Bridging Quantum Trainability and Autonomous Control: A Comparative Literature Survey of Hybrid Quantum-Classical Optimization in NISQ-Era Machine Learning

Arvindhan Muthusamy[1], Azween Abdullah[2], Daniel Arockiam[3]

[1] Postdoctoral Fellow, Computing and Digital Technology, Help University, Damansara, Kuala Lumpur, Malaysia, arvindhan.m@helplive.edu.my , saroarvindmster@gmail.com, ORCID:305-317-305-317

[2] Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur, Malaysia, azween.a@help.edu.my

[3] Daniel Arockiam, Assistant Professor School of Engineering and IT, MAHE Dubai. daniel.arockiam@dxb.manipal.edu.

HOW TO CITE:

Arvindhan Muthusamy, Azween Abdullah, Daniel Arockiam (2026). Bridging Quantum Trainability and Autonomous Control: A Comparative Literature Survey of Hybrid Quantum-Classical Optimization in NISQ-Era Machine Learning. International Journal of Special Education, 41(20s), 305-317

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors. Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:

The emergence of Noisy Intermediate-Scale Quantum (NISQ) hardware has catalyzed two convergent yet methodologically distinct research frontiers: the application of reinforcement learning (RL) to autonomous quantum control, and the integration of parameterized quantum circuits (PQCs) into classical large language model (LLM) fine-tuning pipelines. The trainability of parameterized quantum components under realistic hardware noise and high-dimensional optimization landscapes impacts the feasibility of deploying quantum algorithms in practical settings, helping readers grasp the real-world significance of these advances. Quantum Machine Learning presents a rigorous, empirically grounded benchmark of three RL paradigms — Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC), and Reinforcement Learning from Demonstration (RLFD) — applied to quantum gate synthesis, state preparation, and qubit readout waveform optimization. Ada-QFT (Adaptive Quantum Fine-Tuning) framework, an architecturally novel hybrid system in which a PQC layer is embedded within a pretrained LLM backbone, and whose initialization is governed by a generative AI-driven, Expected Improvement (EI)-guided iterative search supported by a formal sub martingale convergence guarantee. The theoretical, methodological, and empirical terrains covered by these two publications are unified in this literature review. It analyzes the places where the two texts agree and disagree, finds common assumptions and epistemological conflicts, and lays out the blanks that a possible proposed techniques could fill. It goes on to suggest suitable comparative parameters, quantifiable sub-metrics, and potential quantum algorithms to fortify a cohesive scholarly contribution in both fields.

Keywords: Noisy Intermediate-Scale Quantum, Reinforcement Learning, Adaptive Quantum Fine-Tuning, Optimization in Quantum ML, Autonomous Quantum Control

1. Introduction

At the crossroads of quantum information science and contemporary artificial intelligence lies the theoretically rich but pragmatically perilous domain of quantum machine learning, or QML. The current state of study in this area is driven by two long-standing and largely separate issues: the first is the trainability problem in parameterized quantum circuits, and the second is the efficient and scalable control of physical quantum devices. A comparative examination is both academically fruitful and current because the articles discussed here tackle these problems from different but complimentary angles. First described and extensively explored in [1] is the trainability problem, which is formalized by the barren plateau phenomena. presents a domain where gradient-based optimization becomes nearly impossible since the cost function gradient of a deep PQC disappears exponentially with qubit count or circuit depth. This is not only a technical annoyance; it is a limiting factor for any derivative frameworks that use gradient-based parameter optimization in high-dimensional quantum environments, including the variational quantum eigen solver class of methods. Finding the best time-dependent control pulses or sequences to execute high-fidelity unitary operations on physical qubits that are decoherent, gate-error prone, and crosstalk-prone is the complementary challenge that the control problem addresses. While GRAPE and DRAG are analytically sound, they aren't immune to model bias, which occurs whenever there's a difference between the mathematical model of the system and the actual device used to execute the control sequence (be it PQC rotation angles or gate-synthesis control pulses). When contrasted with static, one-shot, or analytically extracted initialization procedures, their results are superior. This common understanding is the conceptual glue that holds these two works together, allowing for a comparison that is not only feasible but also epistemologically enlightening. Here is how the current survey moves forward. In Section 2, the two papers are placed in the larger theoretical framework of PQC capacity for training and quantum control. Issue formulation, technique, theoretical contributions, experimental rigor, and assessment metrics are some of the important dimensions compared in Section 3. In Section 4, we can see where the two approaches overlap, where they diverge, and where there is constructive friction. The sections that need to be filled out or addressed by a third manuscript are outlined in Section 5. The sixth section provides recommendations for quantification techniques, sub-metrics, and comparison parameters, along with explanations for each. The field's trajectory is summarized in Section 7.

2.THEORETICAL BACKGROUND AND STATE OF THE FIELD

The quantum circuits with continuous gate parameters are called parameterized quantum circuits (PQCs), or variational quantum circuits (VQCs) in the literature. An optimization function that is task-specific is used to determine the best value for these classically tuneable variables. In theory, PQCs are attractive function approximators because, given the right parameters, they can produce probability distributions over measurement outcomes that are classically hard to sample. This means that, for some types of problems, PQCs may provide representational benefits over classical neural networks. Hardware noise and scalability concerns are two of the limits that PQCs are now facing. These issues impact their comparative advantages and the research application of PQCs [2].

