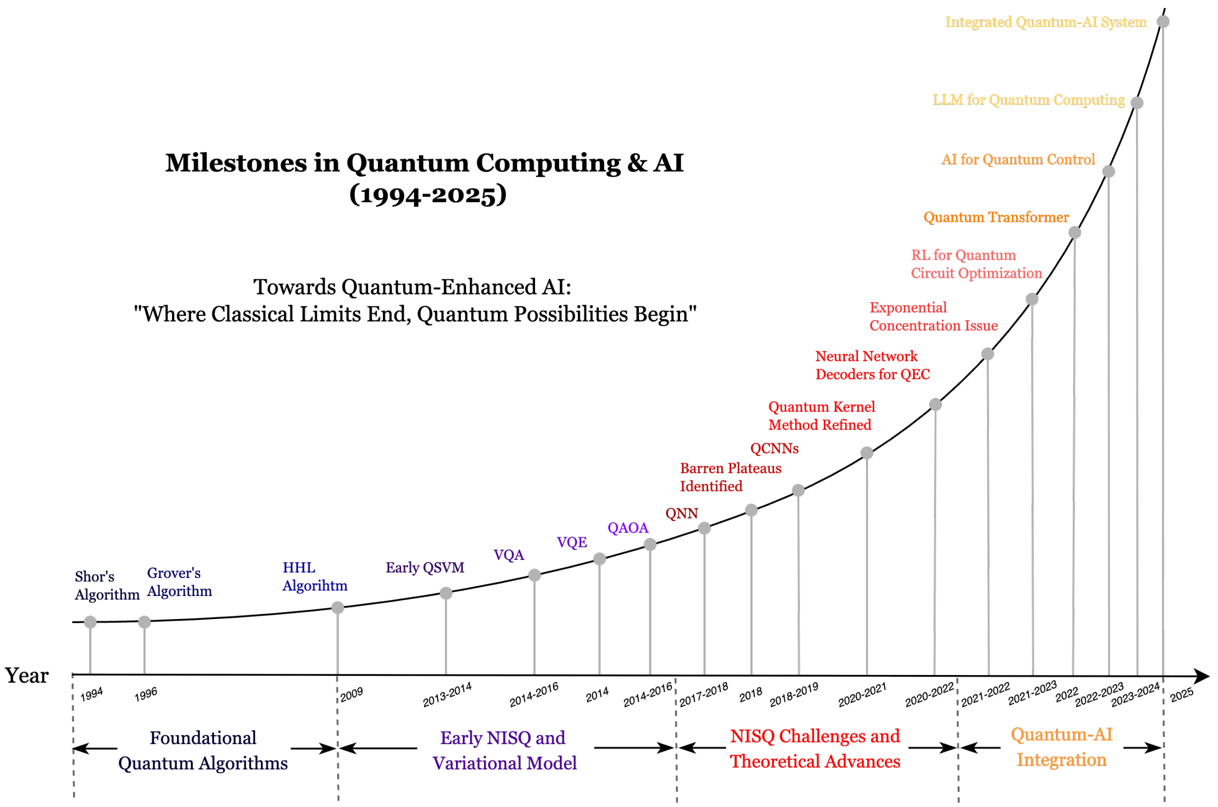


Figure 1: Key algorithmic milestones at the intersection of AI and quantum computing.

1.4. Contributions and positioning relative to recent surveys

Recent surveys on QAI and QML provide complementary but often directional or scope-limited perspectives [1, 8] . On the AI-for-quantum side, high-profile reviews organize how AI supports the quantum stack, including hardware design, calibration and control, compilation, and error mitigation, while explicitly not covering the reciprocal quantum-for-AI direction [4]. On the quantum-for-AI side, recent quantum machine learning surveys comprehensively cover broad methodological overviews ranging from NISQ to fault-tolerant regimes [9], examine near-term hardware-executed supervised and unsupervised demonstrations together with their practical bottlenecks [8], and present systematic or taxonomy-driven perspectives that classify algorithms, software frameworks, and datasets [7, 10].

Our survey is positioned to complement and extend these lines of work in three concrete ways. First, we provide a bidirectional, end-to-end roadmap that treats AI for quantum and quantum for AI under a single set of definitions and evaluation principles. We connect algorithmic performance summaries to NISQ-relevant caveats, including noise and error mitigation overheads, data-encoding costs, trainability and barren-plateau phenomena, and end-to-end wall-clock and resource accounting, to avoid over-interpreting small-scale comparisons. Second, beyond consolidating established paradigms such as variational quantum algorithms, quantum kernels, and hybrid models, we incorporate a 2024–2025 update spanning the hardware landscape, compilation and optimization, quantum error correction and decoding, and emerging model families, including transformer-style architectures and domain-specific pipelines, while emphasizing which claims rely on fault-tolerant assumptions versus what is currently testable on NISQ hardware. Third, distinct from general-purpose QAI and QML surveys and from domain-specific healthcare summaries, we introduce a radiology-anchored translation lens that maps QAI components to imaging workflows and provides deployment-minded guidance on data governance, reporting, subgroup performance, calibration, and regulatory and ethical considerations that is actionable for an imaging AI audience.

In short, while recent surveys provide substantial depth within individual facets of QAI and QML, this article is intended as a practical, compliance-aware roadmap that integrates both directions and makes NISQ realism explicit whenever claims of advantage or utility are discussed.

2. Foundational Concepts and Theoretical Frameworks

A robust understanding of the QAI intersection requires familiarity with the core principles of both quantum computing and artificial intelligence, as well as the mathematical language that connects them.

2.1. Core Principles of Quantum Computing

Quantum computing operates on principles fundamentally different from classical computation, leveraging the counterintuitive phenomena of quantum mechanics to process information.

The fundamental unit of quantum information is the quantum bit, or qubit ~~[?]~~[1]. Unlike a classical bit, which can only represent either 0 or 1, a qubit can exist in a state that is a complex linear combination of these two basis states, denoted $|0\rangle$ and $|1\rangle$. This state, $|\psi\rangle$, is represented as a vector in a two-dimensional complex Hilbert space:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle, \tag{1}$$

where α and β are complex numbers known as probability amplitudes [2]. The squares of the magnitudes of these amplitudes, $|\alpha|^2$ and $|\beta|^2$, represent the probabilities of measuring the qubit in the $|0\rangle$ or $|1\rangle$ state, respectively, and must sum to one ($|\alpha|^2 + |\beta|^2 = 1$). A system of n qubits resides in a 2^n-dimensional Hilbert space, allowing for an exponential growth in the state space capacity compared to classical bits.

The ability of a qubit to exist in a combination of $|0\rangle$ and $|1\rangle$ simultaneously is known as superposition [11]. This principle allows quantum computers to explore multiple computational paths concurrently, forming the basis for quantum parallelism, a key potential source of quantum speedup [12].

Entanglement is a uniquely quantum phenomenon where two or more qubits become inextricably linked, sharing a single quantum state. Measuring the state of one entangled qubit instantaneously influences the state of the others, regardless of the physical distance separating them. This non-local correlation allows for complex information processing and is a crucial resource for many quantum algorithms and communication protocols [13].

Quantum computations are performed by applying sequences of quantum gates to qubits ~~[?]~~[4]. These gates are mathematically represented by unitary matrices, ensuring that the evolution of the quantum state preserves probability (i.e., the state remains normalized). A quantum circuit is a sequence of these gates, analogous to a classical logic circuit, designed to implement a specific algorithm. Universal quantum computation can be achieved using a finite set of gates, such as single-qubit rotations (like $R_x(\theta)$, $R_y(\theta)$, $R_z(\theta)$) and a two-qubit entangling gate like the Controlled-NOT (CNOT) gate [14].

Extracting information from a quantum computer involves measurement [15]. Measurement is a probabilistic process governed by the Born rule: the probability of obtaining a specific outcome (e.g., measuring a qubit as 0) is given by the squared magnitude of the corresponding amplitude ($|\alpha|^2$ for outcome 0). Crucially, the act of measurement irrevocably collapses the quantum state's superposition into the single classical state corresponding to the measurement outcome. This contrasts sharply with classical systems where observation can often be non-intrusive.

Quantum algorithms harness interference, where the probability amplitudes of different computational paths can interfere constructively (reinforcing desired outcomes) or destructively (canceling undesired outcomes) [6]. This wave-like behavior of amplitudes, including the possibility of negative or complex values, is a key differentiator from classical probabilistic algorithms and enables quantum algorithms to find solutions more efficiently for certain problems [16].

Two primary models of quantum computing exist. Gate-based quantum computing uses quantum circuits composed of discrete gates, analogous to classical digital computers, and is considered capable of universal quantum computation [13]. Adiabatic Quantum Computing (AQC) involves preparing a system in the easily achievable ground state of a simple Hamiltonian and slowly evolving the Hamiltonian to one whose ground state encodes the solution to the problem. Quantum Annealing (QA) is a related, more practical approach focused specifically on finding approximate solutions to optimization problems, often formulated as Quadratic Unconstrained Binary Optimization (QUBO) problems, by guiding a system towards a low-energy state [9]. QA devices are typically specialized for optimization rather than universal computation.

2.2. Quantum Hardware Landscape for QAI

Progress in quantum hardware relevant to QAI is best understood by architecture: gate-model superconducting qubits, trapped ions, neutral atoms, photonics, semiconductor spin qubits, and quantum annealers. Their physical characteristics—qubit count, connectivity, coherence, native gate set and fidelity, cycle times/sampling throughput, and calibration automation—bound the feasible regime for variational circuits, quantum kernels, and hybrid optimizers used in QAI workloads.

Gate-model superconducting qubits. Superconducting transmon/fluxonium platforms provide fast gate times and mature control electronics. IBM's Heron family exemplifies the current design point: a heavy-hex connectivity with tunable couplers and TLS-mitigation, with the r1 (133 qubits) advancing in 2023 and the r2 revision scaling to 156 qubits in 2024; IBM also demonstrated the 1,121-qubit *Condor* as a scale vehicle. Performance reporting increasingly emphasizes workload metrics (e.g., layer fidelity, CLOPS-like throughput) over device-level point estimates. Google's latest *Willow* chip (105 qubits) emphasizes surface-code building blocks with sub-threshold error rates and modular control, while Rigetti's *Ankaa-2* (84 qubits) targets higher two-qubit fidelities and compiler-hardware co-design for mid-depth circuits. These systems are the primary testbeds for shallow variational models and error-mitigation studies that matter to QAI.

Trapped-ion processors. Trapped ions offer long coherence and near all-to-all connectivity. Quantinuum's *H*-series reports state-of-the-art algorithmic benchmarks, including very high two-qubit fidelities and record quantum volume on commercial devices (H2: 56 fully connected qubits). IonQ emphasizes an application-oriented metric, *Algorithmic Qubits* (#AQ), recently reporting #AQ 64 on its *Tempo* class systems; all-to-all coupling and high-fidelity entanglers reduce routing overhead for dense token mixers, kernel methods, and compact attention-like quantum layers in hybrid models.

Neutral atoms (Rydberg and alkaline-earth). Neutral-atom arrays scale naturally and support programmable entanglement via Rydberg interactions. Atom Computing (alkaline-earth atoms) has demonstrated $> 10^3$-qubit registers with stable coherence and mid-circuit control features that are attractive for hybrid learning loops. Rydberg-array platforms from QuEra and Pasqal focus on analog/digital-analog simulation and combinatorial optimization; recent milestones include single-shot loading of $\sim$1,000 atoms and 200–300+ atom programmable arrays accessible via cloud backends. The architecture's reconfigurable graphs and native blockade dynamics align with QAI primitives that benefit from structured entanglement patterns and graph-native ansätze.

Photonic quantum computing. Photonic approaches leverage integrated silicon photonics, room-temperature operation, and natural networking. Xanadu's time-multiplexed photonic machines have demonstrated large-scale Gaussian boson sampling (216 entangled modes) as a stress test of photonic programmability; the roadmap is pivoting toward error-corrected GKP-style qubits and modular racks. PsiQuantum focuses on manufacturability, recently unveiling a "utility-scale" photonic chipset fabricated in advanced CMOS foundries, with a program centered on networking many photonic modules into a fault-tolerant system. For QAI, photonics is compelling for batched inference (high shot rates), optical interconnects, and future data-center integration once logical qubits mature.

Semiconductor silicon spin qubits. Spin qubits in silicon aim to inherit semiconductor manufacturing and dense scaling. Intel's 12-qubit *Tunnel Falls* arrays (300 mm wafers, EUV process) exemplify foundry-grade fabrication, while the academic and industrial ecosystem has shown $> 99\%$ two-qubit logic fidelities on small devices. In 2025, UK efforts, like Quantum Motion, began deploying early full-stack silicon-spin testbeds into national facilities, signaling progress toward CMOS-native, tightly integrated quantum tiles. For QAI, silicon spins promise compact, cryo-integrated accelerators once arrays and cryo-control mature.

Quantum annealing (Ising/QUBO). D-Wave's *Advantage2* systems based on the Zephyr graph offer >7,000 qubits with degree-20 connectivity and sub-millisecond anneals, delivering very high sampling throughput for large QUBOs. While not universal gate-model machines, modern annealers and hybrid solvers are practical subroutines for healthcare operations problems (scheduling, allocation) and for discrete search inside hybrid QAI pipelines.

Cross-architecture metrics and their QAI implications. Beyond raw qubit counts, three classes of metrics matter for QAI: (i) *fidelity/connectivity/latency* (how deep/wide a trainable circuit can be before noise dominates and how fast we can iterate); (ii) *workload-oriented benchmarks* (e.g., quantum volume, #AQ, layer-fidelity throughput), which better predict end-to-end training and inference costs; and (iii) *programmability features* such as mid-circuit measurement/reset, feedforward, dynamical decoupling, parametric gates, and robust calibration. Superconducting platforms currently excel at fast iteration and ecosystem tooling; trapped ions lead in fidelity and routing simplicity; neutral atoms lead in reconfigurability and scaling trajectories; photonics leads in manufacturability and networking potential; silicon spins lead in CMOS compatibility; annealers lead in high-throughput discrete optimization.

Hardware outlook (with radiology relevance). Near-term, error-mitigated mid-scale devices with better calibration automation and richer mid-circuit control will expand the reliable window for shallow variational ~~/kernal~~ or kernel circuits and hybrid search, enabling compact representation learning, spectral image reconstruction priors, and uncertainty calibration inside radiology pipelines. In parallel, annealing backends will continue to be useful for hospital operations (OR blocks, bed management, staff rostering) in hybrid solvers. Medium-term milestones likely to unlock larger QAI models are (i) higher-fidelity two-qubit gates and lower-cycle latency on superconducting and trapped-ion nodes, (ii) 10^3–10^4 stable, programmable neutral atoms with improved error suppression, and (iii) first logical-qubit demonstrations in photonics/silicon-spin stacks with practical interconnects. Across all stacks, workload-centric benchmarks and auditable noise/shot budgets should be reported alongside classical baselines to support clinically relevant claims in radiology AI.

2.3. Key AI Concepts for the Intersection

Understanding the AI side of the QAI intersection involves several core concepts. Machine Learning (ML) is a subfield of AI focused on creating systems that learn from data without explicit programming [17]. Key paradigms include supervised learning, where systems learn a mapping from inputs to outputs based on labeled examples (e.g., classification, regression); unsupervised learning, which finds patterns and structure in unlabeled data (e.g., clustering, dimensionality reduction); and Reinforcement Learning (RL), which involves learning optimal behavior through trial-and-error interactions with an environment, guided by reward signals [18]. ML tasks relevant to QAI include pattern recognition, classification, regression, clustering, and optimization ~~[?]~~[7].

Deep Learning (DL) or Deep Neural Networks (DNNs) represent a class of ML algorithms using multi-layered neural networks to learn hierarchical representations of data [6]. DNNs, including architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) ~~/~~or Long Short-Term Memory (LSTM) networks, have achieved state-of-the-art performance in various domains like computer vision and natural language processing but are computationally intensive to train and deploy [9].

Large Language Models (LLMs) are DNNs, typically based on the Transformer architecture [19], trained on vast amounts of text data [20]. Models like GPT and Llama exhibit remarkable capabilities in understanding and generating human language, translating languages, writing different kinds of creative content, and answering questions informatively. Their applications extend to technical domains, including code generation and explanation, making them relevant tools for quantum software development [21].

Optimization represents a fundamental task across AI and QC. In AI, it involves training models (adjusting parameters to minimize a loss function, e.g., via gradient descent) and tuning hyperparameters ~~[?]~~[7]. In QC, optimization problems arise in finding ground states of Hamiltonians (Variational Quantum Eigensolver or VQE), solving combinatorial problems (Quantum Approximate Optimization Algorithm or QAOA, Quantum Annealing or QA), and parameter tuning in Variational Quantum Algorithms (VQAs) [22].

2.4. Mathematical Formalism

The intersection of QC and AI relies on specific mathematical frameworks. Quantum mechanics is inherently described by linear algebra [23]. Quantum states are vectors in complex Hilbert spaces [24]. Quantum operations (gates) are represented by unitary matrices. Key concepts include inner products (for calculating probabilities and overlaps, $\langle \phi \rangle \psi$), outer products ($|\psi\rangle \langle \phi|$), tensor products (for describing multi-qubit systems) [14], and eigenvalues/eigenvectors (representing measurement outcomes and energy levels) [25]. This shared mathematical foundation with many ML algorithms, which also rely heavily on linear algebra for tasks like matrix manipulation and feature representation, facilitates the development of QML algorithms [26]. For example, quantum states in Hilbert space can serve as high-dimensional feature spaces for ML models [27].

In quantum physics, the Hamiltonian ($\hat{H}$) is an operator representing the total energy of a system [15]. Finding the ground state (lowest energy state) of a Hamiltonian is a crucial problem in physics and chemistry, and is the target of

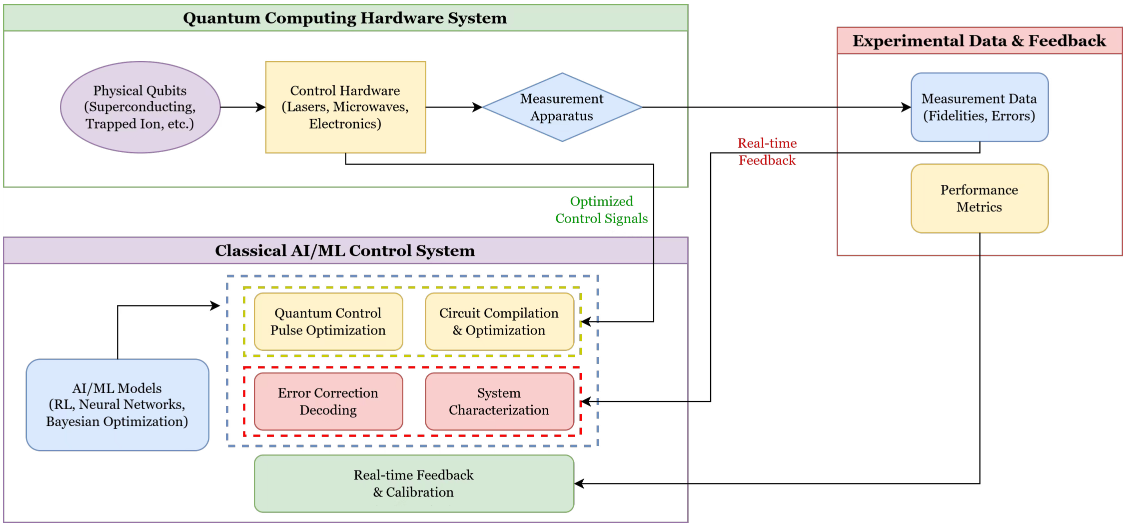


Figure 2: A schematic of the AI-for-QC paradigm. A classical AI control system utilizes real-time experimental data and performance metrics (e.g., fidelities, errors) from the quantum hardware to generate optimized control signals. This creates a feedback loop for continuous calibration, optimization, and error management, addressing key challenges in quantum hardware operation.

algorithms like the Variational Quantum Eigensolver (VQE) [28]. Hamiltonians can also be used to define the cost function in optimization problems mapped to quantum systems~~(e.g., QAOA)~~ [29]. The spectral decomposition,

$$\hat{H} = \sum_k \lambda_k |\phi_k\rangle \langle\phi_k| , \tag{2}$$

expresses the Hamiltonian in terms of its eigenvalues (λ_k, energy levels) and eigenstates ($|\phi_k\rangle$).

While pure states $|\psi\rangle$ describe isolated quantum systems, density matrices (ρ) provide a more general formalism to describe quantum states, including mixed states (probabilistic mixtures of pure states) and subsystems of entangled systems ~~[?]~~[4]. They are essential for describing realistic quantum systems subject to noise and decoherence. A density matrix is a positive semi-definite Hermitian operator with trace equal to 1 ($\mathrm{Tr}(\rho) = 1$) [30].

Evaluating the potential of quantum algorithms requires concepts from computational complexity theory. A key goal is achieving "quantum advantage," where a quantum algorithm solves a problem significantly faster than the best known classical algorithm [31]. Algorithms like Shor's factorization algorithm demonstrate exponential speedups for specific problems. The extended Church-Turing thesis posits that any reasonable computational model can be simulated by a classical Turing machine with polynomial overhead; quantum computers appear to challenge this thesis [16]. Understanding complexity classes, like Bounded-error Quantum Polynomial (BQP) time, helps delineate problems potentially amenable to quantum speedups.

The fundamental principles of QC, particularly superposition, entanglement, and interference, provide the underlying mechanisms that could enable computational advantages for AI tasks. Simultaneously, AI's strengths in handling complex data and optimizing high-dimensional systems are well-suited to address the inherent difficulties in building and controlling quantum computers. This complementary nature suggests a powerful synergy. Furthermore, the shared mathematical language of linear algebra acts as a critical bridge, enabling the direct application of quantum concepts to ML problems and facilitating the development of QML algorithms like the HHL algorithm (Harrow-Hassidim-Lloyd) and kernel methods [32]. However, the probabilistic nature of quantum measurement introduces inherent randomness into QAI algorithms [33]. While this aligns them conceptually with classical stochastic methods, the unique quantum phenomenon of interference allows for manipulating probabilities in ways not possible classically, potentially offering advantages but also requiring distinct analytical and optimization frameworks that account for statistical uncertainty and sampling noise [34].

3. Advancing Quantum Computing with AI and LLMs

The development of practical, large-scale quantum computers faces numerous obstacles related to hardware stability, control precision, circuit efficiency, and error management. Artificial intelligence and large language models are increasingly being employed as powerful tools to address these challenges across the entire quantum computing stack, acting as critical enabling technologies rather than just potential applications. Figure 2 illustrates the general workflow of this paradigm, where a classical AI system actively monitors, controls, and optimizes a quantum processor through a closed feedback loop.

3.1. AI in Quantum Hardware Development

Building robust and scalable quantum hardware remains a fundamental challenge in quantum computing. The complexity of system characterization scales exponentially with system size, as the Hilbert space dimension grows as 2^n for n qubits, making traditional methods like full process tomography computationally prohibitive. Machine learning techniques offer efficient alternatives by learning compressed representations of quantum dynamics.

