QuantumNovelty Review Showcase

Paper audit Trainability and Expressivity of Hamming-Weight Preserving Quantum Circuits for Machine Learning (arXiv:2309.15547)

Generated by the QuantumNovelty paper-audit pipeline

June 2026

Project	QuantumNovelty (QN) — audit-and-falsify framework for quantum-computing research
Repository	https://github.com/boltzmannentropy/QuantumNovelty
Author	Shlomo Kashani (QNeura.ai)
LLM backend	Claude Code CLI (`2.1.1 (Claude Code)`)
Model snapshots used	`claude-haiku-4-5-20251001` `claude-opus-4-5-20251101`
Report generated	2026-06-10 23:18 by `build_reviews.py`

Disclaimer. These reviews are generated end-to-end by AI (the QuantumNovelty agent pipeline running on the Claude Code CLI). They are provided strictly for academic and demonstration purposes — to show what the framework produces on real papers. We make NO claims about the correctness, quality, novelty, or publication-worthiness of the papers under review; four passed real peer review at their respective journals and one is a public arXiv preprint, and nothing here should be read as criticism of the authors or as a substitute for human peer review. Paper copyrights remain with their authors and publishers.

1 Papers under review

Tag	Paper	Venue	arXiv
`hwqml`	Trainability and Expressivity of Hamming-Weight Preserving Quantum Circuits for Machine Learning Léo Monbroussou, Eliott Z. Mamon, Jonas Landman, Alex B. Grilo, Romain Kukla, Elham Kashefi	Quantum 9, 1745 (2025)	2309.15547

2 Token + cost ledger

Every LLM call records the model snapshot ID, exact input/output token counts, and USD cost from the Claude CLI's JSON envelope.

Paper	Stage	Input tk	Output tk	Cost ($)	Elapsed (s)
hwqml	Deep-research review	1	1,546	$0.4351	42.7
hwqml	Reviewer panel 5 voices	1	5,296	$0.5929	124.3
hwqml	Logical-fallacy report	1	2,217	$0.5223	41.4
hwqml	Stage-6 CQE narrative	3,005	1,121	$0.0761	31.9
hwqml total		**3,008**	**10,180**	**$1.6264**	**240.2**
Grand total		**3,008**	**10,180**	**$1.6264**	**240.2**

3 Verdict summary

Paper	Panel score	EIC verdict	CQE composite
`hwqml`	5.5/10	major-revisions	23/100

Panel score is the mean of the five voices' *Verdict: N/10* scores; the EIC verdict comes from the vote table; the CQE composite is the geometric mean of the six process-quality dimensions.

4 hwqml — *Trainability and Expressivity of Hamming-Weight Preserving Quantum Circuits for Machine Learning*

Léo Monbroussou, Eliott Z. Mamon, Jonas Landman, Alex B. Grilo, Romain Kukla, Elham Kashefi arXiv:2309.15547
Quantum 9, 1745 (2025)

4.1 hwqml — Deep-research review

Model: `claude-opus-4-5-20251101` · 1,546 output tokens · 42.7 s

4.1.1 1. One-paragraph summary of what the paper claims

This paper analyzes the trainability and controllability of Hamming-weight (HW) preserving variational quantum circuits (VQCs) for quantum machine learning. The authors make three main contributions: (1) they design and prove the feasibility of new heuristic data loaders that perform quantum amplitude encoding of (n choose k)-dimensional vectors using n-qubit circuits made of Reconfigurable Beam Splitter (RBS) or Fermionic Beam Splitter (FBS) gates, with existence proofs based on controllability arguments via the Quantum Fisher Information Matrix (QFIM) rank; (2) they prove that the rank of the QFIM for any VQC state is almost-everywhere constant (Theorem 1); and (3) they provide trainability analysis showing that the variance of the l2 cost function gradient scales as O(1/(n choose k)), proving conditions for existence/absence of Barren Plateaus for these circuits. Notably, they claim this represents a counterexample to a conjecture from [11] linking controllability (via DLA dimension) to trainability, since HW-preserving circuits can avoid Barren Plateaus even without full controllability or 2-design assumptions.

4.1.2 2. Audit-and-falsify checklist

☐ **Augmented baseline catalog:** PARTIAL — The paper compares against the theoretical framework of [11] (Larocca et al.) and discusses relationships to [17, 18, 25], which are contemporary works on DLA-trainability connections. However, for the data loading/encoding component, comparisons are primarily to [5] (the prior HW-preserving loader in unary basis) without systematic benchmarking against other amplitude encoding methods.

☐ **Strict-domination comparator:** NOT-APPLICABLE — The paper does not make Pareto optimality claims; the main results are theoretical (existence proofs, variance bounds) rather than empirical performance comparisons requiring tolerance calibration.