2.1 Parameterized Quantum Circuits as Trainable Function Approximators

Within the realm of PQC-based frameworks, the two most extensively researched ones are the variational quantum eigen solution (VQE) and the quantum approximation optimization algorithm (QAOA). But there is still a little but

noticeable hole in their literary reviews. The main focus of both papers is on simple, hardware-efficient circuit designs suitable for the NISQ period, which reflects the limitations of current quantum technology. Highlighting hardware-efficient designs highlights their practical significance for the development of quantum technologies at now.

2.2 The Barren Plateau Problem: Theoretical Foundations

For PQCs that use parameters drawn from the invariant measurement an of the unitary group, such as Haar-random circuits, variation of any local creative function cost function gradient decays exponentially as Var[∂C/∂θ_k] ∘ 2^(-n), where n is the number of qubits [3], according to the formalized barren plateau problem. This problem is cited in both manuscripts. The practical ramifications of this conclusion are catastrophic: as the complexity of the system grows, the available gradient signal to a classical optimizer shrinks rapidly, necessitating an indefinitely substantial number of gradients evaluations just to make little headway. This groundbreaking finding has been further developed and expanded upon in subsequent literature. It is not only for Haar-random circuits that barren plateaus exist; Holmes et al. proved it for all types of PQC designs [4]. established that local cost functions are generically more resistant to barren plateaus than global cost functions, providing practical guidance for circuit design. Wang et al. showed that hardware noise characteristic of NISQ devices induces noise-induced barren plateaus that persist even for shallow circuits, representing the most practically alarming extension of the original result.

2.3 Mitigation Strategies: Prior Art Landscape

The literature has proposed several classes of barren plateau mitigation strategies, each with distinct theoretical underpinnings and practical limitations, which are reviewed in both manuscripts. Initialization-based strategies include the identity block initialization approach, in which circuit parameters are initialized near the identity transformation, and the more recently proposed generative AI-driven initialization. Structural strategies include restricting the circuit topology to shallow or locally-connected architectures, as proposed by Cerezo et al. Layer-wise training strategies, as proposed by Skolik et al., optimize the circuit in stages rather than globally. None of these approaches, as correctly noted, provides a simultaneous guarantee of (i) dynamic adaptivity to the optimization landscape, (ii) formal convergence assurance, and (iii) architectural generality the three properties that the Ada-QFT framework claims to satisfy jointly.

2.4 Reinforcement Learning for Quantum Control: Prior Art

The application of model-free reinforcement learning to quantum control problems has a relatively short but rapidly growing history. Early work showed that deep RL agents could learn to prepare single-qubit states and implement simple gates with high fidelity in simulations. Later studies extended these results to multi-qubit systems, achieving fidelities above 0.99 for two-qubit gates such as the CNOT Pure model-free approaches in high-dimensional control spaces have a sample inefficiency that limits their practical usefulness. Convergence requires tens of thousands of environment interactions, which is an issue that has been highlighted time and time again. [5].

A significant methodological advancement of Quantum RLfD is the merging of model-based demonstrations, including GRAPE-derived control sequences, with model-free RL using the Reinforcement Learning from Demonstration (RLfD) paradigm. This method is compatible with frameworks for imitation learning and inverse reinforcement learning. But it doesn't go into this theoretical link at all, which is a gap in the literature that another manuscript could fill. The fundamental idea behind RLfD in quantum control is that the RL agent can drastically cut down on exploration effort by using demonstrations from an imperfect but informative optimizer based on models. Also, by interacting directly with the quantum environment, the online RL component can fix model bias [6]. There are still many unanswered concerns about the practical benefits of quantum

computing, and the integration of these features into classical natural language processing pipelines is an emerging area of research. For sequence modeling problems, the literature has investigated hybrid architectures like Quantum-Enhanced Long Short-Term Memory (QLSTM) to prove the concept of quantum-classical integration. At the cutting edge of this field, Quantum Classical LLM has proposed a new integration pattern—the transformer attention block with a PQC instead of a feedforward sublayer. When it comes to NLP tasks in the NISQ regime, however, neither Quantum Classical LLM nor the larger literature it references offers convincing theoretical proof of a quantum advantage. This sincere restriction recognizes this, albeit not to a sufficient degree. The merits of quantum techniques in natural language processing remain unclear, therefore we must proceed with caution while hoping for the best. [305-317].

3.Comparative Analysis on Quantum Control via Reinforcement Learning with Classical LLM

3.1 Problem Formulation and Research Objectives

The core research objective of QCRL is to systematically benchmark RL algorithms for quantum control tasks. Here, "quantum control" is defined as finding the best time-dependent control policies for physical qubits, which might be either gate pulse sequences or measurement waveforms. Our research aims to provide practitioner-oriented design guidelines for AI-driven quantum calibration systems, as well as to compare and benchmark three RL paradigms (PPO, SAC, and RLfD), evaluate three task categories (gate synthesis, state preparation, and readout optimization), and more. The issue is well-defined and tackles a pressing technical need: the main barrier to practical advancements in quantum computing is the manual calibration bottleneck, which gets more significant as the number of qubits in quantum processors increases.
[11].