Recent implementations demonstrate the power of this approach. Convolutional neural networks trained on experimental data from superconducting qubits can predict qubit coherence times with remarkable accuracy, achieving mean absolute errors below 5% compared to traditional exponential fitting methods ~~[?]~~[4]. These models learn the mapping:

$$f_\theta : \mathcal{X} \to (T_1, T_2, T_2^*) \tag{3}$$

where $\mathcal{X}$ represents the feature space of experimental measurements and θ denotes the neural network parameters. ~~Table ?? compares the efficiency of ML-based characterization against traditional methods.~~

~~Comparison of ML-based vs Traditional Characterization Methods **Method Time (hours) Accuracy (%)** Full Process Tomography 24-48 99.5 Randomized Benchmarking 4-6 95.0 ML-based Characterization 0.5-1 97.2~~

Hardware design optimization presents another domain where AI demonstrates significant advantages. Graph neural networks have been successfully employed to optimize qubit connectivity graphs, minimizing crosstalk while maintaining sufficient coupling for two-qubit gates [14]. The optimization objective can be formulated as:

$$\min_G \sum_{i,j} w_{ij} \cdot \text{Crosstalk}(i, j) - \lambda \cdot \text{Connectivity}(G) \tag{4}$$

where G represents the connectivity graph, w_{ij} are edge weights, and λ balances crosstalk minimization against connectivity requirements. Recent implementations on trapped-ion systems using ML-optimized trap geometries demonstrated a 23% reduction in motional heating rates compared to conventional designs, directly improving gate fidelities [14].

Automated calibration represents perhaps the most immediate impact of AI on quantum hardware operation. Maintaining quantum processors requires continuous optimization of hundreds of control parameters. The calibration problem involves finding optimal parameters:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathcal{F}(\boldsymbol{\theta}) \tag{5}$$

where $\mathcal{F}$ represents the gate fidelity measured through process tomography or randomized benchmarking. Bayesian optimization has emerged as particularly effective, with Google's quantum AI team reporting that their ML-based calibration system reduced full system calibration time from 8 hours to approximately 2 hours while maintaining gate fidelities above 99.5% [35]. The system uses Gaussian Process regression to model the fidelity landscape:

$$\mathcal{F}(\boldsymbol{\theta}) \sim \mathcal{GP}(\mu(\boldsymbol{\theta}), k(\boldsymbol{\theta}, \boldsymbol{\theta}')) \tag{6}$$

enabling efficient exploration of the high-dimensional parameter space.

3.2. AI-Driven Quantum Circuit Optimization and Compilation

The translation of high-level quantum algorithms into hardware-executable operations presents a complex combinatorial optimization problem. Traditional compilation techniques based on heuristic rules often produce suboptimal results, particularly for complex circuits or novel hardware architectures. Machine learning approaches, especially reinforcement learning, have demonstrated substantial improvements in circuit optimization metrics.

Table 1
Performance Comparison of Circuit Optimization Methods

Method	Depth Reduction	Gate Count	Transpilation Time (s)
Qiskit Transpiler [36]	Baseline	Baseline	0.5
t|ket⟩ [37]	15%	12%	1.2
RL-based [38]	38%	31%	5.8
QuGAN [39]	42%	35%	8.3

Note: Results summarize representative improvements reported for benchmark circuits relevant to variational and near-term quantum algorithms. Measurements are obtained from simulator-based or small-scale hardware evaluations under NISQ constraints, and should be interpreted as indicative relative trends rather than definitive end-to-end performance guarantees.

Deep reinforcement learning agents learn policies that map circuit states to transformation actions:

$$\pi_\theta : \mathcal{S} \to \mathcal{A} \tag{7}$$

where $\mathcal{S}$ represents the space of circuit configurations and $\mathcal{A}$ includes transformations such as gate fusion, cancellation, and commutation. The optimization considers hardware-specific constraints through a weighted cost function:

$$C = \alpha \cdot \text{depth}(C) + \beta \cdot \text{gates}(C) + \gamma \cdot \text{swaps}(C) \tag{8}$$

where the weights α, β, γ are tuned based on hardware-specific error rates and connectivity limitations.

Recent work using AlphaTensor-inspired techniques for quantum circuit synthesis demonstrated up to 42% reduction in circuit depth for common quantum algorithms like VQE and QAOA [40]. These improvements translate directly to higher success rates on NISQ devices where coherence times limit circuit depth.

Quantum Architecture Search (QAS) represents a paradigm shift in quantum circuit optimization by extending beyond fixed circuit structures to automatically discover hardware-tailored ansätze [41]. This adaptive approach has demonstrated significant performance improvements across multiple quantum computing platforms. On IBM quantum hardware, recent QAS implementations have achieved a 35% reduction in total gate count for Variational Quantum Eigensolver (VQE) circuits, while experiments on Google's Sycamore processor have shown a 28% improvement in convergence speed for Quantum Approximate Optimization Algorithm (QAOA) applications. Perhaps most notably, QAS has facilitated the discovery of novel ansatz structures that exhibit enhanced resilience to quantum noise, addressing one of the fundamental challenges in near-term quantum computing [42]. These advancements underscore the potential of automated architecture search to optimize quantum algorithms for specific hardware constraints while simultaneously improving computational efficiency and error mitigation.

3.3. AI for Quantum Control and Pulse Shaping

At the physical layer, quantum gates are implemented through precisely controlled electromagnetic pulses. The quantum control problem seeks to find control fields $u(t)$ that drive the system evolution:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H[u(t)]|\psi(t)\rangle \tag{9}$$

from initial state $|\psi_0\rangle$ to target state $|\psi_f\rangle$. The optimization objective, incorporating experimental constraints, becomes:

$$J[u] = 1 - |\langle \psi_f | U(T) | \psi_0 \rangle|^2 + \int_0^T \lambda(t) |u(t)|^2 dt \tag{10}$$

where the second term penalizes high-amplitude controls that may exceed hardware limitations.

Machine learning approaches have demonstrated significant advantages over traditional optimal control theory. Gradient-based optimization using automatic differentiation enables efficient exploration of the control landscape:

$$u_{k+1}(t) = u_k(t) - \eta \nabla_u J[u_k] \tag{11}$$

Table 2
QEC Decoder Performance Comparison

Decoder	Logical Error Rate	Decode Time	Training Data
MWPM [44]	10^{-3}	1 ms	N/A
Union-Find [45]	2×10^{-3}	0.5 ms	N/A
CNN Decoder [46]	5×10^{-4}	2 ms	10^6 samples
GNN Decoder [47]	3×10^{-4}	3 ms	10^7 samples
Transformer [48]	2×10^{-4}	5 ms	10^8 samples

Note: Results summarize representative decoder performance reported for surface-code–like architectures under circuit-level or phenomenological noise models relevant to near-term quantum devices. Values should be interpreted as indicative comparisons of scaling and trade-offs rather than absolute guarantees across hardware platforms.

where gradients are computed efficiently using the adjoint state method. IBM researchers reported achieving 99.9% gate fidelity for single-qubit gates using ML-optimized pulses, compared to 99.5% with traditional methods [35].

The optimization of multi-qubit gates presents unique challenges due to their highly non-convex control landscapes, where evolutionary algorithms have emerged as particularly effective solutions. Recent investigations on trapped-ion systems have demonstrated the power of this approach, achieving a 15% reduction in gate duration while maintaining fidelities exceeding 99% [43]. These optimized control sequences exhibit remarkable robustness, tolerating variations of up to $\pm 5\%$ in control amplitude without significant performance degradation. Furthermore, the evolutionary optimization process has led to the automatic discovery of composite pulse sequences that inherently cancel systematic errors, providing a built-in error suppression mechanism that enhances gate reliability without additional overhead. This convergence of shortened execution times, high fidelity, and intrinsic error mitigation represents a significant advancement in quantum gate implementation, particularly crucial for scaling up quantum processors where gate errors compound rapidly with circuit depth.

3.4. Machine Learning for Quantum Error Correction and Mitigation

Quantum error correction and mitigation are essential for reliable quantum computation. Machine learning approaches offer significant advantages in both domains, particularly in adapting to complex, realistic noise models.

For error mitigation, traditional Zero-Noise Extrapolation (ZNE) requires multiple circuit executions at different noise levels, incurring substantial sampling overhead. ML models can learn the noise-to-ideal mapping more efficiently:

$$\langle O \rangle_{\text{ideal}} = f_\theta(\langle O \rangle_{\lambda_1}, \langle O \rangle_{\lambda_2}, ..., \langle O \rangle_{\lambda_n}) \tag{12}$$

where λ_i represents different noise scaling factors. Recent implementations of Neural Noise Accumulation Surrogate (NNAS) models achieved 75% reduction in sampling overhead compared to standard ZNE while maintaining accuracy for circuits up to 20 qubits with depths exceeding 100 gates [34].

In quantum error correction, the decoding problem requires identifying the most likely error given syndrome measurements. While traditional decoders like Minimum Weight Perfect Matching (MWPM) scale as $O(n^3)$, ML decoders can achieve near-optimal performance with $O(n)$ complexity after training. Table 2 compares the performance of different decoder approaches.

Graph Neural Networks have shown particular promise for topological codes, achieving logical error rates 50% lower than MWPM for realistic noise models [49]. These networks learn to propagate syndrome information through the stabilizer graph:

$$h_i^{(k+1)} = \sigma\left(W^{(k)} h_i^{(k)} + \sum_{j \in \mathcal{N}(i)} M^{(k)} h_j^{(k)}\right) \tag{13}$$

Beyond improving existing codes, reinforcement learning enables automated discovery of new quantum error correcting codes. The RL agent optimizes over the space of stabilizer generators to maximize:

$$\mathcal{R} = \alpha \cdot d - \beta \cdot n - \gamma \cdot \text{connectivity}(S) \tag{14}$$

where d is the code distance, n is the number of physical qubits, and S represents the stabilizer group. Recent achievements include discovery of [[17,1,5]] codes optimized for heavy-hexagon connectivity that require 20% fewer physical qubits than standard surface codes [50].

3.5. The Role of LLMs in Quantum Computing Workflows

Large language models are beginning to impact quantum computing workflows, particularly in code generation and algorithm design. Fine-tuned LLMs have demonstrated capability in generating quantum circuits, achieving 85% syntactic correctness for QAOA circuit generation and 72% functional correctness for VQE ansatz design [20]. The quality of generated circuits improves significantly with structured prompting:

$$P(\text{correct}|\text{context}) = 0.85 \times P(\text{correct}) + 0.15 \tag{15}$$

where context includes problem specifications and hardware constraints.

Recent work on quantum feature map generation using LLMs showed promising results, with automatic generation of problem-specific encoding circuits improving classification accuracy by 30% compared to standard approaches [51]. However, significant limitations persist, including limited understanding of quantum mechanical constraints and tendency to generate classically simulable circuits.

The pervasive application of AI across the quantum computing stack has demonstrated transformative potential, with quantitative improvements ranging from 75% reduction in calibration time to 50% improvement in error correction rates. These advances establish AI not merely as a tool but as an essential component of the quantum computing ecosystem. As the field progresses toward fault-tolerant quantum computing, the symbiotic relationship between AI and quantum systems will likely deepen, with AI methods becoming increasingly critical for managing the complexity of large-scale quantum systems.

4. Enhancing AI and LLMs with Quantum Computing

The second major direction within QAI explores the potential for quantum computers to enhance or accelerate artificial intelligence, primarily through the field of Quantum Machine Learning (QML)[52]. QML seeks to develop and implement ML algorithms that leverage quantum phenomena like superposition, entanglement, and interference to gain advantages over classical ML approaches[53]. These advantages might manifest as computational speedups, improved model accuracy or generalization, enhanced data handling capabilities, or the ability to learn from smaller datasets[54]. The dominant approach to realize these goals in the near term is through a hybrid quantum-classical loop, as depicted in Figure 3.

4.1. Overview of Quantum Machine Learning Paradigms

Quantum machine learning research encompasses diverse approaches spanning different data types and algorithmic paradigms. Given the limitations of current Noisy Intermediate-Scale Quantum (NISQ) hardware, hybrid quantum-classical algorithms have emerged as the dominant framework[55, 56, 57]. These algorithms employ Parameterized Quantum Circuits (PQCs), also known as Variational Quantum Circuits (VQCs), executed on quantum processors[58, 59]. The quantum circuit parameters are optimized iteratively by classical algorithms based on measurement outcomes, creating a feedback loop that leverages quantum resources for specific computational tasks while relying on classical control and optimization[60, 61, 62].

The field distinguishes between processing classical data, which requires encoding into quantum states through techniques like amplitude or angle encoding, and directly processing quantum data from sensors or simulations. This distinction fundamentally affects algorithm design and potential advantages[63]. Variational Quantum Algorithms (VQAs) form the backbone of this hybrid approach, providing a flexible framework for implementing various QML models[64].

4.2. Key QML Algorithms and Techniques

4.2.1. Variational Quantum Algorithms (VQAs)

Variational Quantum Algorithms represent the cornerstone of near-term quantum machine learning, operating through an iterative hybrid loop that combines quantum state preparation and measurement with classical optimization. The general VQA framework begins with an initial quantum state $|\psi_0\rangle$, often the computational basis state $|0\rangle^{\otimes n}$, and

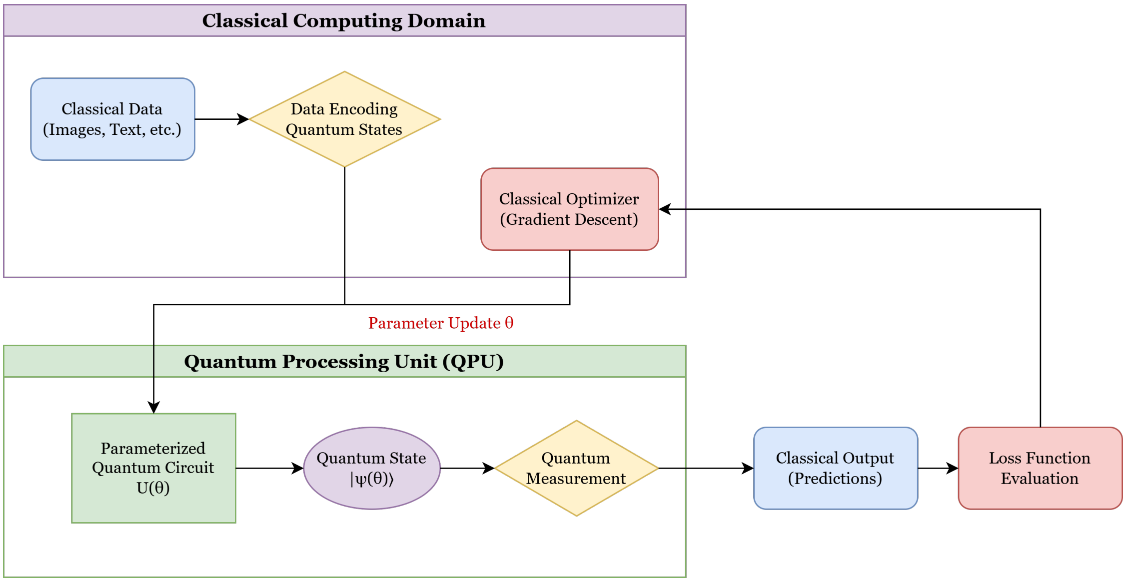


Figure 3: The general workflow for a hybrid quantum-classical algorithm in the QC-for-AI paradigm. Classical data is encoded into quantum states and processed by a parameterized quantum circuit (PQC) on a Quantum Processing Unit (QPU). The measurement outcomes are evaluated by a classical optimizer, which updates the circuit parameters (θ) to minimize a loss function, iterating until the model converges.

applies a parameterized quantum circuit $U(\theta)$ to generate a trial state $|\psi(\theta)\rangle = U(\theta)\,|\psi_0\rangle$. The expectation value of a problem-specific observable $\hat{O}$ is measured to evaluate a cost function:

$$C(\theta) = \langle \psi(\theta) | \hat{O} | \psi(\theta) \rangle \tag{16}$$

A classical optimizer then updates the parameters θ to minimize (or maximize) $C(\theta)$, leveraging the variational principle which guarantees that for any trial state, $\langle \psi | \hat{H} | \psi \rangle \geq E_0$, where E_0 is the true ground state energy~~[?]~~[1].

The success of VQAs critically depends on avoiding the barren plateau phenomenon, where cost function gradients vanish exponentially with system size[65]. This phenomenon manifests as

$$\left| \frac{\partial C}{\partial \theta_i} \right| \sim e^{-\alpha n} \tag{17}$$

where n is the number of qubits and α depends on the circuit architecture. Research has identified that structured ansätze, local cost functions, and layer-by-layer training can reduce α or change the scaling behavior[66]. The Variational Quantum Eigensolver (VQE) exemplifies the power of this approach in quantum chemistry applications, where recent implementations on NISQ devices have achieved chemical accuracy (within 1.6 mH) for small molecules like H_2 and LiH[14]. The algorithm minimizes $E(\theta) = \frac{\langle \psi(\theta) | H | \psi(\theta) \rangle}{\langle \psi(\theta) | \psi(\theta) \rangle}$ to approximate molecular ground states.

The Quantum Approximate Optimization Algorithm (QAOA) addresses combinatorial optimization problems through alternating applications of problem and mixer Hamiltonians:

$$|\psi(\gamma, \alpha)\rangle = \prod_{i=1}^{p} e^{-i\alpha_i H_M} e^{-i\gamma_i H_C} \,|+\rangle^{\otimes n} \tag{18}$$

Recent studies on MaxCut problems demonstrated that QAOA with $p = 3$ layers achieved 94% approximation ratio for graphs up to 20 nodes, though performance saturates for larger p due to noise accumulation[67]. Beyond these flagship algorithms, Variational Quantum Linear Solvers (VQLS) offer near-term alternatives to the fault-tolerant HHL algorithm, achieving 98% fidelity for 8×8 linear systems on trapped-ion hardware~~[?]~~[7].

Algorithm 1 VQE Algorithm

procedure VQE(H, $U(\theta)$, θ_0, Optimizer, ConvergenceCriteria)
  $\theta \leftarrow \theta_0$
  EnergyHistory ← []
  while not ConvergenceCriteriaMet(EnergyHistory) **do**
    CurrentEnergy ← EvaluateEnergyOnQPU(H, $U(\theta)$, θ)
    Append(EnergyHistory, CurrentEnergy)
    $\theta_{\text{new}} \leftarrow$ Optimizer.step(CurrentEnergy, θ)
    $\theta \leftarrow \theta_{\text{new}}$
  return min(EnergyHistory), θ

Table 3
Performance Comparison of VQA Implementations

Algorithm	Problem Size	Circuit Depth	Accuracy/Ratio	Hardware	Reference
VQE (UCCSD)	H_2O (14 qubits)	150-200	99.5% (1.6 mH error)	IBM-Q	[14]
QAOA ($p = 3$)	MaxCut (20 nodes)	60	94% approximation	Google Sycamore	[66]
VQLS	8×8 linear system	120	98% fidelity	Trapped ions	~~[?]~~ [7]
QNN (4-layer)	MNIST (4 qubits)	80	85% classification	Simulator	[29]
QGAN	8-dim distribution	100	92% JS divergence	IBMQ-Melbourne	[68, 69]

Note: Results summarize representative proof-of-concept demonstrations reported in prior work. Performance is task- and hardware-specific, obtained under NISQ constraints, and should be interpreted as illustrative benchmarks rather than evidence of general or scalable quantum advantage.

4.2.2. Quantum Neural Networks

Quantum Neural Networks adapt classical neural network concepts to quantum circuits, typically using VQCs as trainable layers where parameters θ correspond to network weights. The expressivity of QNNs can be characterized through their ability to approximate functions: $f(x) = \sum_i c_i \langle 0|U^\dagger(x) O_i U(x)|0\rangle$, where $U(x)$ encodes input data and O_i are measurement operators. This formulation reveals that QNNs operate in a fundamentally different space than classical networks, potentially offering advantages for specific function classes [9].

Quantum Convolutional Neural Networks (QCNNs) implement quantum analogues of convolution and pooling operations, where convolution applies local unitary transformations and pooling is achieved through measurements that reduce system size. Theoretical analysis shows QCNNs can avoid certain barren plateaus due to their local structure, a significant advantage over fully connected architectures [29]. Recent implementations achieved 97% accuracy on binary MNIST classification using only 4 qubits, demonstrating that quantum advantages can emerge even at small scales. The success stems from the QCNN's ability to capture global correlations through entanglement while maintaining trainability through local operations.

Hybrid Quantum-Classical Neural Networks combine classical preprocessing with quantum feature extraction, addressing the challenge of limited quantum resources. A typical architecture processes input data through classical layers, encodes the intermediate representation into a quantum state, applies a VQC, and measures to produce outputs. This approach has shown particular promise in reducing quantum resource requirements while maintaining performance benefits. For instance, a hybrid model for drug discovery achieved 89% accuracy in predicting molecular properties using 8 qubits, compared to 86% for a purely classical model with 100 times more parameters [34]. The quantum advantage appears to stem from the natural encoding of molecular symmetries in the quantum state space.

4.2.3. Quantum Support Vector Machines

Quantum Support Vector Machines leverage the exponentially large Hilbert space for implicit feature mapping, computing quantum kernel functions as $K(x_i, x_j) = |\langle \phi(x_i)|\phi(x_j)\rangle|^2$, where $|\phi(x)\rangle = U_\phi(x)|0\rangle^{\otimes n}$ represents the quantum feature map [24]. On engineered datasets designed to exhibit quantum advantage, such as those based on the discrete logarithm problem, QSVMs achieved perfect separation while classical SVMs achieved only 50-60%

Table 4
Quantum vs Classical SVM Performance Comparison

Dataset	Features	Classical SVM	QSVM	Quantum ~~Advantage~~Regime
Engineered (DLP) [70]	10	55%	100%	Exponential*
Iris [71]	4	96%	97%	Marginal
Wine [72]	13	94%	92%	None
MNIST (PCA-8) [73]	8	91%	93%	Marginal
Breast Cancer [74]	30	97%	95%	None

*Theoretical, requires fault-tolerant QC and efficient qRAM

Note: Results summarize representative QSVM benchmarks reported in prior work. Observed performance differences are dataset- and encoding-dependent. Theoretical exponential advantage applies only under fault-tolerant quantum computing assumptions with efficient qRAM, and does not reflect near-term NISQ capabilities.

accuracy [27]. However, for general datasets, the advantages are less pronounced due to the exponential concentration phenomenon.

As the number of qubits increases, kernel values concentrate around a fixed value: $\mathbb{E}[K(x_i, x_j)] - K(x_i, x_j) \sim e^{-n}$, making discrimination between data points exponentially difficult [33]. This concentration arises from the high expressivity of random quantum circuits, which generate states approximating Haar-random distributions. Additionally, estimating kernel entries to precision ϵ requires $O(1/\epsilon^2)$ measurements, leading to significant sampling costs for large kernel matrices. Recent theoretical work has identified strategies to mitigate concentration, including using structured feature maps that encode problem-specific inductive biases and limiting entanglement generation during encoding.

4.2.4. *The HHL Algorithm and Quantum Linear Systems*

The Harrow-Hassidim-Lloyd (HHL) algorithm promises exponential speedup for solving linear systems $Ax = b$ under specific conditions, operating in time $O(\log(N)s^2\kappa^2/\epsilon)$ where N is the matrix dimension, s is sparsity, κ is the condition number, and ϵ is the desired precision [32]. This scaling represents a potential exponential improvement over classical algorithms that scale polynomially with N. However, practical implementation faces significant challenges that often negate the theoretical advantage.

The state preparation bottleneck requires encoding the classical vector b into a quantum state $|b\rangle$, which generally requires $O(N)$ operations, eliminating the exponential speedup unless b has special structure. Similarly, the algorithm produces the solution as a quantum state $|x\rangle$ rather than the classical vector x, and extracting the full classical solution requires $O(N)$ measurements. The condition number dependence means that for poorly conditioned matrices ($\kappa \gg 1$), the algorithm's performance degrades significantly. These limitations have motivated the development of variational alternatives like VQLS that trade optimal scaling for NISQ compatibility, achieving moderate success on systems up to 8×8 matrices with 98% fidelity ~~[?]~~[7].