☐ **Recompute-from-raw:** PARTIAL — Figures 4, 6, 9, and 10 display numerical results, and Table 2 shows simulation errors. The theoretical predictions (dotted lines in Figs. 9-10) appear consistent with the plotted data, but no raw data files or recomputation scripts are referenced. The paper text indicates the simulation was done via Numpy/Qiskit but no audit trail is provided.

☐ **Wilson 95% CIs:** FAIL — Table 2 reports "Average Error" and "Variance" for 1000 samples but does not provide confidence intervals. Figures 9-10 show variance comparisons but without error bars or uncertainty quantification on the numerical estimates.

☐ **Cross-LLM falsifiability:** NOT-APPLICABLE — No LLM-in-the-loop methodology is employed; this is a theoretical/numerical quantum computing paper.

☐ **Honest negatives:** PARTIAL — The paper acknowledges limitations: FBS gates have reduced controllability (Section 2.2, Appendix C), classical simulability limits speedup (Section 3.3), and Theorem 2 relies on an unproven conjecture (Conjecture F1, acknowledged in Section 3.2). However, there is no dedicated "Failure Modes" section, and no cases where the proposed data loader fails to converge are reported.

☐ **Simulator precision floor:** NOT-APPLICABLE — The paper does not compare quantum vs. classical energy calculations; numerical simulations are for gradient variance and QFIM rank, which are computed symbolically/analytically where possible.

☐ **Auditable claims:** FAIL — No `audit_claims.py` or equivalent is provided. The paper does not reference a code repository, reproducibility package, or JSON files for numerical claims. Quantum journal requires Code Availability statements, but this appears as an arXiv preprint accepted to Quantum without such artifacts being explicitly linked.

4.1.3 3. Overall assessment

This paper presents rigorous theoretical work on an important topic (trainability of symmetric VQCs), with novel analytical results that bypass the standard 2-design assumptions used in prior work. The mathematical development is careful, with proofs relegated to extensive appendices. The main weakness from an audit-and-falsify perspective is **reproducibility infrastructure**: the numerical claims in Figures 4, 6, 9-10 and Table 2 lack uncertainty quantification, and no code/data artifacts are referenced for independent verification. The reliance on Conjecture F1 for Theorem 2 is honestly disclosed but represents an unresolved gap. The theoretical claims themselves appear internally consistent, and the counterexample to Conjecture 1 from [11] is well-argued. However, strict reviewer-mode auditing would flag the absence of confidence intervals on small-sample statistics and the lack of reproducibility artifacts as significant concerns for a journal requiring Code/Data Availability statements.

Research rigour score: 6/10 — Strong theoretical development with honest acknowledgment of limitations, but lacking the reproducibility infrastructure and statistical rigor expected under the audit-and-falsify framework.

4.1.4 4. Three highest-leverage improvements

1. **Add uncertainty quantification to all numerical results**: For Figures 9-10, report bootstrap or Wilson 95% confidence intervals on the estimated variance values. For Table 2, add confidence intervals on the mean error. The claim "variance follows theoretical values" in Figures 9-10 should be supported by statistical tests (e.g., š goodness-of-fit against the predicted 8k(nk)/(n(n1)d_k) values).

2. **Provide a reproducibility package with `audit_claims.py`**: Create a repository containing (a) code to regenerate all figures from raw simulation data, (b) JSON files storing the numerical data points underlying Figures 4, 6, 9-10 and Table 2, and (c) a script that re-derives every numerical claim (e.g., QFIM ranks, gradient variances, classification accuracies) from these files. This satisfies Quantum journal's Code Availability requirement and enables independent verification.

3. **Either prove Conjecture F1 or provide explicit numerical bounds**: Theorem 2 currently depends on Conjecture F1 about spectral gaps of stochastic matrices. Either (a) complete the proof (the numerical evidence in Appendix F.9 is suggestive but not definitive), or (b) provide explicit numerical values for the spectral gaps of T matrices arising from specific connectivities (e.g., nearest-neighbor, Rigetti ASPEN M2) with rigorous error bounds, so that the theorem's applicability can be verified for concrete cases.

4.2 hwqml — Reviewer panel 5 voices

Model: `claude-haiku-4-5-20251001` · 5,296 output tokens · 124.3 s

4.2.1 Voice 1 — Reviewer 1 (Physics correctness)

This paper investigates Hamming-weight preserving variational quantum circuits, specifically those constructed from Reconfigurable Beam Splitter and Fermionic Beam Splitter gates, analyzing their controllability and trainability properties. The physics foundation of the work rests on the observation that these gates preserve subspaces of fixed Hamming weight, reducing the effective Hilbert space dimension from 2ˆn to the binomial coefficient (n choose k). The authors correctly identify that the RBS gate implements a planar rotation between computational basis states |01 and |10 as shown in Equation 3, which is indeed the standard form of this gate used in photonic and fermionic quantum computing platforms. The Hamiltonian HRBS in Equation 4 is

correctly constructed as the generator of this rotation, though I note the authors do not explicitly verify that exp(-iHRBS) produces the stated unitary form, which would strengthen the presentation.