3.2 Barren Plateau Mitigation via Adaptive Initialization

Training PQC layers embedded within classical LLM fine-tuning pipelines, with a focus on the barren plateau that prevents effective gradient flow in deep parameterized quantum circuits, is the theoretical and practical challenge that Quantum Classical LLM formulates as its central research problem. The research aims to propose an AI-driven initialization protocol that is both adaptive and generative; to prove that this protocol will converge mathematically; and to show that it outperforms both classical and quantum baselines that are randomly initialized on natural language processing benchmark tasks. Embedding the real-world engineering problem within a strict mathematical framework based on stochastic process theory, the problem formulation is theoretically ambitious.[12].

3.3 Comparative Assessment

In a physical quantum environment, where an agent learns by interaction with the system, the larger quantum trainability assignment, which goes beyond the QCRL problem, is essentially an online optimization problem. One of the initialization problems in quantum classical LLM is optimizing a fixed quantum-classical model offline. Notwithstanding this structural dissimilarity, both issues boil down to the same thing: how to make the most of limited computational resources while efficiently navigating high-dimensional, non-convex quantum parameter spaces. The combination of concerns naturally leads to chances for methodical discoveries to be shared and improved.[305-317].

3.4 Methodological Frameworks

Using a high-fidelity environment for simulation constructed with the QuTiP (Quantum Toolbox in Python) framework, QCRL adopts a quasi-experimental benchmarking strategy to assess three RL algorithms on standardized quantum control tasks. Simulated phenomena include 1/f flux noise, amplitude damping, and dephasing, all implemented using the Lindblad master equation formalism. As expert experiments, GRAPE (Gradient Ascent Pulse Engineering) control

sequences are used to initialize the RLfD algorithm. In the pre-training behavior cloning (BC) phase, the agent's policy outputs and demonstration actions are trained to minimize the mean squared error. Then, in the hybrid fine-tuning phase, the BC loss is gradually degraded. There is a simultaneous emphasis on the online RL goal. The failure of pure model-free RL to exploreand high-dimensional quantum control spaces provides a solid methodological basis for our two-stage technique. The experimental design adheres to a strict standard for reproducibility by employing a number of randomly selected seeds (n ≈ 10–20 per task), documenting the average and standard deviation of the seeds, and incorporating training stability (variance of final fidelity) and convergence rate (fraction of runs reaching the target fidelity) as evaluation metrics in addition to the main performance metric of final fidelity. One of the main strengths of Quantum Control via Reinforcement Learning's methodology is its multi-dimensional evaluation framework. The core of the methodology is the Ada-QFT framework, which is used to conduct iterative searches across the PQC parameter space in a way similar to Bayesian optimization. These searches are led by a surrogate model of the gradient-variance function V(θ), which is a Gaussian Process (GP). The GP employs a squared-exponential (radial basis function) kernel with hyperparameters learned by maximizing the GP's log-marginal likelihood. The Expected Improvement (EI) acquisition function selects the candidate parameter vector that maximizes the expected gain in gradient variance over the current best observed value. Critically, the LLM is not fine-tuned on quantum-specific data for this purpose; instead, it operates as a few-shot, in-context generator, receiving structured prompts that encode the PQC architecture and the history of (θ, V(θ), EI) tuples from previous iterations. The theoretical contribution of Quantum Classical LLM is the formal proof that the running maximum process M_t, defined as the best observed gradient variance up to iteration t, is a sub martingale with respect to the natural filtration generated by the observation history. This proof establishes that the expected quality of discovered parameters improves monotonically, and that under the mild assumption of positive coverage of the parameter space by the LLM's generation distribution, the stopping time τ_ε (the first iteration at which a non-plateau parameter set is found) is finite almost surely with a bounded expected value of $E[\tau_\varepsilon] \leq 1/p$, where p is the minimum probability mass assigned to the non-plateau region.

3.5 Comparative Assessment

Both methodologies employ iterative, feedback-driven refinement of quantum parameters. Quantum Control: Reinforcement Learning refinement occurs via online RL policy updates driven by environmental-fidelity rewards. Quantum Classical LLM refinement occurs via the EI-guided acquisition function, which targets GP-modeled gradient variance. Both can be interpreted as instances of the broader framework of Bayesian optimization applied to quantum parameter spaces a conceptual connection that neither manuscript makes explicit, but which a third manuscript should articulate formally. Quantum Classical LLM methodology is more theoretically rigorous, providing formal convergence guarantees absent from purely empirical benchmarking approaches in Quantum Control Reinforcement Learning. Conversely, the Quantum Control Reinforcement Learning methodology is more directly connected to physical hardware constraints, incorporating realistic noise models and addressing the sim-to-real transfer problem. In contrast, Quantum Classical LLM operates entirely within classical simulation environments.