4.3. Quantum Transformers: Architectures and Applications

The integration of quantum computing with transformer architectures represents a frontier in QML research, aiming to leverage quantum parallelism for enhanced expressivity in attention mechanisms. Quantum Self-Attention mechanisms replace classical attention weights with quantum amplitudes, where the attention operation $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$ is implemented using quantum circuits with Q, K, V encoded as quantum states. The Quantum Mixed-State Self-Attention Network (QMSAN) extends this concept by using density matrices to represent token interactions, potentially capturing more complex dependencies than pure state representations [75].

Recent implementations have demonstrated promising results despite hardware limitations. QSANN achieved 89% accuracy on sentiment analysis tasks while using 75% fewer parameters than classical LSTM models, suggesting that quantum encoding can capture linguistic patterns more efficiently [76]. A related line of work introduces an Adaptive Quantum–Classical Fusion approach that dynamically assigns transformer layers to classical or quantum execution according to input complexity, combining entropy-driven circuits and quantum memory mechanisms for stable, efficient hybrid training [77]. In the same vein, CLAQS introduces a compact, fully-quantum token mixer that learns both complex-valued LCU mixing and a QSVT-based nonlinearity end-to-end, stabilizes training via ℓ_1-normalized amplitudes, and uses a dual-stage PQC to deliver strong text-classification accuracy under an 8-data-qubit budget [78]. In computer vision applications, HQViT demonstrated remarkable robustness to quantum noise, maintaining 92% accuracy with up to 1% gate error rates, significantly outperforming noise sensitivity expectations [79, 80]. For unsupervised learning, QClusformer improved clustering purity by 15% on CIFAR-10 compared to classical methods by leveraging quantum superposition to explore multiple cluster assignments simultaneously [81].

However, current quantum transformers face significant scalability challenges. Circuit depth constraints limit practical implementations to sequences shorter than 100 tokens, far below the thousands of tokens processed by classical transformers. The quadratic scaling of attention mechanisms becomes even more pronounced in quantum implementations due to the need for controlled operations between all token pairs. Additionally, the encoding of high-dimensional classical data into quantum states remains resource-intensive, often requiring ancillary qubits that increase circuit complexity without directly contributing to computational capacity.

Algorithm 2 Quantum Self-Attention Mechanism

```
function QuantumSelfAttention(X, W_Q, W_K, W_V)
    Q ← EncodeQuantum(X W_Q)
    K ← EncodeQuantum(X W_K)
    V ← EncodeQuantum(X W_V)
    Initialize quantum register |ψ⟩
    for each token pair (i, j) do
        Apply controlled rotation based on ⟨Q_i|K_j⟩
    |attention⟩ ← QuantumSoftmax(|ψ⟩)
    Output ← Measure(|attention⟩ ⊗ V)
    return Output
```

4.4. Quantum Kernels: Theory, Applications, and Challenges

Quantum kernel methods provide a systematic framework for leveraging quantum computers in machine learning through the kernel trick, where the quantum advantage potential stems from accessing an exponentially large feature space. However, this advantage is fundamentally limited by the concentration phenomenon. For random quantum feature maps, the expected kernel value between different inputs approaches

$$\mathbb{E}[K(x_i, x_j)] = \frac{1}{2^n} + O(2^{-2n}), \tag{19}$$

requiring exponentially many measurements to distinguish between data points [33]. This concentration arises because highly expressive quantum circuits generate states that approximate Haar-random distributions, leading to exponentially small overlaps between different encoded inputs.

Recent theoretical work has identified conditions for avoiding concentration while maintaining expressivity. Structured feature maps that encode data-dependent patterns into the circuit architecture can maintain distinguishability while leveraging quantum resources. For instance, feature maps of the form

$$U_\phi(x) = \prod_i e^{-i x_i P_i}, \tag{20}$$

where P_i are commuting Pauli operators, avoid concentration but may sacrifice expressivity. The trade-off between expressivity and trainability remains a fundamental challenge in quantum kernel design. Low-entanglement encodings that limit entanglement generation during the encoding process can maintain larger kernel values, but may not fully exploit the quantum Hilbert space. Problem-specific designs that incorporate domain knowledge into the feature map construction have shown the most promise, achieving practical advantages on structured datasets while avoiding the worst-case concentration behavior.

4.5. Assessing Quantum Speedups and Advantage in AI Tasks

Determining genuine quantum advantage for AI tasks requires careful analysis beyond theoretical speedup claims. Quantum algorithms offer potential speedups ranging from polynomial improvements, such as Grover's algorithm providing quadratic speedup for unstructured search, to exponential advantages like HHL for specific linear algebra problems. However, realizing these theoretical speedups in practice faces numerous obstacles. The end-to-end complexity must account for data encoding, circuit execution, measurement, and classical post-processing, often revealing that auxiliary requirements negate the core algorithmic advantage.

For near-term variational algorithms, demonstrating advantage is particularly challenging. While VQE has shown practical utility for small molecular simulations, scaling to industrially relevant molecules requires error rates below current capabilities. QAOA's performance on combinatorial optimization shows promise but lacks proven separation from classical algorithms. The best classical algorithms continually improve, raising the bar for quantum advantage claims. Recent benchmarking studies suggest that hybrid quantum-classical approaches may offer practical benefits through improved solution quality or convergence properties rather than pure speedup. For instance, quantum-enhanced feature spaces in QML have demonstrated improved generalization from limited training data, achieving 15-20% better performance than classical models on specific datasets with fewer than 1000 samples [24].

The definition of quantum advantage itself requires nuance beyond runtime considerations. Potential advantages include improved model accuracy, as demonstrated by QNNs achieving higher classification accuracy with fewer parameters; enhanced generalization capabilities, particularly evident in few-shot learning scenarios; the ability to naturally encode and process quantum data from sensors or simulations; and reduced sample complexity for specific learning tasks. These multifaceted advantages suggest that quantum enhancement of AI may manifest differently than traditional notions of computational speedup, requiring new metrics and benchmarks tailored to the unique capabilities of quantum processors.

4.6. Toward Quantumized Large Language Models

Beyond the dual axes of "AI for Quantum" and "Quantum for AI," an emerging frontier envisions the *quantumization of large language models* themselves. Classical LLMs approach the computational and energetic limits of von Neumann hardware, while their probabilistic embeddings and high-dimensional attention manifolds admit a natural formulation in Hilbert space. Embedding language computation into quantum representations could therefore transcend classical scaling constraints and introduce non-classical correlations, enabling superposed reasoning pathways and compact parameterizations.

A pragmatic roadmap is staged: (i) *quantum-inspired* designs that endow classical transformers with mixed-state encoders and amplitude-aware attention; (ii) *hybrid quantum–classical* transformers that execute selected attention heads or routing modules on parameterized quantum circuits under shot budgets and noise mitigation; and ultimately (iii) *fully quantum* state-based models that operate natively in Hilbert space and use measurement for sampling and decoding. In radiology-facing workflows, quantumized LLMs may natively encode diagnostic ambiguity, fuse multi-modal image–text evidence with entangled feature maps, and interface with quantum-enhanced sensing or reconstruction back-ends. While significant challenges remain—data encoding, trainability under noise, and end-to-end I/O latency—the potential for improved generalization under small data, calibrated uncertainty, and reduced carbon cost motivates sustained exploration within clinically grounded, auditable hybrid pipelines.

5. QAI for Healthcare: Opportunities, Challenges, and Deployment

5.1. Opportunities Across Healthcare Data and Tasks

Quantum AI (QAI) can contribute across the major healthcare data modalities and task families, with the most credible near-term pathways being hybrid, involving small, well-scoped quantum subroutines embedded in classical pipelines and evaluated with rigorous, clinically meaningful endpoints. [82, 83, 84]. To illustrate these opportunities, the discussion proceeds from structured electronic records to unstructured text, imaging, multi-omics, and ultimately system-level operations.

The first domain of interest is structured and semi-structured EHR data. Routinely collected labs, vitals, diagnoses, medications, and procedures underpin risk prediction, readmission modeling, and phenotype discovery [85, 86, 87]. In this tabular and often sparse setting, QAI offers several complementary strategies. Quantum-enhanced kernels can map high-dimensional covariates into rich Hilbert feature spaces to support margin-based learning [88]. Variational quantum circuits provide a means of representation learning that can respect structured clinical hierarchies such as ICD or ATC/RxNorm codes [89]. In addition, quantum-assisted feature selection, formulated as a QUBO problem, enables efficient navigation through noisy and correlated predictors [90]. Practically, the quantum component is a drop-in model class or search primitive wrapped by a classical workflow that handles calibration, missingness, and fairness.

Moving from structured records to unstructured narratives, clinical text, including notes and discharge summaries, is central to information extraction, cohort identification, and decision support, yet it is challenging due to domain shift, abbreviations, and long-range dependencies. Here, quantum-inspired or hybrid quantum–classical NLP can target few-shot classification [91] via task-specific quantum kernels, as well as sequence and graph representations using variational circuits or tensor-network–inspired ansätze [92, 93]. In practice, these appear as quantum kernels or hybrid encoders plugged into standard extraction or triage pipelines and evaluated for robustness under cross-site generalization.

Continuing along the data spectrum, medical imaging tasks such as reconstruction, triage, segmentation, and radiomics benefit when data are limited and search spaces are large. QAI can provide quantum-kernel and QSVM baselines in small-sample regimes and support quantum-amenable search over priors and regularizers for inverse problems or compact backbones in architecture design. Clinical grounding remains paramount, and evaluation should focus on patient-level outcomes, calibration, and prospective robustness rather than proxy image metrics alone.

Extending further to high-throughput biology, multi-omics and biomolecular data, including genomics, transcriptomics, proteomics, and metabolomics, pose nonlinear and combinatorial inference challenges. QAI is well suited to this domain in several ways. It can perform structured search over biomarker subsets through QUBO or QAOA formulations, support variational modeling that captures shared latent structure across multiple omics layers, and provide quantum-assisted similarity kernels that facilitate biomarker discovery and patient stratification [94]. These quantum pieces integrate naturally into established multi-omics frameworks, informing hypotheses prior to wet-lab validation.

Finally, stepping back from patient-level modeling to the delivery system itself, healthcare operations such as scheduling, logistics, and resource allocation including operating room blocks, nurse rostering, and bed management are natural fits for quantum optimization. Many such problems can be cast as QUBOs and addressed with QAOA or annealing inside hybrid pipelines that respect real-world constraints and SLAs [95]. Here, the quantum routine serves as a heuristic subsolver; performance should be reported in operational metrics, such as throughput, wait time, overtime, and stress-tested under realistic stochastic arrivals.

Across these settings, common guidance applies: start with tasks where strong classical baselines are hampered by data scarcity or combinatorial search; treat quantum pieces as interchangeable modules; and insist on transparent ablations, calibration, external validation, privacy-preserving training like federated settings, interpretability, and reproducibility, irrespective of the computational substrate.

5.2. Small-Data, High-Dimensionality, and Practical Modeling Challenges in Healthcare

Many clinically valuable problems fall into low-sample-size, high-dimensional regimes, such as rare diseases, pediatric subtypes, and narrowly defined cohorts, where sample efficiency and inductive bias dominate performance. In these settings, QAI is most credible as part of hybrid pipelines: quantum subroutines tackle the statistically or combinatorially brittle pieces, while classical code handles preprocessing, covariate shift checks, and evaluation [96, 97]. Concretely, kernel methods profit from expressive quantum feature maps that lift sparse, high-dimensional tabular or radiomic inputs into richer spaces for margin-based learning [98]. Variational quantum circuits can act as compact representation learners when model capacity must be tightly controlled relative to the available sample size [99, 100]. Quantum or quantum-inspired generative models can augment minority classes under strict privacy and distributional constraints [101, 102]. When cost and complexity are dominated by combinatorial exploration, such as selecting sparse biomarker panels or treatment policy sets, QUBO or QAOA quantum formulations offer a principled search primitive inside a larger heuristic loop [103, 104]. In all cases, the right baseline is not a straw man but the strongest classical method feasible under the same compute and data budgets.

Small-data regimes worsen a long-standing issue in healthcare, namely distribution shift, in which cross-site, cross-vendor, demographic, and temporal variations are routine, and gains on in-distribution test sets often fail to generalize out of distribution [105, 106]. Accordingly, QAI components should be judged under explicit OOD protocols, paired with multi-center external validation and ablations that isolate the incremental value of quantum modules over matched classical surrogates, with robustness readouts. These challenges are compounded by modality heterogeneity, as real-world fusion of EHR, omics, clinical text, and imaging demands normalization, principled handling of missingness, and careful timeline synchronization [107]. Practical stacks will remain hybrid, with classical pipelines taking cleaning, temporal alignment, and early, late, and intermediate fusion, while quantum subroutines are reserved for the most brittle subproblems [108, 60]. Clear interface contracts for I/O shapes, batching, and error handling should make the quantum piece a drop-in module rather than a bespoke one-off. Finally, privacy and governance constrain both augmentation and training [109]. Patient data typically sit behind institutional firewalls, so federated learning with secure aggregation and, where appropriate, differential privacy offers a deployable pattern that QAI can plug into by placing quantum subroutines on the client or server side [110, 111, 112]. When quantum communication or cryptographic primitives are contemplated, they must integrate with existing compliance frameworks and clinical IT rather than replace them.

Across these constraints, a common recipe emerges: choose tasks where classical baselines are strong but hampered by limited sample size and high dimensionality, or combinatorial search. Use quantum pieces as interchangeable modules, evaluate under OOD and multi-center protocols, report calibration and subgroup performance, and document cost–benefit trade-offs in wall-clock time, hardware access, and carbon footprint alongside accuracy. Reproducibility requires releasing code, circuits, seeds, and hardware configurations, plus classical surrogates that make the quantum contribution auditable.

5.3. Fairness, Transparency, Reporting, and Deployment for Healthcare QAI

Building QAI for healthcare is a socio-technical endeavor, not just a modeling exercise. Credible claims must pair accuracy with equity, interpretability, rigorous reporting, and plans for safe deployment. The most defensible near-term pattern is hybrid, with quantum subroutines addressing statistically or combinatorially brittle steps, while classical systems handle data curation, monitoring, and clinician-facing interfaces. A first requirement is fairness auditing beyond pooled metrics. Report clinically and operationally relevant subgroups, conduct disparity analyses with confidence intervals, document mitigations, and stress-test intersectional slices under cross-site and temporal shift [113, 114, 115, 116]. Pooled AUC alone is insufficient and should be complemented by subgroup tables, calibration curves, and clinically meaningful effect sizes [117, 118].

Transparency and reproducibility are essential. Because many quantum components are not directly interpretable, pair them with interpretable surrogates or attributions surfaced on the classical interface, such as SHAP or LIME on hybrid features or prototypes or nearest neighbors induced by quantum kernels [119, 120], while exposing well calibrated uncertainty with every prediction [121, 122]. Reproducibility requires reporting circuit design details, shot counts, noise mitigation settings, hardware identifiers, and classical baselines of comparable capacity, ensuring that any claimed quantum advantage is transparent and auditable. Reporting should align with medical AI guidance, including TRIPOD and TRIPOD-AI with PROBAST for predictive modeling, DECIDE-AI for early clinical evaluation, CONSORT-AI and SPIRIT-AI for randomized trials, and STARD and STARD-AI for diagnostic accuracy [123, 124]. Specify data sources, inclusion and exclusion criteria, preprocessing and handling of missing data, model selection and hyperparameter settings, evaluation protocols for in-distribution and out-of-distribution data, uncertainty estimates, and external validation. Use MRMC for imaging studies and analogous multi clinician evaluations with decision analytic endpoints such as net benefit for non imaging tasks [125, 126].

Real-world deployment introduces systems and regulatory concerns. Plan for hybrid execution, ~~Eelectronic~~ electronic health record and picture archiving and communication systems interoperability, robust machine learning operations, and safe rollout modes that preserve clinical workflow. Treat external multi-center validation as first-class, and predefine monitoring, drift detection, recalibration, and re-approval triggers across sites, coupled with post-market surveillance and periodic fairness and robustness re-audits under distribution shift [127, 128, 129]. Position QAI within SaMD pathways, emphasizing traceability, human oversight, well-scoped indications for use, and—where continuous learning is intended—pre-specified change-control plans aligned with IMDRF SaMD and good ML practice [130]. A minimal dossier should include subgroup performance and calibration plots, uncertainty summaries, model cards and dataset datasheets, external validation results, MRMC or multi-clinician studies as applicable, and a monitored deployment plan with clear rollback and re-approval criteria.

6. Key Challenges and Future Directions

To support consistent interpretation of reported QAI performance claims, we summarize the key bottlenecks and research directions that currently limit scalability and reproducibility. First, hardware noise and decoherence, limited qubit counts, and constrained connectivity impose strict depth budgets and make outcomes highly backend- and calibration-dependent. Second, many variational approaches face trainability limitations, for example barren plateaus, where gradients vanish and optimization becomes unstable without structured ansätze, careful initialization, and hardware-aware training strategies. Third, data encoding and end-to-end I/O costs, including state preparation, measurement, and sampling, can dominate wall-clock runtime and may offset theoretical speedups, particularly when high-dimensional classical data must be repeatedly embedded and read out. Fourth, effective performance is strongly influenced by compilation and routing overhead, which can inflate depth and two-qubit gate counts and thereby amplify noise sensitivity. Finally, the field still lacks standardized benchmarks and consistent resource reporting, including shots, qubits, depth, compilation targets, wall-clock timing, and uncertainty estimates, complicating cross-paper comparisons and slowing reproducible progress. Looking forward, near-term progress will likely come from improved error mitigation, hardware-aware compilation, and trainability-aware model design, alongside hybrid pipelines that explicitly account for end-to-end costs. Longer-term advances in fault-tolerant architectures and scalable quantum memory/IO primitives may unlock more robust forms of quantum advantage for select subroutines, but rigorous resource estimation and standardized evaluation will remain essential.

7. Conclusion

This survey has explored the multifaceted relationship between quantum computing and artificial intelligence, revealing a complex landscape of opportunities and challenges. The bidirectional nature of QAI illustrates how these technologies can mutually enhance each other: AI accelerates quantum computing development through improved hardware design, circuit optimization, and error correction, while quantum computing offers theoretical pathways to enhance AI through potential computational advantages for specific tasks.

The field exhibits a striking contrast between long-term theoretical promise and near-term practical realities. Many quantum algorithms for AI demonstrate compelling theoretical advantages under idealized conditions, yet face substantial implementation hurdles on current hardware. Meanwhile, AI methods are already delivering tangible benefits for quantum computing development, often outperforming traditional approaches for complex tasks like quantum error correction decoding and pulse optimization.

The comparative analysis highlights how different methodologies present distinct trade-offs. Near-term approaches using variational quantum algorithms offer accessibility on NISQ devices but face trainability limitations like barren plateaus. While fault-tolerant algorithms promise exponential speedups, the demanding hardware they require renders their practical implementation a distant prospect. AI techniques for quantum computing present a similar landscape of trade-offs: reinforcement learning excels at discovering novel strategies at the cost of significant training resources, whereas gradient-based methods offer efficiency for well-structured problems.

A domain that illustrates these trade-offs is healthcare. It couples heterogeneous data, like EHRs, multi-omics, imaging, and clinical text with small-sample and high-dimensional regimes and stringent demands for safety, equity, and accountability. In the near term, the most credible progress will come from hybrid quantum–classical pipelines. These frameworks deploy quantum subroutines selectively at points where classical algorithms encounter their sharpest bottlenecks. Representative applications include quantum-enhanced kernels for high-dimensional risk prediction, variational modules for compact representation learning, and quantum optimization for combinatorial biomarker or treatment-policy design. All the while, data curation, multimodal fusion, calibration, and uncertainty quantification remain anchored in mature classical methods. Credible evaluation in this domain also requires multi-center external validation, explicit out-of-distribution testing, subgroup-fairness analysis, privacy-preserving training, and auditable deployment paths.

Looking forward, QAI advancement requires addressing fundamental challenges across hardware, algorithms, and their integration. Critical questions remain about the precise conditions for quantum advantage in AI, effective mitigation strategies for trainability issues, and efficient approaches to classical-quantum interfaces. Progress will likely require innovations that go beyond simply scaling current approaches, instead developing structured, problem-aware methods that can effectively harness quantum phenomena for meaningful computational advantages.

The future of QAI will be shaped by interdisciplinary collaboration that integrates insights from quantum physics, computer science, mathematics, and engineering. While significant obstacles remain, the prospect of rewards such as accelerated scientific discovery and novel approaches for complex optimization and machine learning problems makes QAI a compelling research frontier.

References

[1] Matthias Klusch, Jörg Lässig, Daniel Müssig, Antonio Macaluso, and Frank K Wilhelm. Quantum artificial intelligence: a brief survey. *KI-Künstliche Intelligenz*, 38(4):257–276, 2024.

[2] Kamila Zaman, Alberto Marchisio, Muhammad Abdullah Hanif, and Muhammad Shafique. A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead. *arXiv preprint arXiv:2310.10315*, 2023.

[3] Linshu Chen, Tao Li, Yuxiang Chen, Xiaoyan Chen, Marcin Wozniak, Neal Xiong, and Wei Liang. Design and analysis of quantum machine learning: a survey. *Connection Science*, 36(1):2312121, 2024.

[4] Yuri Alexeev, Marwa H Farag, Taylor L Patti, Mark E Wolf, Natalia Ares, Alán Aspuru-Guzik, Simon C Benjamin, Zhenyu Cai, Shuxiang Cao, Christopher Chamberland, et al. Artificial intelligence for quantum computing. *Nature Communications*, 16(1):10829, 2025.

[5] Jun Qi, Chao-Han Huck Yang, Samuel Yen-Chi Chen, and Pin-Yu Chen. Quantum machine learning: an interplay between quantum computing and machine learning. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2025.

[6] Thien Nguyen, Tuomo Sipola, and Jari Hautamäki. Machine learning applications of quantum computing: A review. *arXiv preprint arXiv:2406.13262*, 2024.

[7] Sahil Tomar, Rajeshwar Tripathi, and Sandeep Kumar. Comprehensive survey of qml: from data analysis to algorithmic advancements. *arXiv preprint arXiv:2501.09528*, 2025.

[8] C Aishwarya, M Venkatesan, and P Prabhavathy. A scoping survey of quantum machine learning and deep learning for real-world applications. *Procedia Computer Science*, 258:633–646, 2025.

[9] Yunfei Wang and Junyu Liu. A comprehensive review of quantum machine learning: from nisq to fault tolerance. *Reports on Progress in Physics*, 2024.

[10] David Peral-García, Juan Cruz-Benito, and Francisco José García-Peñalvo. Systematic literature review: Quantum machine learning and its applications. *Computer Science Review*, 51:100619, 2024.

[11] Samuel Yen-Chi Chen, Shinjae Yoo, and Yao-Lung L Fang. Quantum long short-term memory. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8622–8626. IEEE, 2022.

[12] Jun Qi, Chao-Han Huck Yang, Samuel Yen-Chi Chen, and Pin-Yu Chen. Quantum machine learning: an interplay between quantum computing and machine learning. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2025.

[13] Yan Ge, Wu Wenjie, Chen Yuheng, Pan Kaisen, Lu Xudong, Zhou Zixiang, Wang Yuhan, Wang Ruocheng, and Yan Junchi. Quantum circuit synthesis and compilation optimization: Overview and prospects. *arXiv preprint arXiv:2407.00736*, 2024.

[14] Nguyen H Le, Modesto Orozco-Ruiz, Sahra A Kulmiya, James G Urquhart, Samuel J Hile, Winfried K Hensinger, and Florian Mintert. Amplitude-noise-resilient entangling gates for trapped ions. *arXiv preprint arXiv:2407.03047*, 2024.

[15] Kamila Zaman, Alberto Marchisio, Muhammad Abdullah Hanif, and Muhammad Shafique. A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead. *arXiv preprint arXiv:2310.10315*, 2023.