The treatment of the Fermionic Beam Splitter in Definition 4 deserves particular scrutiny. The authors introduce the fermionic parity factor f = _{i<l<j} s_l, which accounts for the anticommutation relations of fermionic operators when mapped to qubits via Jordan-Wigner transformation. This is physically correct and represents the key distinction between bosonic and fermionic systems. However, the paper's claim in Appendix C that "each block Wˆk for k > 1 is perfectly determined by Wˆ1 as it is the k-compound matrix of Wˆ1" requires more careful justification. While this relationship holds for the fermionic case due to the determinantal structure of fermionic wavefunctions, the authors should clarify whether this restricts the FBS architecture to Jordan-Wigner ordering specifically, or whether the result generalizes to other fermion-to-qubit mappings such as Bravyi-Kitaev or parity encoding. The DLA dimension upper bound of n(n-1)/2 for FBS circuits stated in Figure 4 appears consistent with this compound matrix structure, but the connection to physical symmetries (particle number conservation) could be made more explicit.

The Quantum Fisher Information Matrix analysis in Section 2.3 is mathematically sound, with the definition in Equation 9 matching standard formulations. The authors' claim in Theorem 1 that the QFIM rank is almost-everywhere constant follows from standard results on analytic functions and Jacobian ranks, though the proof in Appendix D could benefit from more explicit connection to the VQC structure. One technical concern is the numerical precision of QFIM rank calculations referenced in Figure 6: the authors should specify whether they used complex128 or complex64 arithmetic, as rank determination near singular matrices is notoriously sensitive to floating-point precision. The claim that "drawing a point uniformly at random and calculating its QFIM rank yields r() = rmax with probability 1" is mathematically correct but practically requires discussion of numerical tolerances used in rank determination.

The trainability analysis in Section 3 connects the gradient variance to the subspace dimension through Lemma 1, with the key result that gradients scale as O(1/d_k) rather than exponentially. The backpropagation formulation in Equations 15 and 16 correctly accounts for the multi-pair rotations induced by single gates in higher Hamming-weight subspaces. However, I am concerned about the cost function choice in Equation 13, which uses squared Euclidean distance between quantum states. The authors acknowledge in a footnote that this is not directly expressible as an observable expectation value, but the proposed ancilla-based workaround from reference [6] adds circuit depth that could affect the practical trainability claims. The proofs in Appendices E through G are technically detailed and appear correct, though the reliance on Conjecture F1 regarding spectral gaps of stochastic matrices represents an acknowledged gap in the rigorous chain of reasoning.

Questions for Authors: 1. Does the compound matrix relationship for FBS gates hold for Bravyi-Kitaev mapping, or is Jordan-Wigner ordering implicit throughout? 2. What numerical tolerance was used for QFIM rank determination in Figure 6, and how sensitive are the reported ranks to this choice? 3. Can you provide any analytical progress on Conjecture F1, or at minimum characterize the class of patterns for which it is provably true?

Verdict: 7/10. Recommendation: minor-revisions.

4.2.2 Voice 2 — Reviewer 2 (Algorithmic novelty)

The central contributions of this work must be evaluated against the rapidly evolving landscape of trainability analysis for variational quantum circuits. The authors position their work relative to Larocca et al. (2022, reference [11]), Ragone et al. (2024, reference [17]), and Fontana et al. (2024, reference [18]), claiming independence from these approaches and applicability in settings where the DLA-based framework does not apply. This positioning is partially justified: the recent work by Ragone et al. established that variance of cost gradients scales inversely with the DLA dimension under 2-design assumptions, while Fontana et al. extended this to include the adjoint representation framework. Both papers explicitly exclude or provide only upper bounds for Hamming-weight preserving circuits, and the present work fills this gap by providing exact variance formulas for RBS/FBS circuits without invoking 2-design assumptions.

However, the novelty must be tempered by examination of concurrent and recent work. Diaz et al. (2023, reference [25]) specifically addressed scenarios evading the DLA framework, providing theory that the authors acknowledge as "consistent" with their results. More critically, the data loading application in Section 2 closely parallels the approach of Johri et al. (2021, reference [5]) and Landman et al. (2022, reference [6]), which already demonstrated HW-preserving amplitude encoding in the unary basis. The extension to arbitrary Hamming

weight k represents an incremental generalization rather than a conceptual breakthrough. The controllability analysis via QFIM is also not novel—Haug et al. (2021, reference [29]) established the connection between QFIM rank and circuit capacity, and the algorithmic approaches in Section 2.3 (Algorithms 1 and 2) are straightforward applications of greedy rank-maximization strategies that have appeared in quantum control literature for decades.