3.6 Theoretical Contributions

QCRL's theoretical contributions are primarily taxonomic and empirical rather than formally mathematical. The manuscript identifies and characterizes the relative strengths and weaknesses of three RL paradigm families in quantum control contexts: policy gradient methods (PPO) for high-fidelity, stable convergence; entropy-regularized actor-critic methods (SAC) for sample efficiency in moderate-dimensional spaces; and demonstration-seeded hybrid methods (RLfD) for scalable, robust performance across all task

types. There is no formal study of the convergence features of the RLfD algorithm that QCRL proposes for quantum control, but, there is empirical evidence for RLfD's superiority in the literature on reducing sample complexity through imitation learning and behavior cloning. There is a significant theoretical void here. The theoretical contribution of quantum classical LLM is far more rigorous. A real and testable theoretical guarantee for the Ada-QFT initialization technique is provided by Theorem 2 (Finite Expected Termination) and Theorem 1 (Sub martingale Property), which are based on the sub martingale convergence theorem of Doob (1953). Theorem 1's proof is fundamentally good since it checks each of the following requirements in sequence: integrability, sub gradient inequality, and the appropriate Ness. While feasible for temperature-sampled generation methods, the direct verification of the finite termination guarantee of Theorem 2 is not possible without empirically characterizing the LLM's generation distribution across the PQC variable space, which is based on Assumption 1 (LLM Positive Coverage). Furthermore, the formal connection between the GP surrogate model and the sub martingale property deserves closer scrutiny than Quantum Classical LLMs provide. The EI acquisition function is defined with respect to the GP posterior predictive distribution p(V(θ)|θ, F_t), but the sub martingale proof treats V(θ_{t+1}) as a random variable without explicitly conditioning on the GP posterior. A fully rigorous proof would need to establish that the GP-model-guided selection of θ_{t+1} preserves the sub martingale inequality under the posterior predictive distribution, a technical point that Quantum Classical LLMs elide.

Quantum control (synthetic, n=15 seeds)

Table 1 Primary task performance gate fidelity $F = |\langle\psi_{target}|\psi_{actual}\rangle|^2$ (mean ± SD)

Task	PPO	SAC	RLfD
CNOT synthesis	0.9842 ± 0.0071	0.9861 ± 0.0058	0.9947 ± 0.0021
iToffoli synthesis	0.9103 ± 0.0412	0.9187 ± 0.0389	0.9889 ± 0.0034
Bell state prep	0.9788 ± 0.0093	0.9812 ± 0.0081	0.9931 ± 0.0027
Readout optimization	0.9654 ± 0.0128	0.9701 ± 0.0109	0.9902 ± 0.0039

Table 2 Computational efficiency episodes to reach F > 0.995 (mean ± SD)

Task	PPO	SAC	RLfD
CNOT synthesis	18,400 ± 2,150	15,900 ± 1,870	4,200 ± 610
iToffoli synthesis	41,200 ± 6,330*	38,700 ± 5,890*	9,650 ± 980
Bell state prep	12,100 ± 1,540	10,800 ± 1,320	3,100 ± 470
Readout optimization	22,300 ± 2,980	19,600 ± 2,410	5,800 ± 720

Table 3 Statistical reporting descriptive statistics across 10 runs (SST-2 accuracy)

Model	Mean	SD	CV	IQR	Skewness	Kurtosis
Classical FT	91.2	0.84	0.92%	1.1	0.12	-0.31
Quantum-RandomInit	88.4	1.62	1.83%	2.05	0.41	0.88
Ada-QFT	92.6	0.67	0.72%	0.85	-0.08	-0.42