[16] Alexander M Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T Hann, Michael J Kastoryano, Emil T Khabiboulline, Aleksander Kubica, et al. Quantum algorithms: A survey of applications and end-to-end complexities. *arXiv preprint arXiv:2310.03011*, 2023.

[17] Yaswitha Gujju, Atsushi Matsuo, and Rudy Raymond. Quantum machine learning on near-term quantum devices: Current state of supervised and unsupervised techniques for real-world applications. *Physical Review Applied*, 21(6):067001, 2024.

[18] Samuel Yen-Chi Chen. An introduction to quantum reinforcement learning (qrl). In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1139–1144. IEEE, 2024.

[19] Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. Large language models: A survey. *arXiv preprint arXiv:2402.06196*, 2024.

[20] Linus Jern, Valter Uotila, Cong Yu, and Bo Zhao. Fine-tuning large language models on quantum optimization problems for circuit generation. *arXiv preprint arXiv:2504.11109*, 2025.

[21] Giordano d'Aloisio, Sophie Fortz, Carol Hanna, Daniel Fortunato, Avner Bensoussan, Eñaut Mendiluze Usandizaga, and Federica Sarro. Exploring llm-driven explanations for quantum algorithms. In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 475–481, 2024.

[22] Kirill Plekhanov, Matthias Rosenkranz, Mattia Fiorentini, and Michael Lubasch. Variational quantum amplitude estimation. *Quantum*, 6:670, 2022.

[23] Samuel Yen-Chi Chen and Zhiding Liang. Introduction to quantum machine learning and quantum architecture search. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–6. IEEE, 2025.

[24] Kuan-Cheng Chen, Tai-Yue Li, Yun-Yuan Wang, Simon See, Chun-Chieh Wang, Robert Wille, Nan-Yow Chen, An-Cheng Yang, and Chun-Yu Lin. Validating large-scale quantum machine learning: Efficient simulation of quantum support vector machines using tensor networks. *Machine Learning: Science and Technology*, 6(1):015047, 2025.

[25] Anderson Melchor Hernandez, Davide Pastorello, and Giacomo De Palma. Mean-field limit from general mixtures of experts to quantum neural networks. *arXiv preprint arXiv:2501.14660*, 2025.

[26] Walid El Maouaki, Taoufik Said, and Mohamed Bennai. Quantum support vector machine for prostate cancer detection: a performance analysis. *arXiv preprint arXiv:2403.07856*, 2024.

[27] Jan Schnabel and Marco Roth. Quantum kernel methods under scrutiny: a benchmarking study. *Quantum Machine Intelligence*, 7(1):58, 2025.

[28] J-Z Zhuang, Y-K Wu, and L-M Duan. Hardware-efficient variational quantum algorithm in a trapped-ion quantum computer. *Physical Review A*, 110(6):062414, 2024.

[29] Prashant Kumar Choudhary, Nouhaila Innan, Muhammad Shafique, and Rajeev Singh. Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction. *arXiv preprint arXiv:2503.15403*, 2025.

[30] Ze Hu, Ziqing Zhu, Linghua Zhu, Xiang Wei, Siqi Bu, and Ka Wing Chan. Advancing hybrid quantum neural network for alternative current optimal power flow. *arXiv preprint arXiv:2410.20275*, 2024.

[31] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[32] Xiaonan Liu, Haoshan Xie, Zhengyu Liu, and Chenyan Zhao. Survey on the improvement and application of hhl algorithm. In *Journal of Physics: Conference Series*, volume 2333, page 012023. IOP Publishing, 2022.

[33] Supanut Thanasilp, Samson Wang, Marco Cerezo, and Zoë Holmes. Exponential concentration in quantum kernel methods. *Nature communications*, 15(1):5200, 2024.

[34] Xiao-Yue Xu, Xin Xue, Tianyu Chen, Chen Ding, Tian Li, Haoyi Zhou, He-Liang Huang, and Wan-Su Bao. Physics-inspired machine learning for quantum error mitigation. *arXiv preprint arXiv:2501.04558*, 2025.

[35] Elizabeth Robertson, Luisa Esguerra, Leon Messner, Guillermo Gallego, and Janik Wolters. Machine-learning optimal control pulses in an optical quantum memory experiment. *Physical Review Applied*, 22(2):024026, 2024.

[36] Qiskit contributors. Qiskit, 2019. Computer program.

[37] Seyon Sivarajah, Silas Dilkes, Alexander Cowtan, Will Simmons, Alec Edgington, and Ross Duncan. t| ket>: a retargetable compiler for nisq devices. *Quantum Science and Technology*, 6(1):014003, 2020.

[38] Thomas Fösel, Petru Tighineanu, Talitha Weiss, and Florian Marquardt. Reinforcement learning with neural networks for quantum feedback. *Physical Review X*, 8(3):031084, 2018.

[39] Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930):47–53, 2022.

[40] Francisco JR Ruiz, Tuomas Laakkonen, Johannes Bausch, Matej Balog, Mohammadamin Barekatain, Francisco JH Heras, Alexander Novikov, Nathan Fitzpatrick, Bernardino Romera-Paredes, John van de Wetering, et al. Quantum circuit optimization with alphatensor. *Nature Machine Intelligence*, pages 1–12, 2025.

[41] 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.

[42] Abhishek Sadhu, Aritra Sarkar, and Akash Kundu. A quantum information theoretic analysis of reinforcement learning-assisted quantum architecture search. *Quantum Machine Intelligence*, 6(2):49, 2024.

[43] Lukas Broers and Ludwig Mathey. Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols. *Physical Review Research*, 6(1):013076, 2024.

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

[45] Nicolas Delfosse and Naomi H Nickerson. Almost-linear time decoding algorithm for topological codes. *Quantum*, 5:595, 2021.

[46] Savvas Varsamopoulos, Ben Criger, and Koen Bertels. Decoding small surface codes with feedforward neural networks. *Quantum Science and Technology*, 3(1):015004, 2017.

[47] Moritz Lange, Pontus Havström, Basudha Srivastava, Isak Bengtsson, Valdemar Bergentall, Karl Hammar, Olivia Heuts, Evert van Nieuwenburg, and Mats Granath. Data-driven decoding of quantum error correcting codes using graph neural networks. *Physical Review Research*, 7(2):023181, 2025.

[48] Johannes Bausch, Andrew W Senior, Francisco JH Heras, Thomas Edlich, Alex Davies, Michael Newman, Cody Jones, Kevin Satzinger, Murphy Yuezhen Niu, Sam Blackwell, et al. Learning high-accuracy error decoding for quantum processors. *Nature*, 635(8040):834–840, 2024.

[49] Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold. *Nature*, 638(8052):920–926, 2025.

[50] Volodymyr Sivak, Michael Newman, and Paul Klimov. Optimization of decoder priors for accurate quantum error correction. *Physical Review Letters*, 133(15):150603, 2024.

[51] Kenya Sakka, Kosuke Mitarai, and Keisuke Fujii. Automating quantum feature map design via large language models. *arXiv preprint arXiv:2504.07396*, 2025.

[52] Raghavendra M Devadas and T Sowmya. Quantum machine learning: A comprehensive review of integrating ai with quantum computing for computational advancements. *MethodsX*, page 103318, 2025.

[53] Maria Schuld and Nathan Killoran. Is quantum advantage the right goal for quantum machine learning? *Prx Quantum*, 3(3):030101, 2022.

[54] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean. Power of data in quantum machine learning. *Nature communications*, 12(1):2631, 2021.

[55] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner 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.

[56] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[57] Bikram Khanal and Pablo Rivas. Evaluating the impact of noise on variational quantum circuits in nisq era devices. In *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, pages 1658–1664. IEEE, 2023.

[58] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum science and technology*, 4(4):043001, 2019.

[59] Wim Lavrijsen, Ana Tudor, Juliane Müller, Costin Iancu, and Wibe De Jong. Classical optimizers for noisy intermediate-scale quantum devices. In *2020 IEEE international conference on quantum computing and engineering (QCE)*, pages 267–277. IEEE, 2020.

[60] Jonathan Wurtz and Peter J Love. Classically optimal variational quantum algorithms. *IEEE Transactions on Quantum Engineering*, 2:1–7, 2021.

[61] Aidan Pellow-Jarman, Ilya Sinayskiy, Anban Pillay, and Francesco Petruccione. A comparison of various classical optimizers for a variational quantum linear solver. *Quantum Information Processing*, 20(6):202, 2021.

[62] Kim Nicoli, Christopher J Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, Klaus-Robert Müller, Paolo Stornati, Pan Kessel, and Shinichi Nakajima. Physics-informed bayesian optimization of variational quantum circuits. *Advances in Neural Information Processing Systems*, 36:18341–18376, 2023.

[63] 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.

[64] 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.

[65] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018.

[66] Jack Cunningham and Jun Zhuang. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24(2):48, 2025.

[67] Jack Cunningham and Jun Zhuang. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24(2):48, 2025.

[68] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1):103, 2019.

[69] Pierre-Luc Dallaire-Demers and Nathan Killoran. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.

[70] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. *Quantum Information Processing*, 13(11):2567–2586, 2014.

[71] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[72] Stefan Aeberhard, Danny Coomans, and Olivier De Vel. Improvements to the classification performance of rda. *Journal of chemometrics*, 7(2):99–115, 1993.

[73] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 2002.

[74] William H Wolberg, W Nick Street, and Olvi L Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer letters*, 77(2-3):163–171, 1994.

[75] Fu Chen, Qinglin Zhao, Li Feng, Chuangtao Chen, Yangbin Lin, and Jianhong Lin. Quantum mixed-state self-attention network. *Neural Networks*, 185:107123, 2025.

[76] Guangxi Li, Xuanqiang Zhao, and Xin Wang. Quantum self-attention neural networks for text classification. *Science China Information Sciences*, 67(4):142501, 2024.

[77] Yi Pan, Hanqi Jiang, Junhao Chen, Yiwei Li, Huaqin Zhao, Lin Zhao, Yohannes Abate, Yingfeng Wang, and Tianming Liu. Bridging classical and quantum computing for next-generation language models. *arXiv preprint arXiv:2508.07026*, 2025.

[78] Junhao Chen, Yifan Zhou, Hanqi Jiang, Yi Pan, Yiwei Li, Huaqin Zhao, Wei Zhang, Yingfeng Wang, and Tianming Liu. Claqs: Compact learnable all-quantum token mixer with shared-ansatz for text classification. *arXiv preprint arXiv:2510.06532*, 2025. Version v1.

[79] Hui Zhang, Qinglin Zhao, Mengchu Zhou, and Li Feng. Hqvit: Hybrid quantum vision transformer for image classification. *arXiv preprint arXiv:2504.02730*, 2025.

[80] Eyup B Unlu, Marçal Comajoan Cara, Gopal Ramesh Dahale, Zhongtian Dong, Roy T Forestano, Sergei Gleyzer, Daniel Justice, Kyoungchul Kong, Tom Magorsch, Konstantin T Matchev, et al. Hybrid quantum vision transformers for event classification in high energy physics. *Axioms*, 13(3):187, 2024.

[81] Xuan-Bac Nguyen, Hoang-Quan Nguyen, Samuel Yen-Chi Chen, Samee U Khan, Hugh Churchill, and Khoa Luu. Qclusformer: A quantum transformer-based framework for unsupervised visual clustering. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 2, pages 347–352. IEEE, 2024.

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

[83] James CL Chow. Quantum computing in medicine. *Medical Sciences*, 12(4):67, 2024.

[84] Thomas JS Durant, Elizabeth Knight, Brent Nelson, Sarah Dudgeon, Seung J Lee, Dominic Walliman, Hobart P Young, Lucila Ohno-Machado, and Wade L Schulz. A primer for quantum computing and its applications to healthcare and biomedical research. *Journal of the American Medical Informatics Association*, 31(8):1774–1784, 2024.

[85] Riddhi S Gupta, Carolyn E Wood, Teyl Engstrom, Jason D Pole, and Sally Shrapnel. A systematic review of quantum machine learning for digital health. *npj Digital Medicine*, 8(1):237, 2025.

[86] Zoran Krunic, Frederik F Flöther, George Seegan, Nathan D Earnest-Noble, and Omar Shehab. Quantum kernels for real-world predictions based on electronic health records. *IEEE Transactions on Quantum Engineering*, 3:1–11, 2022.

[87] 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.

[88] Omar Shehab, Zoran Krunic, Frederik Floether, George Seegan, and Nate Earnest-Noble. Quantum kernels for electronic health records classification. In *APS March Meeting Abstracts*, volume 2022, pages S37–006, 2022.

[89] Kostas Blekos, Dean Brand, Andrea Ceschini, Chiao-Hui Chou, Rui-Hao Li, Komal Pandya, and Alessandro Summer. A review on quantum approximate optimization algorithm and its variants. *Physics Reports*, 1068:1–66, 2024.

[90] Gabriel San Martin Silva and Enrique Lopez Droguett. Quantum-based combinatorial optimization for optimal sensor placement in civil structures. *arXiv preprint arXiv:2305.08738*, 2023.

[91] Juan P Rubio-Perez. A few shot learning scheme for quantum natural language processing. *arXiv preprint arXiv:2410.01832*, 2024.

[92] Lee J O'Riordan, Myles Doyle, Fabio Baruffa, and Venkatesh Kannan. A hybrid classical-quantum workflow for natural language processing. *Machine Learning: Science and Technology*, 2(1):015011, 2020.

[93] SM Tomal, Abdullah Al Shafin, Debojit Bhattacharjee, MD Amin, and Rafiad Sadat Shahir. Quantum-enhanced attention mechanism in nlp: A hybrid classical-quantum approach. *arXiv preprint arXiv:2501.15630*, 2025.

[94] Hakan Doga, Aritra Bose, M Emre Sahin, Joao Bettencourt-Silva, Anh Pham, Eunyoung Kim, Alan Andress, Sudhir Saxena, Laxmi Parida, Jan Lukas Robertus, et al. How can quantum computing be applied in clinical trial design and optimization? *Trends in Pharmacological Sciences*, 45(10):880–891, 2024.

[95] Agostino Marengo and Vito Santamato. Quantum algorithms and complexity in healthcare applications: a systematic review with machine learning-optimized analysis. *Frontiers in Computer Science*, 7:1584114, 2025.

[96] Yudai Suzuki, Hiroshi Yano, Qi Gao, Shumpei Uno, Tomoki Tanaka, Manato Akiyama, and Naoki Yamamoto. Analysis and synthesis of feature map for kernel-based quantum classifier. *Quantum Machine Intelligence*, 2(1):9, 2020.

[97] Elies Gil-Fuster, Jens Eisert, and Vedran Dunjko. On the expressivity of embedding quantum kernels. *Machine Learning: Science and Technology*, 5(2):025003, 2024.

[98] Utkarsh Singh, Jean-Frédéric Laprade, Aaron Z Goldberg, and Khabat Heshami. A resource efficient quantum kernel. *arXiv preprint arXiv:2507.03689*, 2025.

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

[100] Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Yashar Mansoori, and José Millet. Variational quantum circuits for machine learning. an application for the detection of weak signals. *Applied Sciences*, 11(14):6427, 2021.

[101] Graham R Enos, Matthew J Reagor, and Eric Hulburd. Evaluating quantum generative models via imbalanced data classification benchmarks. *arXiv preprint arXiv:2308.10847*, 2023.

[102] Jonathan Romero and Alán Aspuru-Guzik. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. *Advanced Quantum Technologies*, 4(1):2000003, 2021.

[103] Frederik F Flöther, Daniel Blankenberg, Maria Demidik, Karl Jansen, Raga Krishnakumar, Rajiv Krishnakumar, Nouamane Laanait, Laxmi Parida, Carl Y Saab, and Filippo Utro. How quantum computing can enhance biomarker discovery. *Patterns*, 6(6), 2025.

[104] Leo Zhou, Joao Basso, and Song Mei. Statistical estimation in the spiked tensor model via the quantum approximate optimization algorithm. *Advances in Neural Information Processing Systems*, 37:28537–28588, 2024.

[105] Jenny Yang, Andrew AS Soltan, and David A Clifton. Machine learning generalizability across healthcare settings: insights from multi-site covid-19 screening. *NPJ digital medicine*, 5(1):69, 2022.

[106] Jessica Schrouff, Natalie Harris, Sanmi Koyejo, Ibrahim M Alabdulmohsin, Eva Schnider, Krista Opsahl-Ong, Alexander Brown, Subhrajit Roy, Diana Mincu, Christina Chen, et al. Diagnosing failures of fairness transfer across distribution shift in real-world medical settings. *Advances in Neural Information Processing Systems*, 35:19304–19318, 2022.

[107] Sandeep Suresh Cranganore, Vincenzo De Maio, Ivona Brandic, and Ewa Deelman. Paving the way to hybrid quantum–classical scientific workflows. *Future Generation Computer Systems*, 158:346–366, 2024.

[108] Weitang Li, Zhi Yin, Xiaoran Li, Dongqiang Ma, Shuang Yi, Zhenxing Zhang, Chenji Zou, Kunliang Bu, Maochun Dai, Jie Yue, et al. A hybrid quantum computing pipeline for real world drug discovery. *Scientific Reports*, 14(1):16942, 2024.

[109] Rocco Ballester, Jesus Cerquides, and Luis Artiles. Quantum federated learning: a comprehensive literature review of foundations, challenges, and future directions. *Quantum Machine Intelligence*, 7(2):1–29, 2025.

[110] Arnaldo Rafael Camara Araujo, Ogobuchi Daniel Okey, Muhammad Saadi, Pablo Adasme, Renata Lopes Rosa, and Demóstenes Zegarra Rodríguez. Quantum-assisted federated intelligent diagnosis algorithm with variational training supported by 5g networks. *Scientific Reports*, 14(1):26333, 2024.

[111] Ruozhou Zuo, Haibo Tian, Zhiyuan An, and Fangguo Zhang. Post-quantum privacy-preserving aggregation in federated learning based on lattice. In *International Symposium on Cyberspace Safety and Security*, pages 314–326. Springer, 2022.

[112] C Srivenkateswaran, A Jaya Mabel Rani, R Senthil Kumaran, and R Vinston Raja. Securing healthcare data: A federated learning framework with hybrid encryption in cluster environments. *Technology and Health Care*, 33(3):1232–1257, 2025.

[113] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

[114] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew BA McDermott, Irene Y Chen, and Marzyeh Ghassemi. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine*, 27(12):2176–2182, 2021.

[115] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, pages 5637–5664. PMLR, 2021.

[116] Luke Oakden-Rayner, Jared Dunnmon, Gustavo Carneiro, and Christopher Ré. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In *Proceedings of the ACM conference on health, inference, and learning*, pages 151–159, 2020.

[117] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

[118] Andrew J Vickers, Ben van Calster, and Ewout W Steyerberg. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and prognostic research*, 3(1):18, 2019.

[119] Yujia Zhang, Kuangyan Song, Yiming Sun, Sarah Tan, and Madeleine Udell. " why should you trust my explanation?" understanding uncertainty in lime explanations. *arXiv preprint arXiv:1904.12991*, 2019.

[120] Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju. Fooling lime and shap: Adversarial attacks on post hoc explanation methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 180–186, 2020.

[121] Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 2023.

[122] Daniele Pelosi, Diletta Cacciagrano, and Marco Piangerelli. Explainability and interpretability in concept and data drift: A systematic literature review. *Algorithms*, 18(7):443, 2025.

[123] Gary S Collins, Johannes B Reitsma, Douglas G Altman, and Karel GM Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *Journal of British Surgery*, 102(3):148–158, 2015.

[124] Robert F Wolff, Karel GM Moons, Richard D Riley, Penny F Whiting, Marie Westwood, Gary S Collins, Johannes B Reitsma, Jos Kleijnen, Sue Mallett, and PROBAST Group†. Probast: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine*, 170(1):51–58, 2019.

[125] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[126] Andrew J Vickers and Elena B Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574, 2006.

[127] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.

[128] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

[129] John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11):e1002683, 2018.

[130] Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1):30–36, 2019.

Revised Manuscript - Clean Click here to view linked References

Quantum Artificial Intelligence: A Comprehensive Survey

Hanqi Jiang[a,1], Yi Pan[a,1], Junhao Chen[a,1], Zhengliang Liu[a], Lichao Sun[b], Quanzheng Li[c], Lu Zhang[d], Dajiang Zhu[e], Xianqiao Wang[a], Wei Liu[f], Xiang Li[c], Gang Li[h], Wei Zhang[i], Lin Zhao[j], Xiaowei Yu[k], Yingfeng Wang[l] and Tianming Liu[a,*]

[a]School of Computing, University of Georgia, 415 Boyd Graduate Studies Research Center, Athens, 30602, Georgia, United States
[b]Department of Computer Science and Engineering, Lehigh University, 19 Memorial Drive West, Bethlehem, 18015, Pennsylvania, United States
[c]Department of Radiology, Harvard University, 55 Fruit Street, Boston, 02114, Massachusetts, United States
[d]Department of Computer Science, Indiana University Indianapolis, 535 W. Michigan Street, Indianapolis, 46202, Indiana, United States
[e]Department of Computer Science and Engineering, University of Texas at Arlington, 500 UTA Boulevard, Arlington, 76019, Texas, United States
[f]Department of Radiation Oncology, Mayo Clinic, 5881 East Mayo Boulevard, Phoenix, 85054, Arizona, United States
[g]Department of Radiology, Massachusetts General Hospital and Harvard Kempner Institute, 55 Fruit Street, Boston, 02114, Massachusetts, United States
[h]Department of Radiology, University of North Carolina at Chapel Hill, 130 Mason Farm Road, Chapel Hill, 27599, North Carolina, United States
[i]School of Computer and Cyber Sciences, Augusta University, 1120 15th Street, Augusta, 30912, Georgia, United States
[j]Department of Biomedical Engineering, New Jersey Institute of Technology, University Heights, Newark, 07102, New Jersey, United States
[k]Department of Computer Science, Missouri University of Science and Technology, 500 West 15th Street, Rolla, 65409, Missouri, United States
[l]Department of Computer Science and Engineering, University of Tennessee at Chattanooga, 615 McCallie Ave, Chattanooga, 37403, Tennessee, United States

ARTICLE INFO

Keywords:
Quantum artificial intelligence
Hybrid quantum–classical computing
Variational quantum algorithms
Quantum transformers
NISQ devices

Abstract

Quantum Artificial Intelligence (QAI) has emerged at the nexus of quantum computing and AI, promising to redefine computational frontiers. This survey critically synthesizes the state-of-the-art through 2024, elucidating the profound bidirectional synergy between these fields. We analyze how classical machine learning is accelerating quantum hardware control, circuit optimization, and error correction. Conversely, we assess the potential quantum advantage of algorithms, including variational and kernel-based methods, across domains such as drug discovery, financial modeling, and cybersecurity. Our analysis reveals a critical trade-off between the utility of near-term Noisy Intermediate-Scale Quantum (NISQ) devices and the long-term promise of fault-tolerant architectures. We identify fundamental obstacles to QAI's advancement, including hardware decoherence, algorithmic barren plateaus, and data-encoding bottlenecks. While QAI's potential is transformative, achieving practical quantum advantage requires a concerted effort to overcome these core challenges at the hardware-software interface. This work provides a roadmap for navigating the current landscape and prioritizing future research in this rapidly evolving discipline.

1. Introduction

1.1. The Convergence of Quantum Computing and Artificial Intelligence

The confluence of quantum computing (QC) and artificial intelligence (AI), particularly machine learning (ML) and large language models (LLMs), marks the emergence of a transformative field often termed Quantum Artificial Intelligence (QAI) [1]. This interdisciplinary domain investigates the synergistic potential arising from integrating quantum mechanical principles with sophisticated AI algorithms. The core premise of QAI rests on a bidirectional relationship: quantum systems' unique computational capabilities can address intractable problems within AI, while AI's power can overcome significant hurdles in the development and operation of quantum computers.