The trainability results themselves, while technically sound, raise questions about practical relevance. The main theoretical contribution is that gradient variance scales as 8k(n-k)/(n(n-1)d_k) per Theorem 3, avoiding exponential barren plateaus for fixed k. However, this polynomial scaling with d_k = (n choose k) still implies vanishing gradients when k is not held constant. The authors acknowledge this in Section 3.3 but do not adequately address the practical regime of interest. For quantum machine learning applications requiring expressive circuits, small k implies limited expressivity (the authors themselves note in Section 2.2 that reduced k may be necessary "in hopes of achieving full controllability for a smaller subspace"). This creates a fundamental tension between trainability and expressivity that the paper does not resolve, and which was already identified in Holmes et al. (2022, reference [30]). The claimed absence of barren plateaus is therefore conditioned on operating in a regime that may be of limited practical interest.

Regarding computational comparisons, the classical simulability of HW-preserving circuits in polynomial-dimension subspaces is well-established (the authors cite Anschuetz et al. 2023, reference [27]). The running time analysis in Table 1 is straightforward but the claimed "polynomial advantage" for quantum training is misleading: if the circuit is classically simulable, quantum training offers no asymptotic speedup, only constant-factor improvements from parallelization of gate applications. The suggestion that one could "train them classically to represent classical data and then associate them with a quantum circuit that is harder to simulate" (end of Section 2.3) is speculative and would require a separate analysis of how trainability transfers across the transition from simulable to non-simulable regimes. This gap between the theoretical guarantees and practical quantum advantage undermines the paper's positioning as relevant to near-term quantum machine learning.

The comparison with the Conjecture 1 from Larocca et al. is interesting but overstated. The authors claim to "refute" this conjecture "in its full generality," but more accurately, they demonstrate a setting where the conjecture's hypotheses are not satisfied (HW-preserving circuits do not satisfy the full controllability assumption), so the conjecture simply does not apply. This is not a refutation but rather a delineation of the conjecture's domain. The positive contribution is showing that trainability can be analyzed without invoking the conjecture's framework, but this does not invalidate the conjecture for circuits where it does apply.

Questions for Authors: 1. For what range of k values do you expect the polynomial gradient scaling to remain practically useful for gradient-based optimization? 2. Can you quantify the "polynomial advantage" claimed for quantum training in concrete circuit parameters? 3. How do your trainability guarantees extend when HW-preserving blocks are combined with non-HW-preserving operations, as would be needed for practical applications?

Verdict: 6/10. Recommendation: major-revisions.

4.2.3 Voice 3 — Reviewer 3 (Empirical evidence)

The empirical validation presented in this paper is concentrated in Section 4 and the accompanying figures, with the bulk of the technical content being theoretical. While the mathematical results are substantial, the numerical evidence supporting these results suffers from several methodological shortcomings that limit confidence in the practical applicability of the findings. The authors present simulations but do not provide the statistical rigor expected in modern quantum computing research, particularly given the stochastic nature of VQC training and the sensitivity of gradient-based claims to initialization and sampling.

Figure 6 presents the evolution of QFIM rank for a periodic structure ansatz, but the methodology for rank determination is not specified. The authors claim to verify Theorem 1 ("rank is almost-everywhere constant") numerically, but the figure shows only single point evaluations without error bars or repeated sampling. Given that numerical rank computation is notoriously sensitive to condition number and floating-point precision, the authors should report the singular value spectrum and the threshold used to distinguish zero from nonzero singular values. The statement "upon randomly sampling parameter values [0, 2]^p, the value of rank[QFIM()] is independent of " requires empirical verification over multiple samples with statistical characterization of any observed variation. The claim that one random sample suffices to determine maximum rank (Lemma D3) is mathematically correct but practically requires confidence intervals on the numerical rank determination.

The gradient variance simulations in Figures 9 and 10 are more carefully executed, with 100 and 10000 random samples respectively. However, the agreement with theoretical predictions (dotted lines) is claimed but not quantified. The authors should report chi-squared goodness-of-fit statistics or at minimum the maximum observed deviation from theory. The figures show reasonable agreement but there appear to be systematic deviations at higher L values in Figure 9 panels (3) and (4) that are not discussed. More concerning, the experimental setup for Figure 10 fixes the initial state as a basis state rather than sampling uniformly, which the authors acknowledge places this "in between the assumptions of spherical t-designs of t = 1 and t = 2." This makes direct comparison with Theorem 3's predictions problematic, and the claim that "points only roughly follow the dashed 1/d_k values" is a red flag for potential systematic effects that are not captured by the theory.