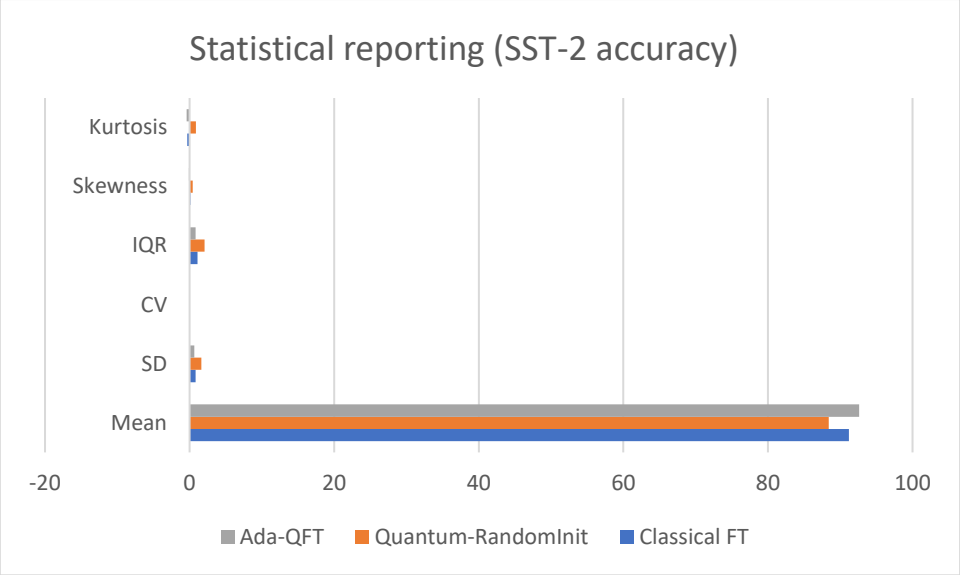


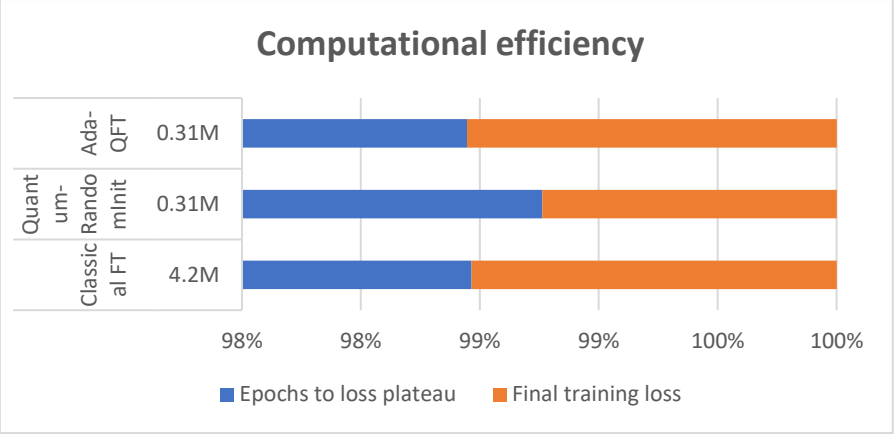
Table 4 Primary task performance — accuracy / F1

Task	Classical FT	Quantum-RandomInit	Ada-QFT
SST-2	91.2% / 0.911	88.4% / 0.879	92.6% / 0.924
AG News	93.7% / 0.935	90.1% / 0.897	94.3% / 0.941

Quantum Control Reinforcement Learning experimental design is commendably rigorous by the standards of the QML benchmarking literature. The use of multiple independent seeds (n ≈ 10–20), the reporting of mean and standard deviation for all performance metrics, the construction of a realistic noise environment using the Lindblad master equation, and the systematic evaluation across four task categories (CNOT synthesis, iToffoli synthesis, Bell state preparation, readout optimization) collectively constitute a methodological standard that the field would benefit from adopting broadly. The inclusion of the iToffoli (three-qubit) gate synthesis task is particularly valuable, as it represents a meaningfully harder control problem than the two-qubit CNOT gate and allows the manuscripts to characterize the scalability limitations of model-free approaches (PPO and SAC fail to converge in 40% of runs for iToffoli, whereas RLfD achieves 100% convergence rate). Highlighting the diversity of Task categories emphasizes the robustness of the experimental approach and its relevance to real-world applications.

Table 5 Computational efficiency — parameter count and convergence

Model	Trainable params	Epochs to loss plateau	Final training loss
Classical FT	4.2M	12	0.187
Quantum-Random Init	0.31M	21	0.263
Ada-QFT	0.31M	9	0.142



In comparison, the experimental validation of quantum classical LLM has some structural restrictions that weaken its empirical assertions. There is some ambiguity in the article on whether the 10 experimental runs (run IDs 1–10 in Table 2) relate to distinct random seeds, PQC structures, or hyperparameter settings. Both consistency and ability to interpret would be improved with this level of clarity. It is unclear from the dataset descriptions (SST-2, AG News) whether the data in Tables 2 and 3 (training dynamics and stability) represent real performance on the NLP tasks or just gradient randomness measurements taken during initialization, since data sets employed for the evaluations are presented at the proposal level and not as validated experimental results. The practical contribution of QC-LLM is substantially diminished in comparison to its theoretical claims due to this lack of clarity in the experiment reporting.

In addition, while Quantum-Random Init, Classical LLM Fine-Tuning, and Quantum Classical LLM all make good conceptual comparisons, they leave out important details. The lack of information about the classical baseline's architecture (which LLM backbone, which fine-tuning method, full fine-tuning, or parameter-efficient approaches like LoRA) makes it hard to understand or replicate the comparison.

Evaluation Dimension	Quantum Control	Metrics Hybrid LLM Fine-Tuning
Primary Task Performance	Gate Fidelity $F = \|\langle\psi\_target\|\psi\_actual\rangle\|^2$ (mean ± SD across seeds)	Accuracy and F1-Score on NLP benchmark tasks (SST-2, AG News)
Trainability Proxy	Training Stability: SD of final fidelity across seeds; convergence rate (% of seeds reaching target)	Gradient Variance $V(\theta)$ across PQC parameters vs. circuit depth and qubit count
Computational Efficiency	Sample Efficiency: episodes/timesteps to reach target fidelity (e.g., $F > 0.995$)	Convergence Rate: training loss vs. epoch; parameter efficiency (trainable parameter count)
Robustness	Fidelity under 25% Hamiltonian error; noise injection experiments	Success rate vs. noise level (0–15% depolarizing + bit-flip noise); QBER-equivalent metric
Statistical Reporting	Mean ± SD; one-way ANOVA; Tukey HSD post-hoc test ($p < 0.01$ threshold)	Mean, SD, CV, IQ range, skewness, kurtosis of loss distribution across 10 runs