Quantum computing harnesses phenomena such as superposition and entanglement to offer fundamentally new paradigms for information processing, promising capabilities that surpass classical machines for specific computational tasks [2]. Concurrently, AI has demonstrated remarkable success in extracting patterns from vast datasets, automating complex decision-making processes, and optimizing intricate systems, particularly through recent advances in ML and the development of powerful LLMs [3].

*Corresponding author.
ORCID(s):
[1]These authors have contributed equally to this work.

The field of QAI explores how these complementary strengths can be synergistically combined through two primary research directions. The first, "Quantum for AI," investigates quantum algorithms that could potentially accelerate or enhance ML tasks, optimization problems, and data analysis. The second, "AI for Quantum," applies AI techniques to address challenges in quantum hardware design, control, error correction, and algorithm discovery [4]. This dual approach signifies QAI's evolution from a purely theoretical concept into a field with tangible, albeit nascent, potential for transformative technological impact.

Two concurrent trends make this an auspicious moment for QAI. On the quantum side, prototype devices with tens to low hundreds of qubits are increasingly stable and programmable, enabling experimental studies of hybrid quantum–classical workflows in the so-called NISQ (noisy intermediate-scale quantum) regime [2]. On the AI side, LLMs and modern ML pipelines have scaled to billions of parameters and industrial deployments, but face escalating computational and energy costs [3]. QAI is motivated by the possibility that carefully chosen quantum subroutines can provide algorithmic or constant-factor advantages for certain ML primitives, while advanced AI methods can in turn accelerate the path toward useful quantum computation by improving calibration, control, and fault management [4, 1].

1.2. Motivation and Potential Impact

The motivation for exploring QAI stems significantly from the recognized limitations of classical computation in the face of escalating demands from modern science and technology. The relentless growth in data volume and the complexity of contemporary AI models, particularly deep learning architectures and LLMs, strain the capabilities of current computing infrastructures. Moore's Law, the historical driver of exponential growth in classical hardware performance, is encountering physical limits, making it increasingly difficult and expensive to achieve further significant gains in computational power through traditional means. Training state-of-the-art LLMs, for instance, requires massive investments in time and hardware resources, accessible only to a few entities. Quantum computing emerges as a potential pathway to transcend these limitations. By exploiting quantum mechanical effects, QC promises to efficiently solve certain classes of problems currently considered intractable for even the most powerful classical supercomputers. This includes complex optimization problems frequently encountered in logistics and finance, simulations of quantum systems crucial for materials science and drug discovery, and specific linear algebra tasks underlying some ML algorithms.

Conversely, the very nature of quantum computing presents formidable challenges arising from its operation based on counterintuitive quantum phenomena, its susceptibility to noise, and the complexity of controlling and scaling quantum hardware. The high-dimensional Hilbert spaces, the intricacies of quantum dynamics, and the need for precise control make designing, calibrating, and operating quantum computers exceptionally difficult. Here, AI, with its proficiency in pattern recognition, data-driven learning, and optimization in high-dimensional spaces, offers essential tools. AI algorithms can analyze experimental data to characterize quantum systems, optimize control pulses for quantum gates, design efficient quantum circuits, develop error correction strategies, and potentially even discover novel quantum algorithms. This interplay suggests that progress in quantum computing may itself depend significantly on advancements in AI.

The potential impact of successful QAI development spans numerous domains. Accelerated drug discovery and materials science through enhanced quantum simulations [5], improved financial modeling and optimization, breakthroughs in optimization problems relevant to logistics and supply chains, enhanced capabilities in cybersecurity through quantum-resistant cryptography and potentially quantum-enhanced attack analysis [6], and fundamentally new approaches to scientific discovery through the analysis of complex datasets [7] represent just some of the anticipated applications.

1.3. Survey Structure

This survey provides a comprehensive overview of the QAI landscape, synthesizing current research and highlighting key developments, challenges, and future trajectories. Section 2 establishes the theoretical groundwork, outlining the fundamental concepts of both quantum computing and relevant AI paradigms, along with the essential mathematical formalism. Section 3 delves into the application of AI and LLMs to advance quantum computing technologies, covering areas from hardware design and calibration to circuit optimization and error correction (AI for QC). Section 4 explores the converse direction: the potential for quantum computing to enhance AI and LLM capabilities, focusing on Quantum Machine Learning (QML) algorithms and their prospective advantages (QC for AI). Section 4.2 provides a comparative analysis of the various methodologies discussed in Sections 3 and 4, evaluating their strengths, weaknesses, and performance characteristics.

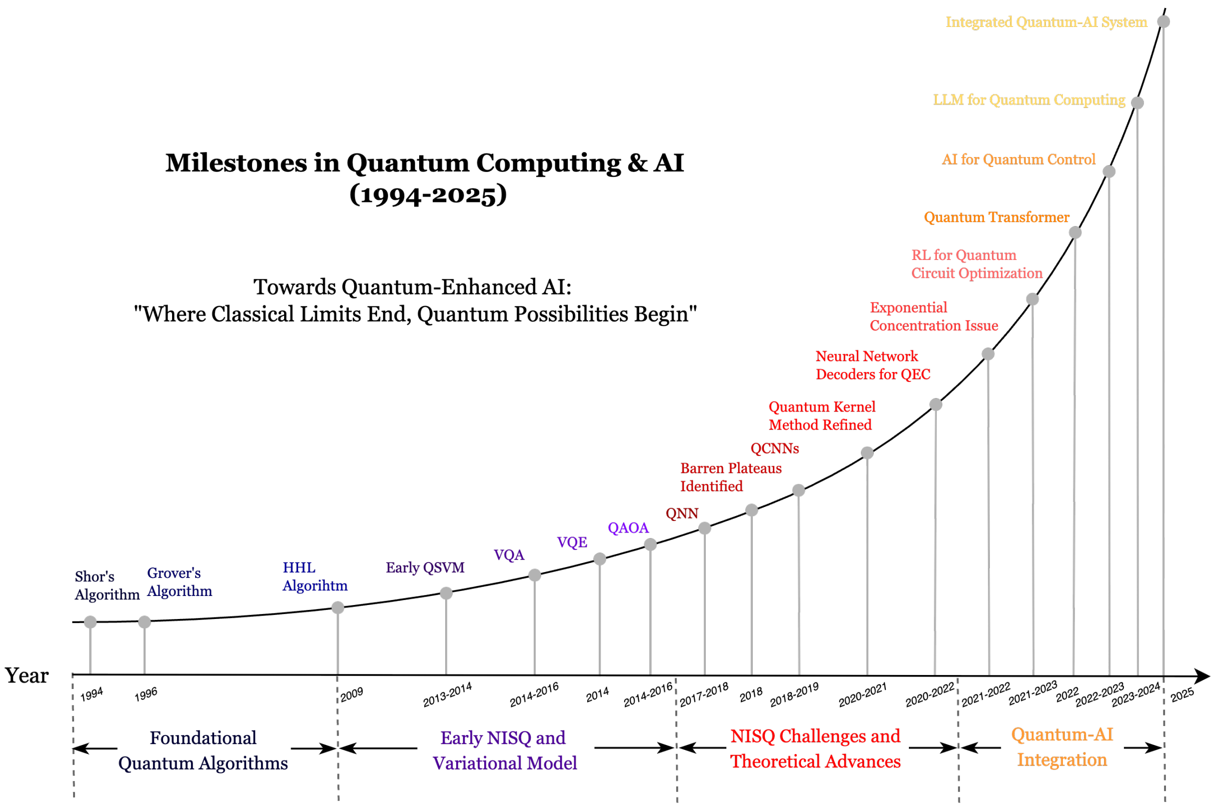


Figure 1: Key algorithmic milestones at the intersection of AI and quantum computing.

1.4. Contributions and positioning relative to recent surveys

Recent surveys on QAI and QML provide complementary but often directional or scope-limited perspectives [1, 8]. On the AI-for-quantum side, high-profile reviews organize how AI supports the quantum stack, including hardware design, calibration and control, compilation, and error mitigation, while explicitly not covering the reciprocal quantum-for-AI direction [4]. On the quantum-for-AI side, recent quantum machine learning surveys comprehensively cover broad methodological overviews ranging from NISQ to fault-tolerant regimes [9], examine near-term hardware-executed supervised and unsupervised demonstrations together with their practical bottlenecks [8], and present systematic or taxonomy-driven perspectives that classify algorithms, software frameworks, and datasets [7, 10].

Our survey is positioned to complement and extend these lines of work in three concrete ways. First, we provide a bidirectional, end-to-end roadmap that treats AI for quantum and quantum for AI under a single set of definitions and evaluation principles. We connect algorithmic performance summaries to NISQ-relevant caveats, including noise and error mitigation overheads, data-encoding costs, trainability and barren-plateau phenomena, and end-to-end wall-clock and resource accounting, to avoid over-interpreting small-scale comparisons. Second, beyond consolidating established paradigms such as variational quantum algorithms, quantum kernels, and hybrid models, we incorporate a 2024–2025 update spanning the hardware landscape, compilation and optimization, quantum error correction and decoding, and emerging model families, including transformer-style architectures and domain-specific pipelines, while emphasizing which claims rely on fault-tolerant assumptions versus what is currently testable on NISQ hardware. Third, distinct from general-purpose QAI and QML surveys and from domain-specific healthcare summaries, we introduce a radiology-anchored translation lens that maps QAI components to imaging workflows and provides deployment-minded guidance on data governance, reporting, subgroup performance, calibration, and regulatory and ethical considerations that is actionable for an imaging AI audience.

In short, while recent surveys provide substantial depth within individual facets of QAI and QML, this article is intended as a practical, compliance-aware roadmap that integrates both directions and makes NISQ realism explicit whenever claims of advantage or utility are discussed.

2. Foundational Concepts and Theoretical Frameworks

A robust understanding of the QAI intersection requires familiarity with the core principles of both quantum computing and artificial intelligence, as well as the mathematical language that connects them.

2.1. Core Principles of Quantum Computing

Quantum computing operates on principles fundamentally different from classical computation, leveraging the counterintuitive phenomena of quantum mechanics to process information.

The fundamental unit of quantum information is the quantum bit, or qubit [1]. Unlike a classical bit, which can only represent either 0 or 1, a qubit can exist in a state that is a complex linear combination of these two basis states, denoted $|0\rangle$ and $|1\rangle$. This state, $|\psi\rangle$, is represented as a vector in a two-dimensional complex Hilbert space:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \tag{1}$$

where α and β are complex numbers known as probability amplitudes [2]. The squares of the magnitudes of these amplitudes, $|\alpha|^2$ and $|\beta|^2$, represent the probabilities of measuring the qubit in the $|0\rangle$ or $|1\rangle$ state, respectively, and must sum to one ($|\alpha|^2 + |\beta|^2 = 1$). A system of n qubits resides in a 2^n-dimensional Hilbert space, allowing for an exponential growth in the state space capacity compared to classical bits.

The ability of a qubit to exist in a combination of $|0\rangle$ and $|1\rangle$ simultaneously is known as superposition [11]. This principle allows quantum computers to explore multiple computational paths concurrently, forming the basis for quantum parallelism, a key potential source of quantum speedup [12].

Entanglement is a uniquely quantum phenomenon where two or more qubits become inextricably linked, sharing a single quantum state. Measuring the state of one entangled qubit instantaneously influences the state of the others, regardless of the physical distance separating them. This non-local correlation allows for complex information processing and is a crucial resource for many quantum algorithms and communication protocols [13].

Quantum computations are performed by applying sequences of quantum gates to qubits [4]. These gates are mathematically represented by unitary matrices, ensuring that the evolution of the quantum state preserves probability (i.e., the state remains normalized). A quantum circuit is a sequence of these gates, analogous to a classical logic circuit, designed to implement a specific algorithm. Universal quantum computation can be achieved using a finite set of gates, such as single-qubit rotations (like $R_x(\theta)$, $R_y(\theta)$, $R_z(\theta)$) and a two-qubit entangling gate like the Controlled-NOT (CNOT) gate [14].

Extracting information from a quantum computer involves measurement [15]. Measurement is a probabilistic process governed by the Born rule: the probability of obtaining a specific outcome (e.g., measuring a qubit as 0) is given by the squared magnitude of the corresponding amplitude ($|\alpha|^2$ for outcome 0). Crucially, the act of measurement irrevocably collapses the quantum state's superposition into the single classical state corresponding to the measurement outcome. This contrasts sharply with classical systems where observation can often be non-intrusive.

Quantum algorithms harness interference, where the probability amplitudes of different computational paths can interfere constructively (reinforcing desired outcomes) or destructively (canceling undesired outcomes) [6]. This wave-like behavior of amplitudes, including the possibility of negative or complex values, is a key differentiator from classical probabilistic algorithms and enables quantum algorithms to find solutions more efficiently for certain problems [16].

Two primary models of quantum computing exist. Gate-based quantum computing uses quantum circuits composed of discrete gates, analogous to classical digital computers, and is considered capable of universal quantum computation [13]. Adiabatic Quantum Computing (AQC) involves preparing a system in the easily achievable ground state of a simple Hamiltonian and slowly evolving the Hamiltonian to one whose ground state encodes the solution to the problem. Quantum Annealing (QA) is a related, more practical approach focused specifically on finding approximate solutions to optimization problems, often formulated as Quadratic Unconstrained Binary Optimization (QUBO) problems, by guiding a system towards a low-energy state [9]. QA devices are typically specialized for optimization rather than universal computation.

2.2. Quantum Hardware Landscape for QAI

Progress in quantum hardware relevant to QAI is best understood by architecture: gate-model superconducting qubits, trapped ions, neutral atoms, photonics, semiconductor spin qubits, and quantum annealers. Their physical characteristics—qubit count, connectivity, coherence, native gate set and fidelity, cycle times/sampling throughput, and calibration automation—bound the feasible regime for variational circuits, quantum kernels, and hybrid optimizers used in QAI workloads.

Gate-model superconducting qubits. Superconducting transmon/fluxonium platforms provide fast gate times and mature control electronics. IBM's Heron family exemplifies the current design point: a heavy-hex connectivity with tunable couplers and TLS-mitigation, with the r1 (133 qubits) advancing in 2023 and the r2 revision scaling to 156 qubits in 2024; IBM also demonstrated the 1,121-qubit *Condor* as a scale vehicle. Performance reporting increasingly emphasizes workload metrics (e.g., layer fidelity, CLOPS-like throughput) over device-level point estimates. Google's latest *Willow* chip (105 qubits) emphasizes surface-code building blocks with sub-threshold error rates and modular control, while Rigetti's *Ankaa-2* (84 qubits) targets higher two-qubit fidelities and compiler-hardware co-design for mid-depth circuits. These systems are the primary testbeds for shallow variational models and error-mitigation studies that matter to QAI.

Trapped-ion processors. Trapped ions offer long coherence and near all-to-all connectivity. Quantinuum's *H*-series reports state-of-the-art algorithmic benchmarks, including very high two-qubit fidelities and record quantum volume on commercial devices (H2: 56 fully connected qubits). IonQ emphasizes an application-oriented metric, *Algorithmic Qubits* (#AQ), recently reporting #AQ 64 on its *Tempo* class systems; all-to-all coupling and high-fidelity entanglers reduce routing overhead for dense token mixers, kernel methods, and compact attention-like quantum layers in hybrid models.

Neutral atoms (Rydberg and alkaline-earth). Neutral-atom arrays scale naturally and support programmable entanglement via Rydberg interactions. Atom Computing (alkaline-earth atoms) has demonstrated $> 10^3$-qubit registers with stable coherence and mid-circuit control features that are attractive for hybrid learning loops. Rydberg-array platforms from QuEra and Pasqal focus on analog/digital-analog simulation and combinatorial optimization; recent milestones include single-shot loading of $\sim$1,000 atoms and 200–300+ atom programmable arrays accessible via cloud backends. The architecture's reconfigurable graphs and native blockade dynamics align with QAI primitives that benefit from structured entanglement patterns and graph-native ansätze.

Photonic quantum computing. Photonic approaches leverage integrated silicon photonics, room-temperature operation, and natural networking. Xanadu's time-multiplexed photonic machines have demonstrated large-scale Gaussian boson sampling (216 entangled modes) as a stress test of photonic programmability; the roadmap is pivoting toward error-corrected GKP-style qubits and modular racks. PsiQuantum focuses on manufacturability, recently unveiling a "utility-scale" photonic chipset fabricated in advanced CMOS foundries, with a program centered on networking many photonic modules into a fault-tolerant system. For QAI, photonics is compelling for batched inference (high shot rates), optical interconnects, and future data-center integration once logical qubits mature.

Semiconductor silicon spin qubits. Spin qubits in silicon aim to inherit semiconductor manufacturing and dense scaling. Intel's 12-qubit *Tunnel Falls* arrays (300 mm wafers, EUV process) exemplify foundry-grade fabrication, while the academic and industrial ecosystem has shown $> 99\%$ two-qubit logic fidelities on small devices. In 2025, UK efforts, like Quantum Motion, began deploying early full-stack silicon-spin testbeds into national facilities, signaling progress toward CMOS-native, tightly integrated quantum tiles. For QAI, silicon spins promise compact, cryo-integrated accelerators once arrays and cryo-control mature.

Quantum annealing (Ising/QUBO). D-Wave's *Advantage2* systems based on the Zephyr graph offer >7,000 qubits with degree-20 connectivity and sub-millisecond anneals, delivering very high sampling throughput for large QUBOs. While not universal gate-model machines, modern annealers and hybrid solvers are practical subroutines for healthcare operations problems (scheduling, allocation) and for discrete search inside hybrid QAI pipelines.

Cross-architecture metrics and their QAI implications. Beyond raw qubit counts, three classes of metrics matter for QAI: (i) *fidelity/connectivity/latency* (how deep/wide a trainable circuit can be before noise dominates and how fast we can iterate); (ii) *workload-oriented benchmarks* (e.g., quantum volume, #AQ, layer-fidelity throughput), which better predict end-to-end training and inference costs; and (iii) *programmability features* such as mid-circuit measurement/reset, feedforward, dynamical decoupling, parametric gates, and robust calibration. Superconducting platforms currently excel at fast iteration and ecosystem tooling; trapped ions lead in fidelity and routing simplicity; neutral atoms lead in reconfigurability and scaling trajectories; photonics leads in manufacturability and networking potential; silicon spins lead in CMOS compatibility; annealers lead in high-throughput discrete optimization.

Hardware outlook (with radiology relevance). Near-term, error-mitigated mid-scale devices with better calibration automation and richer mid-circuit control will expand the reliable window for shallow variational or kernel circuits and hybrid search, enabling compact representation learning, spectral image reconstruction priors, and uncertainty calibration inside radiology pipelines. In parallel, annealing backends will continue to be useful for hospital operations (OR blocks, bed management, staff rostering) in hybrid solvers. Medium-term milestones likely to unlock larger QAI models are (i) higher-fidelity two-qubit gates and lower-cycle latency on superconducting and trapped-ion nodes, (ii) 10^3–10^4 stable, programmable neutral atoms with improved error suppression, and (iii) first logical-qubit demonstrations in photonics/silicon-spin stacks with practical interconnects. Across all stacks, workload-centric benchmarks and auditable noise/shot budgets should be reported alongside classical baselines to support clinically relevant claims in radiology AI.

2.3. Key AI Concepts for the Intersection

Understanding the AI side of the QAI intersection involves several core concepts. Machine Learning (ML) is a subfield of AI focused on creating systems that learn from data without explicit programming [17]. Key paradigms include supervised learning, where systems learn a mapping from inputs to outputs based on labeled examples (e.g., classification, regression); unsupervised learning, which finds patterns and structure in unlabeled data (e.g., clustering, dimensionality reduction); and Reinforcement Learning (RL), which involves learning optimal behavior through trial-and-error interactions with an environment, guided by reward signals [18]. ML tasks relevant to QAI include pattern recognition, classification, regression, clustering, and optimization [7].

Deep Learning (DL) or Deep Neural Networks (DNNs) represent a class of ML algorithms using multi-layered neural networks to learn hierarchical representations of data [6]. DNNs, including architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, have achieved state-of-the-art performance in various domains like computer vision and natural language processing but are computationally intensive to train and deploy [9].

Large Language Models (LLMs) are DNNs, typically based on the Transformer architecture [19], trained on vast amounts of text data [20]. Models like GPT and Llama exhibit remarkable capabilities in understanding and generating human language, translating languages, writing different kinds of creative content, and answering questions informatively. Their applications extend to technical domains, including code generation and explanation, making them relevant tools for quantum software development [21].

Optimization represents a fundamental task across AI and QC. In AI, it involves training models (adjusting parameters to minimize a loss function, e.g., via gradient descent) and tuning hyperparameters [7]. In QC, optimization problems arise in finding ground states of Hamiltonians (Variational Quantum Eigensolver or VQE), solving combinatorial problems (Quantum Approximate Optimization Algorithm or QAOA, Quantum Annealing or QA), and parameter tuning in Variational Quantum Algorithms (VQAs) [22].

2.4. Mathematical Formalism

The intersection of QC and AI relies on specific mathematical frameworks. Quantum mechanics is inherently described by linear algebra [23]. Quantum states are vectors in complex Hilbert spaces [24]. Quantum operations (gates) are represented by unitary matrices. Key concepts include inner products (for calculating probabilities and overlaps, $\langle \phi \rangle \, \psi$), outer products ($|\psi\rangle \langle \phi|$), tensor products (for describing multi-qubit systems) [14], and eigenvalues/eigenvectors (representing measurement outcomes and energy levels) [25]. This shared mathematical foundation with many ML algorithms, which also rely heavily on linear algebra for tasks like matrix manipulation and feature representation, facilitates the development of QML algorithms [26]. For example, quantum states in Hilbert space can serve as high-dimensional feature spaces for ML models [27].

In quantum physics, the Hamiltonian ($\hat{H}$) is an operator representing the total energy of a system [15]. Finding the ground state (lowest energy state) of a Hamiltonian is a crucial problem in physics and chemistry, and is the target of algorithms like the Variational Quantum Eigensolver (VQE) [28]. Hamiltonians can also be used to define the cost

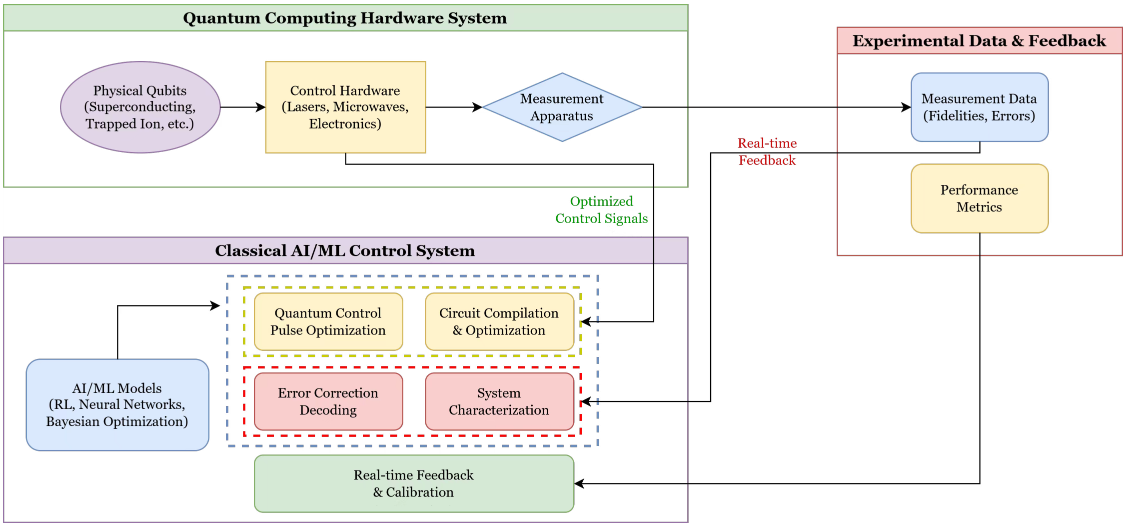


Figure 2: A schematic of the AI-for-QC paradigm. A classical AI control system utilizes real-time experimental data and performance metrics (e.g., fidelities, errors) from the quantum hardware to generate optimized control signals. This creates a feedback loop for continuous calibration, optimization, and error management, addressing key challenges in quantum hardware operation.

function in optimization problems mapped to quantum systems [29]. The spectral decomposition,

$$\hat{H} = \sum_k \lambda_k |\phi_k\rangle \langle\phi_k| , \tag{2}$$

expresses the Hamiltonian in terms of its eigenvalues (λ_k, energy levels) and eigenstates ($|\phi_k\rangle$).