The Fashion MNIST simulations in Section 4.1 and 4.4 represent the only realistic application test in the paper, but the experimental design is inadequate for drawing meaningful conclusions. Table 2 reports mean error and variance over 1000 samples for a single circuit configuration, without comparison to baseline methods or random initialization baselines. The binary classification results in Figure 11 show learning curves but do not report final accuracy values, number of training epochs, or comparison to classical neural networks of equivalent parameter count. The claim that "we do not claim that this plot exhibits any advantage to use RBS based quantum circuits as neural networks" is appropriate but undermines the relevance of including this experiment at all. If the goal is merely to demonstrate that training converges, this should be stated explicitly with appropriate caveats about the toy nature of the demonstration.

The reliance on Conjecture F1 for the main theoretical results (Theorem 2) is acknowledged but inadequately supported. The numerical evidence in Appendix F.9 examines only three specific patterns (line-down, line-downup, pyramid) for n 50 qubits and k {1, 2, 3}. The polynomial decay fits show rš values above 0.999, which is encouraging, but the fitted exponents (slopes of -1.99, -1.96, -0.95 in Figure F1) vary significantly across patterns, suggesting non-universal behavior that the conjecture does not capture. The authors do not examine patterns with higher connectivity or larger gate counts, which would be more representative of practical circuit designs. Furthermore, the conjecture requires spectral gap bounds that scale as (1/poly(n)), but the specific polynomial is not determined from the numerical fits, leaving the dependence on circuit architecture as an open question that directly affects the required circuit depth for Theorem F4 to apply.

Questions for Authors: 1. What singular value threshold was used for QFIM rank determination, and how does changing this threshold affect the reported ranks? 2. Can you provide statistical tests (chi-squared, Kolmogorov-Smirnov) for the agreement between observed gradient variances and theoretical predictions in Figures 9-10? 3. For Conjecture F1, can you characterize the polynomial exponent as a function of pattern properties, or at minimum provide bounds on the constants involved? 4. Is the code for reproducing all numerical results publicly available, and if so, does it include scripts for regenerating all figures from raw data?

Verdict: 5/10. Recommendation: major-revisions.

4.2.4 Voice 4 — Devil's Advocate

This paper represents a competent but ultimately incremental contribution that has been dressed up with elaborate mathematical machinery to obscure fundamental limitations. Let me be direct about the critical weaknesses that my colleagues have treated too gently.

The core claim of avoiding barren plateaus is technically true but practically vacuous. The trainability guarantee requires keeping the Hamming weight k fixed as the number of qubits n grows, which means the subspace dimension d_k = (n choose k) scales polynomially with n. But this polynomial scaling still implies gradient variance decreasing as 8k(n-k)/(n(n-1)d_k), which for k=2 becomes O(1/nş). Even avoiding exponential vanishing, a cubic decrease in gradient magnitude renders optimization impractical for n > 50 qubits without shot counts scaling as n to maintain fixed signal-to-noise. The authors conflate "not exponentially vanishing" with "trainable," which is a fundamental category error. The theoretical contribution is that gradients vanish polynomially rather than exponentially, but this distinction matters only if the polynomial degree is small enough for practical optimization, which the authors never establish.

The dependence on Conjecture F1 is more problematic than acknowledged. The conjecture asserts that spectral gaps of stochastic matrices associated with connected RBS/FBS patterns scale as (1/poly(n)), but the numerical evidence in Appendix F.9 examines only three elementary patterns for modest qubit counts. The

fitted decay exponents vary from -0.95 to -1.99 across patterns, suggesting the polynomial degree is architecture-dependent in ways not captured by the conjecture. More critically, the authors do not bound the constants in this polynomial scaling, so even if the conjecture holds, the required circuit depth L (ź d_k n) from Theorem F4 could be astronomically large for circuits of practical interest. The condition that angles must be located "at any constant fraction of the depth" means trainability is guaranteed only for gates in the middle of the circuit, leaving gates near boundaries potentially trapped in barren plateau-like landscapes. This limitation is buried in the technical conditions rather than highlighted as a fundamental constraint.

The data loading contribution in Section 2 is not novel and not practically useful. The authors acknowledge that their circuits are classically simulable for fixed k, which means the "quantum" data loader offers no advantage over classical preprocessing. The suggestion that one could combine classically-trained HW-preserving blocks with harder-to-simulate components is pure speculation without theoretical or empirical support. How does trainability transfer across this transition? Does adding non-HW-preserving gates destroy the polynomial scaling? These questions are not addressed, making the data loading application an elaborate solution to a non-problem.