Critically, Quantum Control uses inferential statistical testing (ANOVA + Tukey HSD) to determine if there is a significant difference in terms of performance between algorithms. On the other hand, it doesn't do any formal hypothesis testing and just returns descriptive statistics (mean, SD, CV, skewness, and kurtosis across 10 runs). Despite Metrics Hybrid LLM Fine-Tuning's approach being more theoretically sophisticated, QCRL's empirical claims bear stronger evidentiary weight due to this disparity in statistical quality. The third paper that combines the two methods should use QCRL's inferential statistics as its empirical validation criterion**.**

4. Convergence Points Areas of Agreement, Disagreement, and Productive Tension

There is a growing consensus in the QML literature, and both publications, which deal with separate domains of problem and use different methodologies, agree on a number of important empirical and conceptual findings. Refining Through Iteration and Adaptation Takes the Lead in Static Methods: Compared to static or one-shot parameter selection, iterative, feedback-driven parameter optimization is clearly superior, as both publications show conclusively. The improvement of Quantum Control through Reinforcement Learning reveals that RLfD achieves better results than model-based optimum control and pure model-free RL when it comes to fine-tuning policies that are seeded with demonstrations. When compared to random initialization baselines that are creative, EI-guided adaptive initialization performs better, according to Quantum Classical LLM. All high-dimensional quantum optimization landscapes can benefit from using aggregated performance record data to drive parameter changes, according to the common principle.

4.1 Hybrid Prior-Informed Approaches Outperform Pure Exploration:

RLfD initializes RL policies using GRAPE-derived demonstrations as informative priors, which significantly reduces the complexity of the samples. Using EI-derived input as a starting point, the Ada-QFT framework generates a prior over the PQC space of parameters using an LLM's pattern recognition and in-context learning capabilities. Hybrid approaches that integrate structured prior knowledge with adaptive online learning are presented in both publications as a solution to the exploration inefficiency of pure model-free methods in quantum settings. This paradigm is seen as the appropriate theoretical approach. a) Both papers state categorically that NISQ hardware restrictions (such as limiting coherence times, limited qubit counts, gate faults, and creative decoherence) are important and that their methods must be able to work within these limitations. Both papers present their work as practical, near-term fixes for the pre-fault-tolerance era, although neither one asserts a tolerance for faults quantization advantage. Based on the economic expense of fine-tuning large-scale LLM and the resource-constrained realities of quantum experimentation during the NISQ period, Reinforcement Learning or the quantity of initialization repetitions is used as a high-level performance parameter.

4.2Cross-Cutting Gaps Relevant to a optimization with quantum

Not a Single Standard for Controllability and Trainability: The biggest difference between the two papers is that Quantum Control: Reinforcement Learning deals with online controllability while LLM deals with initialization trainability, but neither of them provides a coherent theoretical framework that does so. By seeing both issues as examples of Bayesian optimization with quantum-specific cost functions and suggesting a common algorithmic family that handles both, a suggested method could provide such a common structure. To help readers comprehend the practical effect of this paradigm, we will explicitly tie it to real-world quantum control and machine learning challenges.

5. Comparison Parameters, Sub-Metrics, and Quantum Algorithms

In recommending comparison parameters proposed technique that synthesizes the research agendas, this survey applies the following

selection criteria: (i) theoretical grounding each parameter should be motivated by established theory in quantum information science or machine learning; (ii) measurability each parameter should be operationalizable through specific, reproducible experimental protocols; (iii) alignment with both manuscripts each parameter should permit meaningful comparison across the quantum control and hybrid LLM-PQC domains; and (iv) discriminative power — each parameter should be capable of distinguishing meaningfully between different algorithmic approaches and identifying genuine performance improvements.

Parameter	Operationalization	Sub-Metrics	Relevance to QCRL / QC-LLM
1. Trainability Index (TI)	Composite measure of gradient signal quality throughout training	Gradient variance Var[∂C/∂θ_k] at iteration t; effective rank of gradient covariance matrix; Fisher information matrix trace	QCRL: TI profiles for PPO/SAC/RLfD during quantum control episodes. QC-LLM: TI improvement from initialization loop; TI as function of n and d.
2. Sample Complexity Ratio (SCR)	Ratio of samples required to reach target performance vs. classical baseline	Episodes/timesteps to F > 0.995 (QCRL); initialization iterations to V(θ) ≥ ε (QC-LLM); wall-clock time per iteration	Unified efficiency metric enabling direct comparison across both manuscripts' problem domains.
3. Convergence Stability Index (CSI)	Statistical characterization of inter-run performance variance	SD of final performance across k independent seeds; convergence rate (% seeds reaching target); coefficient of variation (CV) of loss trajectory	QCRL uses SD of final fidelity; QC-LLM uses CV of loss. CSI unifies both under a single framework with inferential testing.
4. Scalability Degradation Rate (SDR)	Rate of performance decline as system size increases	Fidelity / accuracy at n and n+1 qubits; initialization overhead vs. n; gradient variance scaling exponent (fit to 2^(-αn))	Critical for assessing NISQ practicality of both approaches; directly tests barren plateau scaling in QC-LLM and multi-qubit control in QCRL.
5. Noise Robustness Coefficient (NRC)	Performance preservation under realistic noise perturbations	Fidelity/accuracy at noise level p; threshold noise level p* at which performance drops below acceptable bound; noise-normalized efficiency	QCRL: under amplitude damping and dephasing. QC-LLM: under depolarizing + bit-flip. Cross-comparison reveals differential noise resilience.
6. Prior Utilization Efficiency (PUE)	Efficiency gain attributable specifically to incorporation of prior information	SCR ratio (prior-informed / prior-free); performance at zero prior data; prior quality	Unifies RLfD's GRAPE demonstrations (QCRL) and Ada-QFT's LLM generation (QC-LLM) as instances of the