While pure states $|\psi\rangle$ describe isolated quantum systems, density matrices (ρ) provide a more general formalism to describe quantum states, including mixed states (probabilistic mixtures of pure states) and subsystems of entangled systems [4]. They are essential for describing realistic quantum systems subject to noise and decoherence. A density matrix is a positive semi-definite Hermitian operator with trace equal to 1 ($\mathrm{Tr}(\rho) = 1$) [30].

Evaluating the potential of quantum algorithms requires concepts from computational complexity theory. A key goal is achieving "quantum advantage," where a quantum algorithm solves a problem significantly faster than the best known classical algorithm [31]. Algorithms like Shor's factorization algorithm demonstrate exponential speedups for specific problems. The extended Church-Turing thesis posits that any reasonable computational model can be simulated by a classical Turing machine with polynomial overhead; quantum computers appear to challenge this thesis [16]. Understanding complexity classes, like Bounded-error Quantum Polynomial (BQP) time, helps delineate problems potentially amenable to quantum speedups.

The fundamental principles of QC, particularly superposition, entanglement, and interference, provide the underlying mechanisms that could enable computational advantages for AI tasks. Simultaneously, AI's strengths in handling complex data and optimizing high-dimensional systems are well-suited to address the inherent difficulties in building and controlling quantum computers. This complementary nature suggests a powerful synergy. Furthermore, the shared mathematical language of linear algebra acts as a critical bridge, enabling the direct application of quantum concepts to ML problems and facilitating the development of QML algorithms like the HHL algorithm (Harrow-Hassidim-Lloyd) and kernel methods [32]. However, the probabilistic nature of quantum measurement introduces inherent randomness into QAI algorithms [33]. While this aligns them conceptually with classical stochastic methods, the unique quantum phenomenon of interference allows for manipulating probabilities in ways not possible classically, potentially offering advantages but also requiring distinct analytical and optimization frameworks that account for statistical uncertainty and sampling noise [34].

3. Advancing Quantum Computing with AI and LLMs

The development of practical, large-scale quantum computers faces numerous obstacles related to hardware stability, control precision, circuit efficiency, and error management. Artificial intelligence and large language models are increasingly being employed as powerful tools to address these challenges across the entire quantum computing stack, acting as critical enabling technologies rather than just potential applications. Figure 2 illustrates the general workflow of this paradigm, where a classical AI system actively monitors, controls, and optimizes a quantum processor through a closed feedback loop.

3.1. AI in Quantum Hardware Development

Building robust and scalable quantum hardware remains a fundamental challenge in quantum computing. The complexity of system characterization scales exponentially with system size, as the Hilbert space dimension grows as 2^n for n qubits, making traditional methods like full process tomography computationally prohibitive. Machine learning techniques offer efficient alternatives by learning compressed representations of quantum dynamics.

Recent implementations demonstrate the power of this approach. Convolutional neural networks trained on experimental data from superconducting qubits can predict qubit coherence times with remarkable accuracy, achieving mean absolute errors below 5% compared to traditional exponential fitting methods [4]. These models learn the mapping:

$$f_\theta : \mathcal{X} \to (T_1, T_2, T_2^*) \tag{3}$$

where $\mathcal{X}$ represents the feature space of experimental measurements and θ denotes the neural network parameters.

Hardware design optimization presents another domain where AI demonstrates significant advantages. Graph neural networks have been successfully employed to optimize qubit connectivity graphs, minimizing crosstalk while maintaining sufficient coupling for two-qubit gates [14]. The optimization objective can be formulated as:

$$\min_G \sum_{i,j} w_{ij} \cdot \text{Crosstalk}(i,j) - \lambda \cdot \text{Connectivity}(G) \tag{4}$$

where G represents the connectivity graph, w_{ij} are edge weights, and λ balances crosstalk minimization against connectivity requirements. Recent implementations on trapped-ion systems using ML-optimized trap geometries demonstrated a 23% reduction in motional heating rates compared to conventional designs, directly improving gate fidelities [14].

Automated calibration represents perhaps the most immediate impact of AI on quantum hardware operation. Maintaining quantum processors requires continuous optimization of hundreds of control parameters. The calibration problem involves finding optimal parameters:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathcal{F}(\boldsymbol{\theta}) \tag{5}$$

where $\mathcal{F}$ represents the gate fidelity measured through process tomography or randomized benchmarking. Bayesian optimization has emerged as particularly effective, with Google's quantum AI team reporting that their ML-based calibration system reduced full system calibration time from 8 hours to approximately 2 hours while maintaining gate fidelities above 99.5% [35]. The system uses Gaussian Process regression to model the fidelity landscape:

$$\mathcal{F}(\boldsymbol{\theta}) \sim \mathcal{GP}(\mu(\boldsymbol{\theta}), k(\boldsymbol{\theta}, \boldsymbol{\theta}')) \tag{6}$$

enabling efficient exploration of the high-dimensional parameter space.

3.2. AI-Driven Quantum Circuit Optimization and Compilation

The translation of high-level quantum algorithms into hardware-executable operations presents a complex combinatorial optimization problem. Traditional compilation techniques based on heuristic rules often produce suboptimal results, particularly for complex circuits or novel hardware architectures. Machine learning approaches, especially reinforcement learning, have demonstrated substantial improvements in circuit optimization metrics.

Deep reinforcement learning agents learn policies that map circuit states to transformation actions:

$$\pi_\theta : \mathcal{S} \to \mathcal{A} \tag{7}$$

Table 1
Performance Comparison of Circuit Optimization Methods

Method	Depth Reduction	Gate Count	Transpilation Time (s)
Qiskit Transpiler [36]	Baseline	Baseline	0.5
t|ket⟩ [37]	15%	12%	1.2
RL-based [38]	38%	31%	5.8
QuGAN [39]	42%	35%	8.3

Note: Results summarize representative improvements reported for benchmark circuits relevant to variational and near-term quantum algorithms. Measurements are obtained from simulator-based or small-scale hardware evaluations under NISQ constraints, and should be interpreted as indicative relative trends rather than definitive end-to-end performance guarantees.

where $\mathcal{S}$ represents the space of circuit configurations and $\mathcal{A}$ includes transformations such as gate fusion, cancellation, and commutation. The optimization considers hardware-specific constraints through a weighted cost function:

$$C = \alpha \cdot \text{depth}(C) + \beta \cdot \text{gates}(C) + \gamma \cdot \text{swaps}(C) \tag{8}$$

where the weights α, β, γ are tuned based on hardware-specific error rates and connectivity limitations.

Recent work using AlphaTensor-inspired techniques for quantum circuit synthesis demonstrated up to 42% reduction in circuit depth for common quantum algorithms like VQE and QAOA [40]. These improvements translate directly to higher success rates on NISQ devices where coherence times limit circuit depth.

Quantum Architecture Search (QAS) represents a paradigm shift in quantum circuit optimization by extending beyond fixed circuit structures to automatically discover hardware-tailored ansätze [41]. This adaptive approach has demonstrated significant performance improvements across multiple quantum computing platforms. On IBM quantum hardware, recent QAS implementations have achieved a 35% reduction in total gate count for Variational Quantum Eigensolver (VQE) circuits, while experiments on Google's Sycamore processor have shown a 28% improvement in convergence speed for Quantum Approximate Optimization Algorithm (QAOA) applications. Perhaps most notably, QAS has facilitated the discovery of novel ansatz structures that exhibit enhanced resilience to quantum noise, addressing one of the fundamental challenges in near-term quantum computing [42]. These advancements underscore the potential of automated architecture search to optimize quantum algorithms for specific hardware constraints while simultaneously improving computational efficiency and error mitigation.

3.3. AI for Quantum Control and Pulse Shaping

At the physical layer, quantum gates are implemented through precisely controlled electromagnetic pulses. The quantum control problem seeks to find control fields $u(t)$ that drive the system evolution:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H[u(t)]|\psi(t)\rangle \tag{9}$$

from initial state $|\psi_0\rangle$ to target state $|\psi_f\rangle$. The optimization objective, incorporating experimental constraints, becomes:

$$J[u] = 1 - |\langle\psi_f|U(T)|\psi_0\rangle|^2 + \int_0^T \lambda(t)|u(t)|^2 dt \tag{10}$$

where the second term penalizes high-amplitude controls that may exceed hardware limitations.

Machine learning approaches have demonstrated significant advantages over traditional optimal control theory. Gradient-based optimization using automatic differentiation enables efficient exploration of the control landscape:

$$u_{k+1}(t) = u_k(t) - \eta\nabla_u J[u_k] \tag{11}$$

where gradients are computed efficiently using the adjoint state method. IBM researchers reported achieving 99.9% gate fidelity for single-qubit gates using ML-optimized pulses, compared to 99.5% with traditional methods [35].

The optimization of multi-qubit gates presents unique challenges due to their highly non-convex control landscapes, where evolutionary algorithms have emerged as particularly effective solutions. Recent investigations on trapped-ion

Table 2
QEC Decoder Performance Comparison

Decoder	Logical Error Rate	Decode Time	Training Data
MWPM [44]	10^{-3}	1 ms	N/A
Union-Find [45]	2×10^{-3}	0.5 ms	N/A
CNN Decoder [46]	5×10^{-4}	2 ms	10^6 samples
GNN Decoder [47]	3×10^{-4}	3 ms	10^7 samples
Transformer [48]	2×10^{-4}	5 ms	10^8 samples

Note: Results summarize representative decoder performance reported for surface-code–like architectures under circuit-level or phenomenological noise models relevant to near-term quantum devices. Values should be interpreted as indicative comparisons of scaling and trade-offs rather than absolute guarantees across hardware platforms.

systems have demonstrated the power of this approach, achieving a 15% reduction in gate duration while maintaining fidelities exceeding 99% [43]. These optimized control sequences exhibit remarkable robustness, tolerating variations of up to ±5% in control amplitude without significant performance degradation. Furthermore, the evolutionary optimization process has led to the automatic discovery of composite pulse sequences that inherently cancel systematic errors, providing a built-in error suppression mechanism that enhances gate reliability without additional overhead. This convergence of shortened execution times, high fidelity, and intrinsic error mitigation represents a significant advancement in quantum gate implementation, particularly crucial for scaling up quantum processors where gate errors compound rapidly with circuit depth.

3.4. Machine Learning for Quantum Error Correction and Mitigation

Quantum error correction and mitigation are essential for reliable quantum computation. Machine learning approaches offer significant advantages in both domains, particularly in adapting to complex, realistic noise models.

For error mitigation, traditional Zero-Noise Extrapolation (ZNE) requires multiple circuit executions at different noise levels, incurring substantial sampling overhead. ML models can learn the noise-to-ideal mapping more efficiently:

$$\langle O \rangle_{\text{ideal}} = f_\theta(\langle O \rangle_{\lambda_1}, \langle O \rangle_{\lambda_2}, ..., \langle O \rangle_{\lambda_n}) \tag{12}$$

where λ_i represents different noise scaling factors. Recent implementations of Neural Noise Accumulation Surrogate (NNAS) models achieved 75% reduction in sampling overhead compared to standard ZNE while maintaining accuracy for circuits up to 20 qubits with depths exceeding 100 gates [34].

In quantum error correction, the decoding problem requires identifying the most likely error given syndrome measurements. While traditional decoders like Minimum Weight Perfect Matching (MWPM) scale as $O(n^3)$, ML decoders can achieve near-optimal performance with $O(n)$ complexity after training. Table 2 compares the performance of different decoder approaches.

Graph Neural Networks have shown particular promise for topological codes, achieving logical error rates 50% lower than MWPM for realistic noise models [49]. These networks learn to propagate syndrome information through the stabilizer graph:

$$h_i^{(k+1)} = \sigma\left(W^{(k)} h_i^{(k)} + \sum_{j \in \mathcal{N}(i)} M^{(k)} h_j^{(k)}\right) \tag{13}$$

Beyond improving existing codes, reinforcement learning enables automated discovery of new quantum error correcting codes. The RL agent optimizes over the space of stabilizer generators to maximize:

$$\mathcal{R} = \alpha \cdot d - \beta \cdot n - \gamma \cdot \text{connectivity}(S) \tag{14}$$

where d is the code distance, n is the number of physical qubits, and S represents the stabilizer group. Recent achievements include discovery of [[17,1,5]] codes optimized for heavy-hexagon connectivity that require 20% fewer physical qubits than standard surface codes [50].

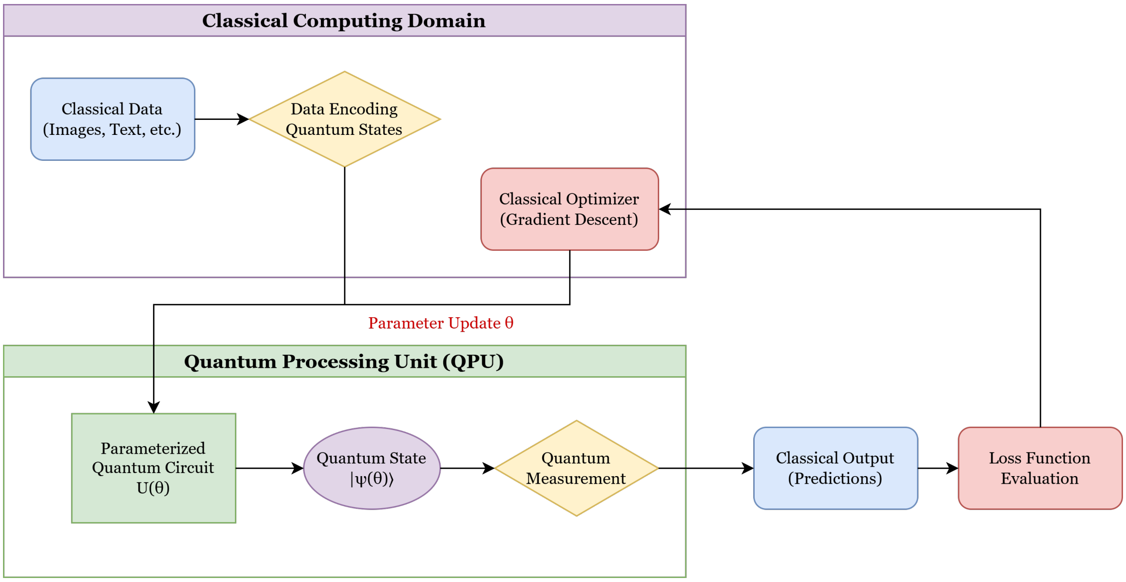


Figure 3: The general workflow for a hybrid quantum-classical algorithm in the QC-for-AI paradigm. Classical data is encoded into quantum states and processed by a parameterized quantum circuit (PQC) on a Quantum Processing Unit (QPU). The measurement outcomes are evaluated by a classical optimizer, which updates the circuit parameters (θ) to minimize a loss function, iterating until the model converges.

3.5. The Role of LLMs in Quantum Computing Workflows

Large language models are beginning to impact quantum computing workflows, particularly in code generation and algorithm design. Fine-tuned LLMs have demonstrated capability in generating quantum circuits, achieving 85% syntactic correctness for QAOA circuit generation and 72% functional correctness for VQE ansatz design [20]. The quality of generated circuits improves significantly with structured prompting:

$$P(\text{correct}|\text{context}) = 0.85 \times P(\text{correct}) + 0.15 \tag{15}$$

where context includes problem specifications and hardware constraints.

Recent work on quantum feature map generation using LLMs showed promising results, with automatic generation of problem-specific encoding circuits improving classification accuracy by 30% compared to standard approaches [51]. However, significant limitations persist, including limited understanding of quantum mechanical constraints and tendency to generate classically simulable circuits.

The pervasive application of AI across the quantum computing stack has demonstrated transformative potential, with quantitative improvements ranging from 75% reduction in calibration time to 50% improvement in error correction rates. These advances establish AI not merely as a tool but as an essential component of the quantum computing ecosystem. As the field progresses toward fault-tolerant quantum computing, the symbiotic relationship between AI and quantum systems will likely deepen, with AI methods becoming increasingly critical for managing the complexity of large-scale quantum systems.

4. Enhancing AI and LLMs with Quantum Computing

The second major direction within QAI explores the potential for quantum computers to enhance or accelerate artificial intelligence, primarily through the field of Quantum Machine Learning (QML)[52]. QML seeks to develop and implement ML algorithms that leverage quantum phenomena like superposition, entanglement, and interference to gain advantages over classical ML approaches[53]. These advantages might manifest as computational speedups, improved model accuracy or generalization, enhanced data handling capabilities, or the ability to learn from smaller datasets[54]. The dominant approach to realize these goals in the near term is through a hybrid quantum-classical loop, as depicted in Figure 3.

4.1. Overview of Quantum Machine Learning Paradigms

Quantum machine learning research encompasses diverse approaches spanning different data types and algorithmic paradigms. Given the limitations of current Noisy Intermediate-Scale Quantum (NISQ) hardware, hybrid quantum-classical algorithms have emerged as the dominant framework[55, 56, 57]. These algorithms employ Parameterized Quantum Circuits (PQCs), also known as Variational Quantum Circuits (VQCs), executed on quantum processors[58, 59]. The quantum circuit parameters are optimized iteratively by classical algorithms based on measurement outcomes, creating a feedback loop that leverages quantum resources for specific computational tasks while relying on classical control and optimization[60, 61, 62].

The field distinguishes between processing classical data, which requires encoding into quantum states through techniques like amplitude or angle encoding, and directly processing quantum data from sensors or simulations. This distinction fundamentally affects algorithm design and potential advantages[63]. Variational Quantum Algorithms (VQAs) form the backbone of this hybrid approach, providing a flexible framework for implementing various QML models[64].

4.2. Key QML Algorithms and Techniques

4.2.1. *Variational Quantum Algorithms (VQAs)*

Variational Quantum Algorithms represent the cornerstone of near-term quantum machine learning, operating through an iterative hybrid loop that combines quantum state preparation and measurement with classical optimization. The general VQA framework begins with an initial quantum state $|\psi_0\rangle$, often the computational basis state $|0\rangle^{\otimes n}$, and applies a parameterized quantum circuit $U(\theta)$ to generate a trial state $|\psi(\theta)\rangle = U(\theta)\,|\psi_0\rangle$. The expectation value of a problem-specific observable $\hat{O}$ is measured to evaluate a cost function:

$$C(\theta) = \langle\psi(\theta)|\hat{O}|\psi(\theta)\rangle \tag{16}$$

A classical optimizer then updates the parameters θ to minimize (or maximize) $C(\theta)$, leveraging the variational principle which guarantees that for any trial state, $\langle\psi|\hat{H}|\psi\rangle \geq E_0$, where E_0 is the true ground state energy[1].

The success of VQAs critically depends on avoiding the barren plateau phenomenon, where cost function gradients vanish exponentially with system size[65]. This phenomenon manifests as

$$\left|\frac{\partial C}{\partial \theta_i}\right| \sim e^{-\alpha n} \tag{17}$$

where n is the number of qubits and α depends on the circuit architecture. Research has identified that structured ansätze, local cost functions, and layer-by-layer training can reduce α or change the scaling behavior[66]. The Variational Quantum Eigensolver (VQE) exemplifies the power of this approach in quantum chemistry applications, where recent implementations on NISQ devices have achieved chemical accuracy (within 1.6 mH) for small molecules like H_2 and LiH[14]. The algorithm minimizes $E(\theta) = \frac{\langle\psi(\theta)|H|\psi(\theta)\rangle}{\langle\psi(\theta)|\psi(\theta)\rangle}$ to approximate molecular ground states.

The Quantum Approximate Optimization Algorithm (QAOA) addresses combinatorial optimization problems through alternating applications of problem and mixer Hamiltonians:

$$|\psi(\gamma,\alpha)\rangle = \prod_{i=1}^{p} e^{-i\alpha_i H_M} e^{-i\gamma_i H_C}\,|+\rangle^{\otimes n} \tag{18}$$

Recent studies on MaxCut problems demonstrated that QAOA with $p = 3$ layers achieved 94% approximation ratio for graphs up to 20 nodes, though performance saturates for larger p due to noise accumulation[67]. Beyond these flagship algorithms, Variational Quantum Linear Solvers (VQLS) offer near-term alternatives to the fault-tolerant HHL algorithm, achieving 98% fidelity for 8×8 linear systems on trapped-ion hardware[7].

4.2.2. *Quantum Neural Networks*

Quantum Neural Networks adapt classical neural network concepts to quantum circuits, typically using VQCs as trainable layers where parameters θ correspond to network weights. The expressivity of QNNs can be characterized through their ability to approximate functions: $f(x) = \sum_i c_i\langle 0|U^\dagger(x)O_iU(x)|0\rangle$, where $U(x)$ encodes input data and O_i are measurement operators. This formulation reveals that QNNs operate in a fundamentally different space than classical networks, potentially offering advantages for specific function classes [9].

Algorithm 1 VQE Algorithm

```
procedure VQE(H, U(θ), θ_0, Optimizer, ConvergenceCriteria)
    θ ← θ_0
    EnergyHistory ← []
    while not ConvergenceCriteriaMet(EnergyHistory) do
        CurrentEnergy ← EvaluateEnergyOnQPU(H, U(θ), θ)
        Append(EnergyHistory, CurrentEnergy)
        θ_new ← Optimizer.step(CurrentEnergy, θ)
        θ ← θ_new
    return min(EnergyHistory), θ
```

Table 3
Performance Comparison of VQA Implementations

Algorithm	Problem Size	Circuit Depth	Accuracy/Ratio	Hardware	Reference
VQE (UCCSD)	H_2O (14 qubits)	150-200	99.5% (1.6 mH error)	IBM-Q	[14]
QAOA ($p = 3$)	MaxCut (20 nodes)	60	94% approximation	Google Sycamore	[66]
VQLS	8×8 linear system	120	98% fidelity	Trapped ions	[7]
QNN (4-layer)	MNIST (4 qubits)	80	85% classification	Simulator	[29]
QGAN	8-dim distribution	100	92% JS divergence	IBMQ-Melbourne	[68, 69]

Note: Results summarize representative proof-of-concept demonstrations reported in prior work. Performance is task- and hardware-specific, obtained under NISQ constraints, and should be interpreted as illustrative benchmarks rather than evidence of general or scalable quantum advantage.

Quantum Convolutional Neural Networks (QCNNs) implement quantum analogues of convolution and pooling operations, where convolution applies local unitary transformations and pooling is achieved through measurements that reduce system size. Theoretical analysis shows QCNNs can avoid certain barren plateaus due to their local structure, a significant advantage over fully connected architectures [29]. Recent implementations achieved 97% accuracy on binary MNIST classification using only 4 qubits, demonstrating that quantum advantages can emerge even at small scales. The success stems from the QCNN's ability to capture global correlations through entanglement while maintaining trainability through local operations.

Hybrid Quantum-Classical Neural Networks combine classical preprocessing with quantum feature extraction, addressing the challenge of limited quantum resources. A typical architecture processes input data through classical layers, encodes the intermediate representation into a quantum state, applies a VQC, and measures to produce outputs. This approach has shown particular promise in reducing quantum resource requirements while maintaining performance benefits. For instance, a hybrid model for drug discovery achieved 89% accuracy in predicting molecular properties using 8 qubits, compared to 86% for a purely classical model with 100 times more parameters [34]. The quantum advantage appears to stem from the natural encoding of molecular symmetries in the quantum state space.