The comparison with prior work is misleading. The authors claim to show that Conjecture 1 from Larocca et al. "does not apply" to HW-preserving circuits, but this is because the conjecture's hypotheses (full controllability, 2-design property) are not satisfied. This is not a refutation but a scope limitation. The positive contribution is providing an alternative analysis for circuits outside the conjecture's domain, but the authors overstate this as revealing "a setting where the link between controllability and trainability does not apply." In fact, there is still a link: lower controllability (smaller DLA dimension) correlates with better trainability (larger gradient variance), which is exactly what the DLA framework predicts for circuits that don't achieve full controllability. The HW-preserving case is consistent with, not contradictory to, the broader theory.

Finally, the experimental validation is embarrassingly thin for a paper of this length. The Fashion MNIST experiments use a 5-qubit circuit with k=2, yielding a 10-dimensional encoding space, and report only average errors without statistical tests, baselines, or any connection to the theoretical results. The claim in Table 1 of "polynomial advantage" for quantum training is unsupported by any timing experiments. The code availability is not addressed, making the numerical claims unverifiable. This is not the empirical standard expected for publication in Quantum.

Verdict: 4/10. Recommendation: major-revisions bordering on reject.

4.2.5 Voice 5 — Editor-in-Chief synthesis

Having carefully considered all four reviews, I find substantial agreement on the paper's strengths and weaknesses, with the Devil's Advocate raising important concerns that sharpen the critique without fundamentally changing the assessment. The paper makes genuine theoretical contributions to understanding trainability of Hamming-weight preserving quantum circuits, but suffers from overselling of practical implications and inadequate empirical validation.

Reviewer 1 confirms the physics foundations are sound, with appropriate treatment of RBS/FBS gates and their relationship to fermionic systems. The concern about fermion-to-qubit mapping dependence (Jordan-Wigner vs. Bravyi-Kitaev) is technical but important for the completeness of the results. Reviewer 2's critique of novelty is partially valid: the data loading application extends prior work incrementally, but the trainability analysis without 2-design assumptions does represent a methodological contribution. However, I agree with Reviewer 2 that the practical regime where fixed-k trainability guarantees are useful remains unclear. The fundamental tension between trainability (requiring small d_k) and expressivity (requiring large d_k) is acknowledged but not resolved, limiting the paper's impact.

Reviewer 3 and the Devil's Advocate both emphasize the inadequacy of empirical validation, and I concur this is a significant weakness. The reliance on Conjecture F1 without rigorous proof or comprehensive numerical support is concerning for a theoretical paper. The Fashion MNIST experiments are too superficial to demonstrate practical relevance. The authors must either strengthen the empirical evidence substantially or reframe the paper as purely theoretical, accepting that practical implications remain speculative. The Devil's Advocate's point about polynomial-but-still-impractical gradient scaling deserves explicit acknowledgment: showing O(1/poly(n)) instead of O(1/exp(n)) is theoretically interesting but practically meaningful only if the polynomial degree permits optimization within reasonable shot budgets.

The comparison with the DLA-based framework is nuanced. The authors do not "refute" Conjecture 1 from Larocca et al. but rather demonstrate a complementary setting where alternative analysis techniques are needed. This should be reframed from confrontational to complementary positioning. The treatment of FBS gates' reduced controllability and its implications for trainability is one of the paper's genuine insights, showing that the controllability-trainability relationship is more subtle than simple inverse proportionality to DLA dimension would suggest.

Regarding the target journal's requirements, the paper does include Author Contributions (Section 6) and Acknowledgments (Section 7), but lacks explicit Data Availability and Code Availability statements. The Methods content is distributed throughout the paper and appendices rather than consolidated, which is acceptable given the theoretical nature but should be rationalized. The paper fits Quantum's scope and the detailed appendices are appropriate for this venue.

Must-fix items before resubmission (ordered by severity):

1. Explicitly quantify the practical regime where trainability guarantees are useful: for what (n, k) combinations does the polynomial gradient scaling permit optimization within feasible shot counts (e.g., 10 shots)?

2. Either prove Conjecture F1 or substantially expand numerical evidence to include diverse pattern architectures, larger qubit counts (n 100), and characterization of the polynomial exponent as a function of circuit structure.

3. Add proper statistical analysis to gradient variance simulations: report chi-squared goodness-of-fit values, confidence intervals, and discuss any systematic deviations from theory.

4. Clarify the fermion-to-qubit mapping dependence of FBS results, specifically whether compound matrix relationships hold beyond Jordan-Wigner.

5. Add explicit Data Availability and Code Availability statements with repository links for reproducing all numerical results.

6. Reframe comparison with DLA framework from "refutation" to "complementary analysis," accurately characterizing scope limitations rather than contradictions.

7. Either remove Fashion MNIST experiments or expand them to include proper baselines, statistical tests, and explicit connection to theoretical predictions.