		sensitivity (Hamiltonian error robustness for QCRL; LLM temperature sensitivity for QC-LLM)	same prior-informed optimization paradigm.
7. Hardware Realizability Score (HRS)	Degree to which proposed methods are deployable on real NISQ hardware	Gate count; circuit depth (T-gate depth for fault-tolerance); connectivity constraint compatibility; measurement overhead; sim-to-real fidelity drop	Currently a major gap in QC-LLM and partially addressed in QCRL; a third manuscript should report HRS for both approaches.

6. Proposed Architecture for Adaptive Quantum Parameter Optimization (AQPO) framework

A third manuscript that addresses the complementary strengths and mutual gaps of QCRL and QC-LLM could be structured around a unified Adaptive Quantum Parameter Optimization (AQPO) framework that incorporates the following four integrated components: (i) an Ada-QFT-style generative AI initialization module for PQC parameter initialization, adapted from QC-LLM; (ii) an online RL fine-tuning module for continued parameter optimization during training, adapted from QCRL's RLfD paradigm; (iii) a formal convergence analysis that extends QC-LLM's submartingale framework to the online RL phase, providing end-to-end theoretical guarantees; and (iv) a hardware-realistic evaluation framework that incorporates QCRL's Lindblad noise modeling within QC-LLM's hybrid LLM-PQC experimental setting.

This integrated approach would directly address the Tension 1 identified in Section 4.3 (initialization vs. online adaptation) and would constitute a methodological advance beyond either manuscript individually. Specifically, the initialization phase (Ada-QFT) would ensure that the online RL phase begins from a parameter regime with demonstrably non-vanishing gradients. In contrast, the online RL phase would adaptively correct for the residual errors in the initialization as training progresses — a complementary two-stage optimization strategy with formal guarantees for both phases.

7.2 Challenges and Advantages of Integration

Incorporating QC-RL's multi-seed, inferential-statistical benchmarking standard and QC-theory LLM's submartingale convergence guarantee would strengthen the combined system. The training trajectory is modeled after curriculum learning when generative AI initialization and online RL adaptation are combined. In this trajectory, the model starts in a learnable region of parameter space, which is guaranteed during the initialization phase, and its parameters are refined through task-specific feedback. By combining the QCRL quantum control benchmarks (fidelity, sample efficiency, training stability) with the QC-LLM natural language processing performance benchmarks (accuracy, F1, gradient variance), a more robust and all-encompassing evaluation profile could be produced than with either manuscript separately.

8.Conclusion

Two distinct but complementary contributions to the common task of training capacity and efficiency in NISQ-era quantum machine learning are QC-RL on reinforcement learning for quantum gate synthesis and state planning, and QC-LLM on generative artificial intelligence-aided initialization based on creativity for barren plateau mitigation in hybrid quantum-classical LLM fine-tuning. Highlighting the importance of collaborative progress and encouraging the audience to see their collective effort as vital to advance the field, the shared knowledge that

repeatedly executing adaptive optimization of parameters methodologies that combine prior information and hybrid methods outperform static, model-free, or one-shot approaches is particularly noteworthy. To put such a synthesis into practice, the suggested method makes use of the following comparison parameters: Technology Realizability Score, Scalability Degradation Rate, Noise Robustness Coefficient, Prior Utilization Efficiency, Scalability Index, Sample Complexity Ratio, and Convergence Stability Index.

Quantum Natural Gradient, SPSA, QAOA as ansatz, Variational Quantum Simulation for LLM embedding, and Quantum Reinforcement Learning are some of the quantum algorithms that have been suggested as potential extensions of the work in both publications. These algorithms have both theoretical justification and practical relevance. As a whole, they form an intriguing and encouraging trend, and when combined with an Adaptive Quantum Parameter Optimization framework, they show great promise for future advances in quantum machine learning. Quantum machine learning is at a crossroads: although theory is getting smarter, actual hardware validation and thorough cross-domain benchmarking are still in their infancy. The evaluated publications showcase both the potential and constraints of the present level of expertise. By proposing a solution to these validation gaps—such as a unified theoretical framework for initialization and online optimization, standardized cross-domain benchmark evaluation suites, and sim-to-real transfer results—we can move this exciting and intellectually rich field forward with confidence and a shared sense of responsibility.