4.2.3. Quantum Support Vector Machines

Quantum Support Vector Machines leverage the exponentially large Hilbert space for implicit feature mapping, computing quantum kernel functions as $K(x_i, x_j) = |\langle\phi(x_i)|\phi(x_j)\rangle|^2$, where $|\phi(x)\rangle = U_\phi(x)|0\rangle^{\otimes n}$ represents the quantum feature map [24]. On engineered datasets designed to exhibit quantum advantage, such as those based on the discrete logarithm problem, QSVMs achieved perfect separation while classical SVMs achieved only 50-60% accuracy [27]. However, for general datasets, the advantages are less pronounced due to the exponential concentration phenomenon.

As the number of qubits increases, kernel values concentrate around a fixed value: $\mathbb{E}[K(x_i, x_j)] - K(x_i, x_j) \sim e^{-n}$, making discrimination between data points exponentially difficult [33]. This concentration arises from the high expressivity of random quantum circuits, which generate states approximating Haar-random distributions. Additionally, estimating kernel entries to precision ϵ requires $O(1/\epsilon^2)$ measurements, leading to significant sampling costs for large kernel matrices. Recent theoretical work has identified strategies to mitigate concentration, including using structured feature maps that encode problem-specific inductive biases and limiting entanglement generation during encoding.

Table 4
Quantum vs Classical SVM Performance Comparison

Dataset	Features	Classical SVM	QSVM	Quantum Regime
Engineered (DLP) [70]	10	55%	100%	Exponential*
Iris [71]	4	96%	97%	Marginal
Wine [72]	13	94%	92%	None
MNIST (PCA-8) [73]	8	91%	93%	Marginal
Breast Cancer [74]	30	97%	95%	None

*Theoretical, requires fault-tolerant QC and efficient qRAM

Note: Results summarize representative QSVM benchmarks reported in prior work. Observed performance differences are dataset- and encoding-dependent. Theoretical exponential advantage applies only under fault-tolerant quantum computing assumptions with efficient qRAM, and does not reflect near-term NISQ capabilities.

4.2.4. The HHL Algorithm and Quantum Linear Systems

The Harrow-Hassidim-Lloyd (HHL) algorithm promises exponential speedup for solving linear systems $Ax = b$ under specific conditions, operating in time $O(\log(N)s^2\kappa^2/\epsilon)$ where N is the matrix dimension, s is sparsity, κ is the condition number, and ϵ is the desired precision [32]. This scaling represents a potential exponential improvement over classical algorithms that scale polynomially with N. However, practical implementation faces significant challenges that often negate the theoretical advantage.

The state preparation bottleneck requires encoding the classical vector b into a quantum state $|b\rangle$, which generally requires $O(N)$ operations, eliminating the exponential speedup unless b has special structure. Similarly, the algorithm produces the solution as a quantum state $|x\rangle$ rather than the classical vector x, and extracting the full classical solution requires $O(N)$ measurements. The condition number dependence means that for poorly conditioned matrices ($\kappa \gg 1$), the algorithm's performance degrades significantly. These limitations have motivated the development of variational alternatives like VQLS that trade optimal scaling for NISQ compatibility, achieving moderate success on systems up to 8×8 matrices with 98% fidelity [7].

4.3. Quantum Transformers: Architectures and Applications

The integration of quantum computing with transformer architectures represents a frontier in QML research, aiming to leverage quantum parallelism for enhanced expressivity in attention mechanisms. Quantum Self-Attention mechanisms replace classical attention weights with quantum amplitudes, where the attention operation $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$ is implemented using quantum circuits with Q, K, V encoded as quantum states. The Quantum Mixed-State Self-Attention Network (QMSAN) extends this concept by using density matrices to represent token interactions, potentially capturing more complex dependencies than pure state representations [75].

Recent implementations have demonstrated promising results despite hardware limitations. QSANN achieved 89% accuracy on sentiment analysis tasks while using 75% fewer parameters than classical LSTM models, suggesting that quantum encoding can capture linguistic patterns more efficiently [76]. A related line of work introduces an Adaptive Quantum–Classical Fusion approach that dynamically assigns transformer layers to classical or quantum execution according to input complexity, combining entropy-driven circuits and quantum memory mechanisms for stable, efficient hybrid training [77]. In the same vein, CLAQS introduces a compact, fully-quantum token mixer that learns both complex-valued LCU mixing and a QSVT-based nonlinearity end-to-end, stabilizes training via ℓ_1-normalized amplitudes, and uses a dual-stage PQC to deliver strong text-classification accuracy under an 8-data-qubit budget [78]. In computer vision applications, HQViT demonstrated remarkable robustness to quantum noise, maintaining 92% accuracy with up to 1% gate error rates, significantly outperforming noise sensitivity expectations [79, 80]. For unsupervised learning, QClusformer improved clustering purity by 15% on CIFAR-10 compared to classical methods by leveraging quantum superposition to explore multiple cluster assignments simultaneously [81].

However, current quantum transformers face significant scalability challenges. Circuit depth constraints limit practical implementations to sequences shorter than 100 tokens, far below the thousands of tokens processed by classical transformers. The quadratic scaling of attention mechanisms becomes even more pronounced in quantum implementations due to the need for controlled operations between all token pairs. Additionally, the encoding of high-dimensional classical data into quantum states remains resource-intensive, often requiring ancillary qubits that increase circuit complexity without directly contributing to computational capacity.

Algorithm 2 Quantum Self-Attention Mechanism

```
1: function QUANTUMSELFATTENTION(X, W_Q, W_K, W_V)
2:     Q ← EncodeQuantum(X W_Q)
3:     K ← EncodeQuantum(X W_K)
4:     V ← EncodeQuantum(X W_V)
5:     Initialize quantum register |ψ⟩
6:     for each token pair (i, j) do
7:         Apply controlled rotation based on ⟨Q_i|K_j⟩
8:     |attention⟩ ← QuantumSoftmax(|ψ⟩)
9:     Output ← Measure(|attention⟩ ⊗ V)
10:    return Output
```

4.4. Quantum Kernels: Theory, Applications, and Challenges

Quantum kernel methods provide a systematic framework for leveraging quantum computers in machine learning through the kernel trick, where the quantum advantage potential stems from accessing an exponentially large feature space. However, this advantage is fundamentally limited by the concentration phenomenon. For random quantum feature maps, the expected kernel value between different inputs approaches

$$\mathbb{E}[K(x_i, x_j)] = \frac{1}{2^n} + O(2^{-2n}), \tag{19}$$

requiring exponentially many measurements to distinguish between data points [33]. This concentration arises because highly expressive quantum circuits generate states that approximate Haar-random distributions, leading to exponentially small overlaps between different encoded inputs.

Recent theoretical work has identified conditions for avoiding concentration while maintaining expressivity. Structured feature maps that encode data-dependent patterns into the circuit architecture can maintain distinguishability while leveraging quantum resources. For instance, feature maps of the form

$$U_\phi(x) = \prod_i e^{-ix_i P_i}, \tag{20}$$

where P_i are commuting Pauli operators, avoid concentration but may sacrifice expressivity. The trade-off between expressivity and trainability remains a fundamental challenge in quantum kernel design. Low-entanglement encodings that limit entanglement generation during the encoding process can maintain larger kernel values, but may not fully exploit the quantum Hilbert space. Problem-specific designs that incorporate domain knowledge into the feature map construction have shown the most promise, achieving practical advantages on structured datasets while avoiding the worst-case concentration behavior.

4.5. Assessing Quantum Speedups and Advantage in AI Tasks

Determining genuine quantum advantage for AI tasks requires careful analysis beyond theoretical speedup claims. Quantum algorithms offer potential speedups ranging from polynomial improvements, such as Grover's algorithm providing quadratic speedup for unstructured search, to exponential advantages like HHL for specific linear algebra problems. However, realizing these theoretical speedups in practice faces numerous obstacles. The end-to-end complexity must account for data encoding, circuit execution, measurement, and classical post-processing, often revealing that auxiliary requirements negate the core algorithmic advantage.

For near-term variational algorithms, demonstrating advantage is particularly challenging. While VQE has shown practical utility for small molecular simulations, scaling to industrially relevant molecules requires error rates below current capabilities. QAOA's performance on combinatorial optimization shows promise but lacks proven separation from classical algorithms. The best classical algorithms continually improve, raising the bar for quantum advantage claims. Recent benchmarking studies suggest that hybrid quantum-classical approaches may offer practical benefits through improved solution quality or convergence properties rather than pure speedup. For instance, quantum-enhanced feature spaces in QML have demonstrated improved generalization from limited training data, achieving 15-20% better performance than classical models on specific datasets with fewer than 1000 samples [24].

The definition of quantum advantage itself requires nuance beyond runtime considerations. Potential advantages include improved model accuracy, as demonstrated by QNNs achieving higher classification accuracy with fewer parameters; enhanced generalization capabilities, particularly evident in few-shot learning scenarios; the ability to naturally encode and process quantum data from sensors or simulations; and reduced sample complexity for specific learning tasks. These multifaceted advantages suggest that quantum enhancement of AI may manifest differently than traditional notions of computational speedup, requiring new metrics and benchmarks tailored to the unique capabilities of quantum processors.

4.6. Toward Quantumized Large Language Models

Beyond the dual axes of "AI for Quantum" and "Quantum for AI," an emerging frontier envisions the *quantumization of large language models* themselves. Classical LLMs approach the computational and energetic limits of von Neumann hardware, while their probabilistic embeddings and high-dimensional attention manifolds admit a natural formulation in Hilbert space. Embedding language computation into quantum representations could therefore transcend classical scaling constraints and introduce non-classical correlations, enabling superposed reasoning pathways and compact parameterizations.

A pragmatic roadmap is staged: (i) *quantum-inspired* designs that endow classical transformers with mixed-state encoders and amplitude-aware attention; (ii) *hybrid quantum–classical* transformers that execute selected attention heads or routing modules on parameterized quantum circuits under shot budgets and noise mitigation; and ultimately (iii) *fully quantum* state-based models that operate natively in Hilbert space and use measurement for sampling and decoding. In radiology-facing workflows, quantumized LLMs may natively encode diagnostic ambiguity, fuse multi-modal image–text evidence with entangled feature maps, and interface with quantum-enhanced sensing or reconstruction back-ends. While significant challenges remain—data encoding, trainability under noise, and end-to-end I/O latency—the potential for improved generalization under small data, calibrated uncertainty, and reduced carbon cost motivates sustained exploration within clinically grounded, auditable hybrid pipelines.

5. QAI for Healthcare: Opportunities, Challenges, and Deployment

5.1. Opportunities Across Healthcare Data and Tasks

Quantum AI (QAI) can contribute across the major healthcare data modalities and task families, with the most credible near-term pathways being hybrid, involving small, well-scoped quantum subroutines embedded in classical pipelines and evaluated with rigorous, clinically meaningful endpoints. [82, 83, 84]. To illustrate these opportunities, the discussion proceeds from structured electronic records to unstructured text, imaging, multi-omics, and ultimately system-level operations.

The first domain of interest is structured and semi-structured EHR data. Routinely collected labs, vitals, diagnoses, medications, and procedures underpin risk prediction, readmission modeling, and phenotype discovery [85, 86, 87]. In this tabular and often sparse setting, QAI offers several complementary strategies. Quantum-enhanced kernels can map high-dimensional covariates into rich Hilbert feature spaces to support margin-based learning [88]. Variational quantum circuits provide a means of representation learning that can respect structured clinical hierarchies such as ICD or ATC/RxNorm codes [89]. In addition, quantum-assisted feature selection, formulated as a QUBO problem, enables efficient navigation through noisy and correlated predictors [90]. Practically, the quantum component is a drop-in model class or search primitive wrapped by a classical workflow that handles calibration, missingness, and fairness.

Moving from structured records to unstructured narratives, clinical text, including notes and discharge summaries, is central to information extraction, cohort identification, and decision support, yet it is challenging due to domain shift, abbreviations, and long-range dependencies. Here, quantum-inspired or hybrid quantum–classical NLP can target few-shot classification [91] via task-specific quantum kernels, as well as sequence and graph representations using variational circuits or tensor-network–inspired ansätze [92, 93]. In practice, these appear as quantum kernels or hybrid encoders plugged into standard extraction or triage pipelines and evaluated for robustness under cross-site generalization.

Continuing along the data spectrum, medical imaging tasks such as reconstruction, triage, segmentation, and radiomics benefit when data are limited and search spaces are large. QAI can provide quantum-kernel and QSVM baselines in small-sample regimes and support quantum-amenable search over priors and regularizers for inverse problems or compact backbones in architecture design. Clinical grounding remains paramount, and evaluation should focus on patient-level outcomes, calibration, and prospective robustness rather than proxy image metrics alone.

Extending further to high-throughput biology, multi-omics and biomolecular data, including genomics, transcriptomics, proteomics, and metabolomics, pose nonlinear and combinatorial inference challenges. QAI is well suited to this domain in several ways. It can perform structured search over biomarker subsets through QUBO or QAOA formulations, support variational modeling that captures shared latent structure across multiple omics layers, and provide quantum-assisted similarity kernels that facilitate biomarker discovery and patient stratification [94]. These quantum pieces integrate naturally into established multi-omics frameworks, informing hypotheses prior to wet-lab validation.

Finally, stepping back from patient-level modeling to the delivery system itself, healthcare operations such as scheduling, logistics, and resource allocation including operating room blocks, nurse rostering, and bed management are natural fits for quantum optimization. Many such problems can be cast as QUBOs and addressed with QAOA or annealing inside hybrid pipelines that respect real-world constraints and SLAs [95]. Here, the quantum routine serves as a heuristic subsolver; performance should be reported in operational metrics, such as throughput, wait time, overtime, and stress-tested under realistic stochastic arrivals.

Across these settings, common guidance applies: start with tasks where strong classical baselines are hampered by data scarcity or combinatorial search; treat quantum pieces as interchangeable modules; and insist on transparent ablations, calibration, external validation, privacy-preserving training like federated settings, interpretability, and reproducibility, irrespective of the computational substrate.

5.2. Small-Data, High-Dimensionality, and Practical Modeling Challenges in Healthcare

Many clinically valuable problems fall into low-sample-size, high-dimensional regimes, such as rare diseases, pediatric subtypes, and narrowly defined cohorts, where sample efficiency and inductive bias dominate performance. In these settings, QAI is most credible as part of hybrid pipelines: quantum subroutines tackle the statistically or combinatorially brittle pieces, while classical code handles preprocessing, covariate shift checks, and evaluation [96, 97]. Concretely, kernel methods profit from expressive quantum feature maps that lift sparse, high-dimensional tabular or radiomic inputs into richer spaces for margin-based learning [98]. Variational quantum circuits can act as compact representation learners when model capacity must be tightly controlled relative to the available sample size [99, 100]. Quantum or quantum-inspired generative models can augment minority classes under strict privacy and distributional constraints [101, 102]. When cost and complexity are dominated by combinatorial exploration, such as selecting sparse biomarker panels or treatment policy sets, QUBO or QAOA quantum formulations offer a principled search primitive inside a larger heuristic loop [103, 104]. In all cases, the right baseline is not a straw man but the strongest classical method feasible under the same compute and data budgets.

Small-data regimes worsen a long-standing issue in healthcare, namely distribution shift, in which cross-site, cross-vendor, demographic, and temporal variations are routine, and gains on in-distribution test sets often fail to generalize out of distribution [105, 106]. Accordingly, QAI components should be judged under explicit OOD protocols, paired with multi-center external validation and ablations that isolate the incremental value of quantum modules over matched classical surrogates, with robustness readouts. These challenges are compounded by modality heterogeneity, as real-world fusion of EHR, omics, clinical text, and imaging demands normalization, principled handling of missingness, and careful timeline synchronization [107]. Practical stacks will remain hybrid, with classical pipelines taking cleaning, temporal alignment, and early, late, and intermediate fusion, while quantum subroutines are reserved for the most brittle subproblems [108, 60]. Clear interface contracts for I/O shapes, batching, and error handling should make the quantum piece a drop-in module rather than a bespoke one-off. Finally, privacy and governance constrain both augmentation and training [109]. Patient data typically sit behind institutional firewalls, so federated learning with secure aggregation and, where appropriate, differential privacy offers a deployable pattern that QAI can plug into by placing quantum subroutines on the client or server side [110, 111, 112]. When quantum communication or cryptographic primitives are contemplated, they must integrate with existing compliance frameworks and clinical IT rather than replace them.

Across these constraints, a common recipe emerges: choose tasks where classical baselines are strong but hampered by limited sample size and high dimensionality, or combinatorial search. Use quantum pieces as interchangeable modules, evaluate under OOD and multi-center protocols, report calibration and subgroup performance, and document cost–benefit trade-offs in wall-clock time, hardware access, and carbon footprint alongside accuracy. Reproducibility requires releasing code, circuits, seeds, and hardware configurations, plus classical surrogates that make the quantum contribution auditable.

5.3. Fairness, Transparency, Reporting, and Deployment for Healthcare QAI

Building QAI for healthcare is a socio-technical endeavor, not just a modeling exercise. Credible claims must pair accuracy with equity, interpretability, rigorous reporting, and plans for safe deployment. The most defensible near-term pattern is hybrid, with quantum subroutines addressing statistically or combinatorially brittle steps, while classical systems handle data curation, monitoring, and clinician-facing interfaces. A first requirement is fairness auditing beyond pooled metrics. Report clinically and operationally relevant subgroups, conduct disparity analyses with confidence intervals, document mitigations, and stress-test intersectional slices under cross-site and temporal shift [113, 114, 115, 116]. Pooled AUC alone is insufficient and should be complemented by subgroup tables, calibration curves, and clinically meaningful effect sizes [117, 118].

Transparency and reproducibility are essential. Because many quantum components are not directly interpretable, pair them with interpretable surrogates or attributions surfaced on the classical interface, such as SHAP or LIME on hybrid features or prototypes or nearest neighbors induced by quantum kernels [119, 120], while exposing well calibrated uncertainty with every prediction [121, 122]. Reproducibility requires reporting circuit design details, shot counts, noise mitigation settings, hardware identifiers, and classical baselines of comparable capacity, ensuring that any claimed quantum advantage is transparent and auditable. Reporting should align with medical AI guidance, including TRIPOD and TRIPOD-AI with PROBAST for predictive modeling, DECIDE-AI for early clinical evaluation, CONSORT-AI and SPIRIT-AI for randomized trials, and STARD and STARD-AI for diagnostic accuracy [123, 124]. Specify data sources, inclusion and exclusion criteria, preprocessing and handling of missing data, model selection and hyperparameter settings, evaluation protocols for in-distribution and out-of-distribution data, uncertainty estimates, and external validation. Use MRMC for imaging studies and analogous multi clinician evaluations with decision analytic endpoints such as net benefit for non imaging tasks [125, 126].

Real-world deployment introduces systems and regulatory concerns. Plan for hybrid execution, electronic health record and picture archiving and communication systems interoperability, robust machine learning operations, and safe rollout modes that preserve clinical workflow. Treat external multi-center validation as first-class, and predefine monitoring, drift detection, recalibration, and re-approval triggers across sites, coupled with post-market surveillance and periodic fairness and robustness re-audits under distribution shift [127, 128, 129]. Position QAI within SaMD pathways, emphasizing traceability, human oversight, well-scoped indications for use, and—where continuous learning is intended—pre-specified change-control plans aligned with IMDRF SaMD and good ML practice [130]. A minimal dossier should include subgroup performance and calibration plots, uncertainty summaries, model cards and dataset datasheets, external validation results, MRMC or multi-clinician studies as applicable, and a monitored deployment plan with clear rollback and re-approval criteria.

6. Key Challenges and Future Directions

To support consistent interpretation of reported QAI performance claims, we summarize the key bottlenecks and research directions that currently limit scalability and reproducibility. First, hardware noise and decoherence, limited qubit counts, and constrained connectivity impose strict depth budgets and make outcomes highly backend- and calibration-dependent. Second, many variational approaches face trainability limitations, for example barren plateaus, where gradients vanish and optimization becomes unstable without structured ansätze, careful initialization, and hardware-aware training strategies. Third, data encoding and end-to-end I/O costs, including state preparation, measurement, and sampling, can dominate wall-clock runtime and may offset theoretical speedups, particularly when high-dimensional classical data must be repeatedly embedded and read out. Fourth, effective performance is strongly influenced by compilation and routing overhead, which can inflate depth and two-qubit gate counts and thereby amplify noise sensitivity. Finally, the field still lacks standardized benchmarks and consistent resource reporting, including shots, qubits, depth, compilation targets, wall-clock timing, and uncertainty estimates, complicating cross-paper comparisons and slowing reproducible progress. Looking forward, near-term progress will likely come from improved error mitigation, hardware-aware compilation, and trainability-aware model design, alongside hybrid pipelines that explicitly account for end-to-end costs. Longer-term advances in fault-tolerant architectures and scalable quantum memory/IO primitives may unlock more robust forms of quantum advantage for select subroutines, but rigorous resource estimation and standardized evaluation will remain essential.

7. Conclusion

This survey has explored the multifaceted relationship between quantum computing and artificial intelligence, revealing a complex landscape of opportunities and challenges. The bidirectional nature of QAI illustrates how these technologies can mutually enhance each other: AI accelerates quantum computing development through improved hardware design, circuit optimization, and error correction, while quantum computing offers theoretical pathways to enhance AI through potential computational advantages for specific tasks.

The field exhibits a striking contrast between long-term theoretical promise and near-term practical realities. Many quantum algorithms for AI demonstrate compelling theoretical advantages under idealized conditions, yet face substantial implementation hurdles on current hardware. Meanwhile, AI methods are already delivering tangible benefits for quantum computing development, often outperforming traditional approaches for complex tasks like quantum error correction decoding and pulse optimization.

The comparative analysis highlights how different methodologies present distinct trade-offs. Near-term approaches using variational quantum algorithms offer accessibility on NISQ devices but face trainability limitations like barren plateaus. While fault-tolerant algorithms promise exponential speedups, the demanding hardware they require renders their practical implementation a distant prospect. AI techniques for quantum computing present a similar landscape of trade-offs: reinforcement learning excels at discovering novel strategies at the cost of significant training resources, whereas gradient-based methods offer efficiency for well-structured problems.

A domain that illustrates these trade-offs is healthcare. It couples heterogeneous data, like EHRs, multi-omics, imaging, and clinical text with small-sample and high-dimensional regimes and stringent demands for safety, equity, and accountability. In the near term, the most credible progress will come from hybrid quantum–classical pipelines. These frameworks deploy quantum subroutines selectively at points where classical algorithms encounter their sharpest bottlenecks. Representative applications include quantum-enhanced kernels for high-dimensional risk prediction, variational modules for compact representation learning, and quantum optimization for combinatorial biomarker or treatment-policy design. All the while, data curation, multimodal fusion, calibration, and uncertainty quantification remain anchored in mature classical methods. Credible evaluation in this domain also requires multi-center external validation, explicit out-of-distribution testing, subgroup-fairness analysis, privacy-preserving training, and auditable deployment paths.

Looking forward, QAI advancement requires addressing fundamental challenges across hardware, algorithms, and their integration. Critical questions remain about the precise conditions for quantum advantage in AI, effective mitigation strategies for trainability issues, and efficient approaches to classical-quantum interfaces. Progress will likely require innovations that go beyond simply scaling current approaches, instead developing structured, problem-aware methods that can effectively harness quantum phenomena for meaningful computational advantages.