4.2.6 Vote table

Voice	Recommendation	Confidence 1-10
Reviewer 1	minor-revisions	7
Reviewer 2	major-revisions	7
Reviewer 3	major-revisions	6
Devil's Advocate	major-revisions	8
Editor-in-Chief	major-revisions	8

4.3 hwqml — Logical-fallacy report

Model: `claude-opus-4-5-20251101` · 2,217 output tokens · 41.4 s

Fallacy: hasty-generalization **Severity:** medium **Location:** Section 1 (Introduction), paragraph 2 — "Quantum machine learning (QML) has become a promising area for real world applications of quantum computers" **Why it's the fallacy:** The abstract and introduction claim QML is "promising" for "real world applications" without citing evidence of demonstrated real-world utility. This generalizes from theoretical or small-scale demonstrations to broad practical applicability without sufficient supporting evidence. **Suggested fix:** Replace with a more measured claim: "Quantum machine learning (QML) has attracted significant research

interest as a potential application area for quantum computers, though practical demonstrations remain limited to small-scale problems."

Fallacy: conflated-regimes **Severity:** medium **Location:** Section 3.2, Theorem 2 and surrounding discussion — "Thus, after some polynomial amount of repetitions, and for angles located at any constant fraction of the depth, there is an absence of Barren Plateaus for CPSA ansätze." **Why it's the fallacy:** The theorem requires L to grow "at least as fast as nˆq" (polynomial in qubit count), but the practical implications for scalability to large n are not rigorously addressed. The proof relies on Conjecture F1 about spectral gaps, with numerical evidence only provided for n [4, 50] and k = 1, 2, 3. The claim of "absence of Barren Plateaus" is stated generally while the supporting evidence is limited to small system sizes. **Suggested fix:** Add explicit caveats: "For the system sizes tested numerically (n 50, k 3), we observe behavior consistent with absence of Barren Plateaus. Extension to larger systems relies on Conjecture F1, which remains unproven."

Fallacy: asymptotic-only-claim **Severity:** medium **Location:** Section 3.2, Theorem 3 — "Var_[_{_} C()] = k(n-k)/(n(n-1)) ů 8/d_k" **Why it's the fallacy:** The theorem establishes asymptotic scaling of gradient variance but does not provide finite-n bounds or discuss what values of n are sufficient for the asymptotic regime to accurately describe practical behavior. The verification plots (Figs. 9, 10) only show n up to 6 qubits. **Suggested fix:** Add a discussion of finite-size effects and the range of n for which the asymptotic formula provides accurate predictions, including error bounds.

Fallacy: active-space-handwave **Severity:** medium **Location:** Section 4.4 — "We do not claim that this plot exhibits any advantage to use RBS based quantum circuits as neural networks, but it illustrates that we can easily use such an architecture. Large simulation for more complex HW-preserving quantum neural networks for large value of k must be tackled in future work." **Why it's the fallacy:** The paper claims generalizability of the QNN approach to larger Hamming weights and more complex problems while explicitly acknowledging they have not actually run these experiments. This is a claim of generalization without supporting evidence. **Suggested fix:** Remove or weaken claims about generalizability to larger k until such experiments are conducted. The current disclaimer is appropriate but should be made more prominent.

Fallacy: circular-reasoning **Severity:** medium **Location:** Section 2.3, Algorithm 1 justification — "When the maximal rank (over parameter space) of the QFIM of a quantum data loader circuit is equal to dim(Sˆ{d_k1}) = d_k 1, we take it as evidence that it may achieve any state in Sˆ{d_k1}, i.e., achieve the amplitude encoding on the subspace of HW k." **Why it's the fallacy:** The paper uses QFIM rank as evidence of controllability, then uses controllability to justify the data loader design. However, high QFIM rank is a necessary but not sufficient condition for achieving arbitrary states—it indicates local controllability but does not prove global reachability. The reasoning is somewhat circular in using rank as both the design criterion and the validation. **Suggested fix:** Clarify that QFIM rank is a necessary condition that provides "support in favor of" but does not prove reachability. Add: "While maximal QFIM rank is necessary for full controllability, it is not sufficient; we therefore verify data loader capability empirically."

Fallacy: appeal-to-authority **Severity:** medium **Location:** Section 5 (Discussion), paragraph 2 — "In recent work [11], authors have shown that if a subspace-preserving VQC satisfies the assumption of full controllability of the subspace... Our results are independent and consistent with these papers" **Why it's the fallacy:** The paper repeatedly references the theoretical framework of [11, 17, 18] to position its results, but the relationship between these works and the current paper's claims is not always clearly delineated. The appeal to these prior works is used to bolster credibility without fully establishing the logical connection. **Suggested fix:** More clearly delineate which results are proven independently versus which rely on the framework of prior work. Specify: "Our Theorem 2 and 3 are proven using direct analytical methods that do not invoke the DLA-based framework of [11, 17, 18]."