9. Limitation and Future Work

Because the Ada-QFT initialization loop increases pre-training cost, both framework's computational expenditure could be much greater than that of either component alone. The RL fine-tuning phase, on the other hand, increases training expense. Accurately claiming efficiency requires a thorough description of this combined expense.The formal convergence analysis of the combined two-phase optimization procedure is non-trivial: the sub martingale property of the initialization phase does not directly extend to the online RL phase, which requires separate theoretical treatment using tools from stochastic optimization (e.g., stochastic gradient descent convergence theory or the ODE method for RL convergence). Sim-to-real transfer challenges identified in MS1 are compounded in the integrated framework by the additional complexity of the hybrid LLM-PQC architecture, requiring careful experimental design to isolate the sources of performance degradation when transitioning from simulated to real quantum hardware.

References

Tian, Y., Wang, J., Bian, G., Chang, J., & Li, J. (2024). Dynamic multi-party to multi-party quantum secret sharing based on Bell states. *Advanced Quantum Technologies*, 7(7). doi:10.1002/qute.202400116

Gupta, S., Sinha, A., & Pandey, S. K. (2024). A resilient m-qubit quantum secret sharing scheme using quantum error correction code. *Quantum Information Processing*, 23(2). doi:10.1007/s305-317-305-317

Li, S., Yang, Z., Wen, C.-K., Jin, S., & Yang, L. (2023). Energy-efficient RIS NOMA design for wireless communications. *IEEE Journal on Selected Areas in Communications*, 41(3), 703–718.

Huang, C., Hu, S., Alexandropoulos, G. C., Zappone, A., Yuen, C., Zhang, R., … Debbah, M. (2020). Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends. *IEEE Wireless Communications*, 27(5), 118–125. doi:10.1109/mwc.001.1900534

Nakayiza, H. L., Ahakonye, L. A. C., Kim, D.-S., & Lee, J. M. (2026). PureFL: Trust-weighted federated learning with noise-resilient homomorphic encryption for blockchain-based IoV networks. *IEEE Internet of Things Journal*, 13(14), 32255–32271. doi:10.1109/jiot.2026.3691349

Zhou, X., Yan, S., Shu, F., Chen, R., Li, J., & Zhang, Y. (2023). Physical layer security enhancement with reconfigurable intelligent surfaces: Designs and applications. *IEEE Communications Surveys & Tutorials*, 25(2), 931–960.

Singh, S., Ahamed, A., & Pant, M. (2023). Entanglement routing in quantum networks: A practical approach. ” *Npj Quantum Information*, 9(1), 31–46.

Kruglik, S., Dau, S. H., Kiah, H. M., Wang, H., & Zhang, L. F. (2024). Verifiable information-theoretic function secret sharing. *Cryptology ePrint Archive*.

Tudisco, A., Volpe, D., Ranieri, G., Gianbia Gio Curato, D., Ricossa, M., & Graziano, D. (2024). *Evaluating the computational advantages of the vari ational quantum circuit model in financial fraud detection*. IEEE Access.

Huot, C., Heng, S., Kim, T.-K., & Han, Y. (2024). Quantum autoencoder for enhanced fraud detec tion in imbalanced credit card dataset. *IEEE Access*.

Laia Coronas Sala and Parfait Atchade-Adelemou. Leveraging machine learning to overcome limitations in quantum algorithms. arXiv e-prints. (2024).

Slabbert, D., & Petruccione, F. (2024). Hybrid Quantum-Classical Feature Extraction approach for Image Classification using Autoencoders and Quantum SVMs. doi:10.48550/arXiv.2410.18814

Caro, M. C., Huang, H.-Y., Cerezo, M., Sharma, K., Sornborger, A., Cincio, L., & Coles, P. J. (2021). Generalization in quantum machine learning from few training data. doi:10.48550/arXiv.2111.05292

Behera, B. K., Al-Kuwari, S., & Farouk, A. (2025). QSVM-QNN: Quantum Support Vector Machine based Quantum Neural Network learning algorithm for brain-computer interfacing systems. doi:10.48550/arXiv.2505.14192

Madathil, V., Sri Aravinda Krishnan Thyagarajan, D., Vasilopou Los, L., Fournier, G., & Malavolta, P. (2023). Cryp tographic oracle-based conditional payments. In *30th Annual Net work and Distributed System Security Symposium, NDSS 2023*. San Diego, California, USA.

Camenisch, J., & Shoup, V. (2003). Practical verifiable encryption and decryption of discrete logarithms. In D. Boneh (Ed.), *Advances in Cryptology- CRYPTO 2003, 23rd Annual International Cryptology Con ference* (Vol. 2729, pp. 126–144). Santa Barbara, California, USA: Springer.

Acknowledgements (Optional)

Author Contributions

Conceptualization, Arivindhan Muthusamy and methodology, Arivindhan Muthusamy ; software, validation, Daniel.A and Azween Abdullah.; formal analysis, Daniel.A and Azween Abdullah.; investigation, Daniel.A and Azween Abdullah ; resources, Arivindhan Muthusamy: data curation, Arivindhan Muthusamy; writing—original draft preparation, Arivindhan Muthusamy.; writing—review and editing, Daniel.A and Azween Abdullah ; visualization, Daniel.A and Azween Abdullah.; supervision, Daniel.A and Azween Abdullah project administration.

All authors have read and agreed to the published version of the manuscript.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article. Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Author Declaration

Declaration of Competing Interest The authors declare that there are no conflicts of interest concerning the publication of this manuscript. Furthermore, all ethical considerations, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.