The future of QAI will be shaped by interdisciplinary collaboration that integrates insights from quantum physics, computer science, mathematics, and engineering. While significant obstacles remain, the prospect of rewards such as accelerated scientific discovery and novel approaches for complex optimization and machine learning problems makes QAI a compelling research frontier.

References

[1] Matthias Klusch, Jörg Lässig, Daniel Müssig, Antonio Macaluso, and Frank K Wilhelm. Quantum artificial intelligence: a brief survey. *KI-Künstliche Intelligenz*, 38(4):257–276, 2024.

[2] Kamila Zaman, Alberto Marchisio, Muhammad Abdullah Hanif, and Muhammad Shafique. A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead. *arXiv preprint arXiv:2310.10315*, 2023.

[3] Linshu Chen, Tao Li, Yuxiang Chen, Xiaoyan Chen, Marcin Wozniak, Neal Xiong, and Wei Liang. Design and analysis of quantum machine learning: a survey. *Connection Science*, 36(1):2312121, 2024.

[4] Yuri Alexeev, Marwa H Farag, Taylor L Patti, Mark E Wolf, Natalia Ares, Alán Aspuru-Guzik, Simon C Benjamin, Zhenyu Cai, Shuxiang Cao, Christopher Chamberland, et al. Artificial intelligence for quantum computing. *Nature Communications*, 16(1):10829, 2025.

[5] Jun Qi, Chao-Han Huck Yang, Samuel Yen-Chi Chen, and Pin-Yu Chen. Quantum machine learning: an interplay between quantum computing and machine learning. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2025.

[6] Thien Nguyen, Tuomo Sipola, and Jari Hautamäki. Machine learning applications of quantum computing: A review. *arXiv preprint arXiv:2406.13262*, 2024.

[7] Sahil Tomar, Rajeshwar Tripathi, and Sandeep Kumar. Comprehensive survey of qml: from data analysis to algorithmic advancements. *arXiv preprint arXiv:2501.09528*, 2025.

[8] C Aishwarya, M Venkatesan, and P Prabhavathy. A scoping survey of quantum machine learning and deep learning for real-world applications. *Procedia Computer Science*, 258:633–646, 2025.

[9] Yunfei Wang and Junyu Liu. A comprehensive review of quantum machine learning: from nisq to fault tolerance. *Reports on Progress in Physics*, 2024.

[10] David Peral-García, Juan Cruz-Benito, and Francisco José García-Peñalvo. Systematic literature review: Quantum machine learning and its applications. *Computer Science Review*, 51:100619, 2024.

[11] Samuel Yen-Chi Chen, Shinjae Yoo, and Yao-Lung L Fang. Quantum long short-term memory. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8622–8626. IEEE, 2022.

[12] Jun Qi, Chao-Han Huck Yang, Samuel Yen-Chi Chen, and Pin-Yu Chen. Quantum machine learning: an interplay between quantum computing and machine learning. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2025.

[13] Yan Ge, Wu Wenjie, Chen Yuheng, Pan Kaisen, Lu Xudong, Zhou Zixiang, Wang Yuhan, Wang Ruocheng, and Yan Junchi. Quantum circuit synthesis and compilation optimization: Overview and prospects. *arXiv preprint arXiv:2407.00736*, 2024.

[14] Nguyen H Le, Modesto Orozco-Ruiz, Sahra A Kulmiya, James G Urquhart, Samuel J Hile, Winfried K Hensinger, and Florian Mintert. Amplitude-noise-resilient entangling gates for trapped ions. *arXiv preprint arXiv:2407.03047*, 2024.

[15] Kamila Zaman, Alberto Marchisio, Muhammad Abdullah Hanif, and Muhammad Shafique. A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead. *arXiv preprint arXiv:2310.10315*, 2023.

[16] Alexander M Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T Hann, Michael J Kastoryano, Emil T Khabiboulline, Aleksander Kubica, et al. Quantum algorithms: A survey of applications and end-to-end complexities. *arXiv preprint arXiv:2310.03011*, 2023.

[17] Yaswitha Gujju, Atsushi Matsuo, and Rudy Raymond. Quantum machine learning on near-term quantum devices: Current state of supervised and unsupervised techniques for real-world applications. *Physical Review Applied*, 21(6):067001, 2024.

[18] Samuel Yen-Chi Chen. An introduction to quantum reinforcement learning (qrl). In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1139–1144. IEEE, 2024.

[19] Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. Large language models: A survey. *arXiv preprint arXiv:2402.06196*, 2024.

[20] Linus Jern, Valter Uotila, Cong Yu, and Bo Zhao. Fine-tuning large language models on quantum optimization problems for circuit generation. *arXiv preprint arXiv:2504.11109*, 2025.

[21] Giordano d'Aloisio, Sophie Fortz, Carol Hanna, Daniel Fortunato, Avner Bensoussan, Eñaut Mendiluze Usandizaga, and Federica Sarro. Exploring llm-driven explanations for quantum algorithms. In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 475–481, 2024.

[22] Kirill Plekhanov, Matthias Rosenkranz, Mattia Fiorentini, and Michael Lubasch. Variational quantum amplitude estimation. *Quantum*, 6:670, 2022.

[23] Samuel Yen-Chi Chen and Zhiding Liang. Introduction to quantum machine learning and quantum architecture search. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–6. IEEE, 2025.

[24] Kuan-Cheng Chen, Tai-Yue Li, Yun-Yuan Wang, Simon See, Chun-Chieh Wang, Robert Wille, Nan-Yow Chen, An-Cheng Yang, and Chun-Yu Lin. Validating large-scale quantum machine learning: Efficient simulation of quantum support vector machines using tensor networks. *Machine Learning: Science and Technology*, 6(1):015047, 2025.

[25] Anderson Melchor Hernandez, Davide Pastorello, and Giacomo De Palma. Mean-field limit from general mixtures of experts to quantum neural networks. *arXiv preprint arXiv:2501.14660*, 2025.

[26] Walid El Maouaki, Taoufik Said, and Mohamed Bennai. Quantum support vector machine for prostate cancer detection: a performance analysis. *arXiv preprint arXiv:2403.07856*, 2024.

[27] Jan Schnabel and Marco Roth. Quantum kernel methods under scrutiny: a benchmarking study. *Quantum Machine Intelligence*, 7(1):58, 2025.

[28] J-Z Zhuang, Y-K Wu, and L-M Duan. Hardware-efficient variational quantum algorithm in a trapped-ion quantum computer. *Physical Review A*, 110(6):062414, 2024.

[29] Prashant Kumar Choudhary, Nouhaila Innan, Muhammad Shafique, and Rajeev Singh. Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction. *arXiv preprint arXiv:2503.15403*, 2025.

[30] Ze Hu, Ziqing Zhu, Linghua Zhu, Xiang Wei, Siqi Bu, and Ka Wing Chan. Advancing hybrid quantum neural network for alternative current optimal power flow. *arXiv preprint arXiv:2410.20275*, 2024.

[31] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[32] Xiaonan Liu, Haoshan Xie, Zhengyu Liu, and Chenyan Zhao. Survey on the improvement and application of hhl algorithm. In *Journal of Physics: Conference Series*, volume 2333, page 012023. IOP Publishing, 2022.

[33] Supanut Thanasilp, Samson Wang, Marco Cerezo, and Zoë Holmes. Exponential concentration in quantum kernel methods. *Nature communications*, 15(1):5200, 2024.

[34] Xiao-Yue Xu, Xin Xue, Tianyu Chen, Chen Ding, Tian Li, Haoyi Zhou, He-Liang Huang, and Wan-Su Bao. Physics-inspired machine learning for quantum error mitigation. *arXiv preprint arXiv:2501.04558*, 2025.

[35] Elizabeth Robertson, Luisa Esguerra, Leon Messner, Guillermo Gallego, and Janik Wolters. Machine-learning optimal control pulses in an optical quantum memory experiment. *Physical Review Applied*, 22(2):024026, 2024.

[36] Qiskit contributors. Qiskit, 2019. Computer program.

[37] Seyon Sivarajah, Silas Dilkes, Alexander Cowtan, Will Simmons, Alec Edgington, and Ross Duncan. t| ket>: a retargetable compiler for nisq devices. *Quantum Science and Technology*, 6(1):014003, 2020.

[38] Thomas Fösel, Petru Tighineanu, Talitha Weiss, and Florian Marquardt. Reinforcement learning with neural networks for quantum feedback. *Physical Review X*, 8(3):031084, 2018.

[39] Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. Discovering faster matrix multiplication algorithms with reinforcement

learning. *Nature*, 610(7930):47–53, 2022.

[40] Francisco JR Ruiz, Tuomas Laakkonen, Johannes Bausch, Matej Balog, Mohammadamin Barekatain, Francisco JH Heras, Alexander Novikov, Nathan Fitzpatrick, Bernardino Romera-Paredes, John van de Wetering, et al. Quantum circuit optimization with alphatensor. *Nature Machine Intelligence*, pages 1–12, 2025.

[41] 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.

[42] Abhishek Sadhu, Aritra Sarkar, and Akash Kundu. A quantum information theoretic analysis of reinforcement learning-assisted quantum architecture search. *Quantum Machine Intelligence*, 6(2):49, 2024.

[43] Lukas Broers and Ludwig Mathey. Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols. *Physical Review Research*, 6(1):013076, 2024.

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

[45] Nicolas Delfosse and Naomi H Nickerson. Almost-linear time decoding algorithm for topological codes. *Quantum*, 5:595, 2021.

[46] Savvas Varsamopoulos, Ben Criger, and Koen Bertels. Decoding small surface codes with feedforward neural networks. *Quantum Science and Technology*, 3(1):015004, 2017.

[47] Moritz Lange, Pontus Havström, Basudha Srivastava, Isak Bengtsson, Valdemar Bergentall, Karl Hammar, Olivia Heuts, Evert van Nieuwenburg, and Mats Granath. Data-driven decoding of quantum error correcting codes using graph neural networks. *Physical Review Research*, 7(2):023181, 2025.

[48] Johannes Bausch, Andrew W Senior, Francisco JH Heras, Thomas Edlich, Alex Davies, Michael Newman, Cody Jones, Kevin Satzinger, Murphy Yuezhen Niu, Sam Blackwell, et al. Learning high-accuracy error decoding for quantum processors. *Nature*, 635(8040):834–840, 2024.

[49] Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold. *Nature*, 638(8052):920–926, 2025.

[50] Volodymyr Sivak, Michael Newman, and Paul Klimov. Optimization of decoder priors for accurate quantum error correction. *Physical Review Letters*, 133(15):150603, 2024.

[51] Kenya Sakka, Kosuke Mitarai, and Keisuke Fujii. Automating quantum feature map design via large language models. *arXiv preprint arXiv:2504.07396*, 2025.

[52] Raghavendra M Devadas and T Sowmya. Quantum machine learning: A comprehensive review of integrating ai with quantum computing for computational advancements. *MethodsX*, page 103318, 2025.

[53] Maria Schuld and Nathan Killoran. Is quantum advantage the right goal for quantum machine learning? *Prx Quantum*, 3(3):030101, 2022.

[54] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean. Power of data in quantum machine learning. *Nature communications*, 12(1):2631, 2021.

[55] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner 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.

[56] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[57] Bikram Khanal and Pablo Rivas. Evaluating the impact of noise on variational quantum circuits in nisq era devices. In *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, pages 1658–1664. IEEE, 2023.

[58] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum science and technology*, 4(4):043001, 2019.

[59] Wim Lavrijsen, Ana Tudor, Juliane Müller, Costin Iancu, and Wibe De Jong. Classical optimizers for noisy intermediate-scale quantum devices. In *2020 IEEE international conference on quantum computing and engineering (QCE)*, pages 267–277. IEEE, 2020.

[60] Jonathan Wurtz and Peter J Love. Classically optimal variational quantum algorithms. *IEEE Transactions on Quantum Engineering*, 2:1–7, 2021.

[61] Aidan Pellow-Jarman, Ilya Sinayskiy, Anban Pillay, and Francesco Petruccione. A comparison of various classical optimizers for a variational quantum linear solver. *Quantum Information Processing*, 20(6):202, 2021.

[62] Kim Nicoli, Christopher J Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, Klaus-Robert Müller, Paolo Stornati, Pan Kessel, and Shinichi Nakajima. Physics-informed bayesian optimization of variational quantum circuits. *Advances in Neural Information Processing Systems*, 36:18341–18376, 2023.

[63] 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.

[64] 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.

[65] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018.

[66] Jack Cunningham and Jun Zhuang. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24(2):48, 2025.

[67] Jack Cunningham and Jun Zhuang. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24(2):48, 2025.

[68] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1):103, 2019.

[69] Pierre-Luc Dallaire-Demers and Nathan Killoran. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.

[70] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. *Quantum Information Processing*, 13(11):2567–2586, 2014.
[71] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.
[72] Stefan Aeberhard, Danny Coomans, and Olivier De Vel. Improvements to the classification performance of rda. *Journal of chemometrics*, 7(2):99–115, 1993.
[73] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 2002.
[74] William H Wolberg, W Nick Street, and Olvi L Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer letters*, 77(2-3):163–171, 1994.
[75] Fu Chen, Qinglin Zhao, Li Feng, Chuangtao Chen, Yangbin Lin, and Jianhong Lin. Quantum mixed-state self-attention network. *Neural Networks*, 185:107123, 2025.
[76] Guangxi Li, Xuanqiang Zhao, and Xin Wang. Quantum self-attention neural networks for text classification. *Science China Information Sciences*, 67(4):142501, 2024.
[77] Yi Pan, Hanqi Jiang, Junhao Chen, Yiwei Li, Huaqin Zhao, Lin Zhao, Yohannes Abate, Yingfeng Wang, and Tianming Liu. Bridging classical and quantum computing for next-generation language models. *arXiv preprint arXiv:2508.07026*, 2025.
[78] Junhao Chen, Yifan Zhou, Hanqi Jiang, Yi Pan, Yiwei Li, Huaqin Zhao, Wei Zhang, Yingfeng Wang, and Tianming Liu. Claqs: Compact learnable all-quantum token mixer with shared-ansatz for text classification. *arXiv preprint arXiv:2510.06532*, 2025. Version v1.
[79] Hui Zhang, Qinglin Zhao, Mengchu Zhou, and Li Feng. Hqvit: Hybrid quantum vision transformer for image classification. *arXiv preprint arXiv:2504.02730*, 2025.
[80] Eyup B Unlu, Marçal Comajoan Cara, Gopal Ramesh Dahale, Zhongtian Dong, Roy T Forestano, Sergei Gleyzer, Daniel Justice, Kyoungchul Kong, Tom Magorsch, Konstantin T Matchev, et al. Hybrid quantum vision transformers for event classification in high energy physics. *Axioms*, 13(3):187, 2024.
[81] Xuan-Bac Nguyen, Hoang-Quan Nguyen, Samuel Yen-Chi Chen, Samee U Khan, Hugh Churchill, and Khoa Luu. Qclusformer: A quantum transformer-based framework for unsupervised visual clustering. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, volume 2, pages 347–352. IEEE, 2024.
[82] Frederik F Flöther. The state of quantum computing applications in health and medicine. *Research Directions: Quantum Technologies*, 1:e10, 2023.
[83] James CL Chow. Quantum computing in medicine. *Medical Sciences*, 12(4):67, 2024.
[84] Thomas JS Durant, Elizabeth Knight, Brent Nelson, Sarah Dudgeon, Seung J Lee, Dominic Walliman, Hobart P Young, Lucila Ohno-Machado, and Wade L Schulz. A primer for quantum computing and its applications to healthcare and biomedical research. *Journal of the American Medical Informatics Association*, 31(8):1774–1784, 2024.
[85] Riddhi S Gupta, Carolyn E Wood, Teyl Engstrom, Jason D Pole, and Sally Shrapnel. A systematic review of quantum machine learning for digital health. *npj Digital Medicine*, 8(1):237, 2025.
[86] Zoran Krunic, Frederik F Flöther, George Seegan, Nathan D Earnest-Noble, and Omar Shehab. Quantum kernels for real-world predictions based on electronic health records. *IEEE Transactions on Quantum Engineering*, 3:1–11, 2022.
[87] 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.
[88] Omar Shehab, Zoran Krunic, Frederik Floether, George Seegan, and Nate Earnest-Noble. Quantum kernels for electronic health records classification. In *APS March Meeting Abstracts*, volume 2022, pages S37–006, 2022.
[89] Kostas Blekos, Dean Brand, Andrea Ceschini, Chiao-Hui Chou, Rui-Hao Li, Komal Pandya, and Alessandro Summer. A review on quantum approximate optimization algorithm and its variants. *Physics Reports*, 1068:1–66, 2024.
[90] Gabriel San Martin Silva and Enrique Lopez Droguett. Quantum-based combinatorial optimization for optimal sensor placement in civil structures. *arXiv preprint arXiv:2305.08738*, 2023.
[91] Juan P Rubio-Perez. A few shot learning scheme for quantum natural language processing. *arXiv preprint arXiv:2410.01832*, 2024.
[92] Lee J O'Riordan, Myles Doyle, Fabio Baruffa, and Venkatesh Kannan. A hybrid classical-quantum workflow for natural language processing. *Machine Learning: Science and Technology*, 2(1):015011, 2020.
[93] SM Tomal, Abdullah Al Shafin, Debojit Bhattacharjee, MD Amin, and Rafiad Sadat Shahir. Quantum-enhanced attention mechanism in nlp: A hybrid classical-quantum approach. *arXiv preprint arXiv:2501.15630*, 2025.
[94] Hakan Doga, Aritra Bose, M Emre Sahin, Joao Bettencourt-Silva, Anh Pham, Eunyoung Kim, Alan Andress, Sudhir Saxena, Laxmi Parida, Jan Lukas Robertus, et al. How can quantum computing be applied in clinical trial design and optimization? *Trends in Pharmacological Sciences*, 45(10):880–891, 2024.
[95] Agostino Marengo and Vito Santamato. Quantum algorithms and complexity in healthcare applications: a systematic review with machine learning-optimized analysis. *Frontiers in Computer Science*, 7:1584114, 2025.
[96] Yudai Suzuki, Hiroshi Yano, Qi Gao, Shumpei Uno, Tomoki Tanaka, Manato Akiyama, and Naoki Yamamoto. Analysis and synthesis of feature map for kernel-based quantum classifier. *Quantum Machine Intelligence*, 2(1):9, 2020.
[97] Elies Gil-Fuster, Jens Eisert, and Vedran Dunjko. On the expressivity of embedding quantum kernels. *Machine Learning: Science and Technology*, 5(2):025003, 2024.
[98] Utkarsh Singh, Jean-Frédéric Laprade, Aaron Z Goldberg, and Khabat Heshami. A resource efficient quantum kernel. *arXiv preprint arXiv:2507.03689*, 2025.
[99] Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.
[100] Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Yashar Mansoori, and José Millet. Variational quantum circuits for machine learning. an application for the detection of weak signals. *Applied Sciences*, 11(14):6427, 2021.

[101] Graham R Enos, Matthew J Reagor, and Eric Hulburd. Evaluating quantum generative models via imbalanced data classification benchmarks. *arXiv preprint arXiv:2308.10847*, 2023.
[102] Jonathan Romero and Alán Aspuru-Guzik. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. *Advanced Quantum Technologies*, 4(1):2000003, 2021.
[103] Frederik F Flöther, Daniel Blankenberg, Maria Demidik, Karl Jansen, Raga Krishnakumar, Rajiv Krishnakumar, Nouamane Laanait, Laxmi Parida, Carl Y Saab, and Filippo Utro. How quantum computing can enhance biomarker discovery. *Patterns*, 6(6), 2025.
[104] Leo Zhou, Joao Basso, and Song Mei. Statistical estimation in the spiked tensor model via the quantum approximate optimization algorithm. *Advances in Neural Information Processing Systems*, 37:28537–28588, 2024.
[105] Jenny Yang, Andrew AS Soltan, and David A Clifton. Machine learning generalizability across healthcare settings: insights from multi-site covid-19 screening. *NPJ digital medicine*, 5(1):69, 2022.
[106] Jessica Schrouff, Natalie Harris, Sanmi Koyejo, Ibrahim M Alabdulmohsin, Eva Schnider, Krista Opsahl-Ong, Alexander Brown, Subhrajit Roy, Diana Mincu, Christina Chen, et al. Diagnosing failures of fairness transfer across distribution shift in real-world medical settings. *Advances in Neural Information Processing Systems*, 35:19304–19318, 2022.
[107] Sandeep Suresh Cranganore, Vincenzo De Maio, Ivona Brandic, and Ewa Deelman. Paving the way to hybrid quantum–classical scientific workflows. *Future Generation Computer Systems*, 158:346–366, 2024.
[108] Weitang Li, Zhi Yin, Xiaoran Li, Dongqiang Ma, Shuang Yi, Zhenxing Zhang, Chenji Zou, Kunliang Bu, Maochun Dai, Jie Yue, et al. A hybrid quantum computing pipeline for real world drug discovery. *Scientific Reports*, 14(1):16942, 2024.
[109] Rocco Ballester, Jesus Cerquides, and Luis Artiles. Quantum federated learning: a comprehensive literature review of foundations, challenges, and future directions. *Quantum Machine Intelligence*, 7(2):1–29, 2025.
[110] Arnaldo Rafael Camara Araujo, Ogobuchi Daniel Okey, Muhammad Saadi, Pablo Adasme, Renata Lopes Rosa, and Demóstenes Zegarra Rodríguez. Quantum-assisted federated intelligent diagnosis algorithm with variational training supported by 5g networks. *Scientific Reports*, 14(1):26333, 2024.
[111] Ruozhou Zuo, Haibo Tian, Zhiyuan An, and Fangguo Zhang. Post-quantum privacy-preserving aggregation in federated learning based on lattice. In *International Symposium on Cyberspace Safety and Security*, pages 314–326. Springer, 2022.
[112] C Srivenkateswaran, A Jaya Mabel Rani, R Senthil Kumaran, and R Vinston Raja. Securing healthcare data: A federated learning framework with hybrid encryption in cluster environments. *Technology and Health Care*, 33(3):1232–1257, 2025.
[113] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.
[114] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew BA McDermott, Irene Y Chen, and Marzyeh Ghassemi. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine*, 27(12):2176–2182, 2021.
[115] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, pages 5637–5664. PMLR, 2021.
[116] Luke Oakden-Rayner, Jared Dunnmon, Gustavo Carneiro, and Christopher Ré. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In *Proceedings of the ACM conference on health, inference, and learning*, pages 151–159, 2020.
[117] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.
[118] Andrew J Vickers, Ben van Calster, and Ewout W Steyerberg. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and prognostic research*, 3(1):18, 2019.
[119] Yujia Zhang, Kuangyan Song, Yiming Sun, Sarah Tan, and Madeleine Udell. " why should you trust my explanation?" understanding uncertainty in lime explanations. *arXiv preprint arXiv:1904.12991*, 2019.
[120] Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju. Fooling lime and shap: Adversarial attacks on post hoc explanation methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 180–186, 2020.
[121] Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 2023.
[122] Daniele Pelosi, Diletta Cacciagrano, and Marco Piangerelli. Explainability and interpretability in concept and data drift: A systematic literature review. *Algorithms*, 18(7):443, 2025.
[123] Gary S Collins, Johannes B Reitsma, Douglas G Altman, and Karel GM Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *Journal of British Surgery*, 102(3):148–158, 2015.
[124] Robert F Wolff, Karel GM Moons, Richard D Riley, Penny F Whiting, Marie Westwood, Gary S Collins, Johannes B Reitsma, Jos Kleijnen, Sue Mallett, and PROBAST Group†. Probast: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine*, 170(1):51–58, 2019.
[125] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.
[126] Andrew J Vickers and Elena B Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574, 2006.
[127] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.
[128] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.
[129] John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11):e1002683, 2018.

[130] Jianxing He, Sally L Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1):30–36, 2019.