4.4 hwqml — Stage-6 CQE narrative

Model: `claude-opus-4-5-20251101` · 1,121 output tokens · 31.9 s

Process Summary: QuantumNovelty Run Evaluation

4.4.1 Composite Verdict

The run achieved a **composite score of 23 out of 100**, calculated via geometric mean across six dimensions. Per the SKILL.md scale interpretation, this places the collaboration firmly in the **"Inadequate"** tier (scores 20-39), indicating that while some structural elements were attempted, the run failed to produce the core artifacts required for a credible scientific contribution. A score of 23 suggests the process stalled early—likely during setup or initial exploration—before any substantive methodology could be executed.

This is not a borderline failure. The geometric mean's sensitivity to low outliers means that multiple dimensions dragged the composite down, and indeed, the lowest-scoring dimension (Novelty rigour at 8) exerts disproportionate downward pressure. The run did not cross the threshold into "Marginally Acceptable" (40-59) because it lacks the fundamental outputs that would demonstrate even partial scientific validity.

4.4.2 Strongest Dimension: Communication (Score: 40)

Communication emerges as the relative high point, though "strongest" here is damning with faint praise. The score of 40 barely crosses into marginal acceptability. Both probes—**logical fallacies absent** and **reviewer panel verdict**—scored 40, but the evidence reveals why this dimension scored higher by default rather than by merit: the logical_fallacies skill was never run, and no review_panel.md was generated.

What this tells us about the run: the process never advanced far enough to produce text that could contain logical fallacies or receive reviewer scrutiny. The Communication score reflects an absence of failure rather than presence of quality. No claims were made poorly because no substantive claims were made at all. This is the scoring equivalent of a student receiving partial credit for writing their name on an otherwise blank exam.

4.4.3 Weakest Dimension: Novelty Rigour (Score: 8)

Novelty rigour scored a catastrophic 8 out of 100, anchoring the entire run in failure territory. The two probes reveal a complete breakdown in the baseline comparison stage:

- **Augmented baseline catalog present**: Scored 10/100 because "baseline_catalog has 0 rows." This means the run failed to populate any prior work against which novelty could be measured. Without a baseline catalog, any claim of novelty would be unfounded—there's nothing to demonstrate the work improves upon.
- **Strict-domination comparator run**: Scored 5/100 because "novelty_verdict.json not found." The comparator that would formally establish whether proposed methods strictly dominate existing approaches never executed.

The stage that produced this failure is unambiguous: **literature synthesis and baseline establishment**. This is typically an early-phase activity, suggesting the run collapsed before meaningful research could begin. The absence of a populated baseline catalog indicates either a failure to retrieve relevant prior work, a breakdown in the catalog construction pipeline, or premature termination during the preparation phase.

A Novelty rigour score of 8 is not recoverable through downstream work—it represents a foundational gap that invalidates any subsequent findings.

4.4.4 Three Highest-Leverage Improvements

1. Prioritize Baseline Catalog Population Before Any Experimental Work The most critical fix is ensuring baseline_catalog is populated with non-zero rows before proceeding. This requires explicit checkpointing: the pipeline should halt and surface an error if the catalog remains empty after the literature synthesis stage. The next run should allocate dedicated resources to querying arxiv, semantic scholar, or domain-specific repositories for relevant prior work in the quantum computing space. Without baselines, novelty claims are scientifically meaningless.

2. Implement Artifact-Existence Gates for Stage Transitions Multiple probes failed because expected artifacts simply don't exist: paper.tex (False), audit_claims.py (False), ablation_results.json (False), ratio_recompute.md (False), wilson_annotations.md (False). The next run should enforce hard gates requiring each artifact's presence before advancing to subsequent stages. This prevents the cascade failure observed here, where early-stage incompleteness propagated through to universal artifact absence.

3. Execute Domain Specification Probes Early and Explicitly Domain depth scored 30 with all three probes—**active space stated explicitly**, **fermion-to-qubit mapping stated**, and **simulator precision floor disclosed**—failing due to absent references. These are foundational quantum chemistry specifications that should be declared in project setup, not discovered missing at evaluation time. The next run should include a mandatory domain specification template completed before any simulation work begins, ensuring active space definition, mapping choice (Jordan-Wigner, Bravyi-Kitaev, etc.), and precision constraints are documented upfront.

Bottom line: This run produced essentially no usable scientific output. The composite of 23 reflects systemic early-stage failure, not isolated weakness. Recovery requires rebuilding from baseline establishment forward, with artifact gates preventing silent failures from propagating.