METHOD STUDY

The Configuration Atlas of Variational Classifiers

Which settings of the trainable quantum model can be trained: most of them, if you measure locally - and a one-minute screen tells you which. Five ansätze, four entanglement patterns, three measurements, four depths, three initializations, two readout heads and an encoding slice with data re-uploading, across four register sizes: 7,488 pre-training trainability measurements, 80 refereed trainings, eleven pre-registered hypotheses with the design tagged publicly before the first run. The product's default region is trainable and wins where it should; one measurement is dead at scale; most of the community's structural advice does not hold here. Eight of eleven registered hypotheses fall, our own warm-start bet among them, and they are printed with the findings. The results ship as a pre-training screen and guardrails. A method study behind the QML RQP.

September 2026

qubit-lab.ch

qubit-lab.ch
contact@qubit-lab.ch

Contents

The study on one page

A variational quantum classifier is a circuit with adjustable rotation angles: data goes in through an encoding, a stack of trainable layers turns it, a measurement reads numbers out, and a small classical head turns those numbers into a class. Training moves the angles by gradient descent. The product lets you choose the layer type (five ansätze), how the qubits are connected (the entanglement pattern), what is measured (two local observables or one global one), how deep the stack is, where the angles start, and the readout head - and the theory says some of these choices make the gradient vanish before training starts, so that hours of training move nothing. This study measured, for every combination the product offers, the size of the gradient and the spread of the output before training (7,488 measurements), then trained 80 pre-registered configurations against a full classical referee (the best classical models we can mount, which any quantum claim must beat), to see whether "trainable" means "wins".

The short answer is encouraging. In the region the product actually operates in - local measurements, its own small-angle initialization - most configurations are trainable up to 16 qubits, the default configuration beats the classical referee on the labels it was designed for at the largest size this series has ever won, and a one-minute pre-training screen separates the configurations worth training from the ones that are not. What fails is narrower than the theory's warnings suggest: one measurement choice is dead at scale, and most of the structural folklore (entanglement pattern, depth, re-uploading, warm start) does not hold under the product's initialization.

Everything was registered before the first run: the design was locked and tagged in the public repository; eleven hypotheses with decision rules were written down. Eight fall as registered and are printed with the same prominence as the findings - including our bet that a warm start would rescue almost every dead configuration. Two product knobs the study needed (data re-uploading, a global measurement) were added before the lock, so the study measures the product's real configuration space.

Throughout, "advantage" means a statistically significant predictive advantage against the specified classical referee under the stated conditions; never a speed-up, a hardware result, or advantage over all classical algorithms. This is not a new theory of barren plateaus - the mechanisms are known [1-4] - but a measured operating envelope for configuring one product, in a single chain: configuration → pre-training screen → refereed verdict → software behaviour.

The study produced five main findings:

1. **The product's default region is trainable, and it wins where it should.** With a local measurement and the product's default initialization, 86% of configurations screen as trainable at 16 qubits, and the gradient stays flat with depth. The default configuration beat the classical referee on its native labels at 4, 8 and 16 qubits - at 16 qubits with macro-F1 0.93 against 0.86 - and outscored every other trainable configuration on those labels in 14 of 14 comparisons. The default is the right default.
2. **The protection is the product's initialization.** In the theory's regime - random angles over the full circle - local gradients fall about 100-fold at 16 qubits; the product's small-angle start keeps them where they are. The explicit warm start adds little on top: our registered bet that it would rescue at least 90% of the dead configurations lost (29%), because the product had already built the protection in.
3. **One measurement is dead at scale, and nothing rescues it.** With the global parity observable the initial gradient halves with every qubit (3,912-fold from 4 to 16) under every initialization; the warm start lifts 20 of 336 such configurations, 2 of 8 rescue trainings improve, and 0 of 22 refereed runs with it beat the referee. The product now limits it to 11 qubits.
4. **Structure matters less than the folklore says - which simplifies configuration.** The entanglement pattern does not order the gradients (17 of 72 cells monotone), more parameters per layer only weakly, data re-uploading lowers the output variance at every size, and the quadratic readout head raises the loss gradient by 22% - a readout-scale effect. Measure locally, keep the default start, and most other knobs are free design choices rather than risks.
5. **A one-minute screen tells you which configurations are worth hours of training.** 9 of 80 trainings beat the referee; none of the 18 configurations that screened as concentrated did, 11 of 42 trainable ones did or reached parity with the kernel method. The registered yes/no cutoff predicts the outcome poorly (AUC 0.68); the graded gradient reading (0.88) and a 25-iteration probe (0.82) predict it well - at 58 seconds per 16-qubit screen against 2.1 hours of training. Both readings now ship in every run.

Scope, once and meant everywhere: exact simulation, no hardware, no shot noise; synthetic label families from the island studies - best-case envelopes, not real-world claims; one training recipe, not tuned per configuration; 4-16 qubits.

1. Management summary

Written for decision-makers; the measured basis follows in sections 2-7. Terms: a "configuration" is one combination of ansatz, entanglement, measurement, depth, initialization and readout; "trainable" means the pre-training screen finds a usable gradient and an output that still depends on the input; the "referee" is the set of classical models a quantum result must beat.

Q1 - Which settings can be trained at all? Most of them, if you measure locally: under the product's own initialization 87%, 69%, 65% and 61% of all configurations at 4, 8, 12 and 16 qubits, and the dead ones are almost all the global measurement. With a local observable 86% of configurations screen as trainable at 16 qubits; with the global parity 15% at 8 qubits, 11% at 12 and 5% at 16. The product now rejects the global measurement from 12 qubits and warns from 8 (section 6).

Q2 - Does the theory's fine print hold in our product? Where it warns about the measurement, yes; where it warns about structure, mostly no - because the product does not start where the theory's warnings apply. The exponential decay of the global observable's gradient is exactly as predicted [1]. But the theory's statements are made for random circuits, and the product does not start random: its small-angle initialization keeps local gradients usable to 16 qubits where uniform random angles would not. Under that initialization the entanglement pattern does not order the gradients at all, and depth does not hurt local measurements.

Q3 - If a configuration is trainable, does it win? When it should: on its native labels the default configuration wins up to 16 qubits; elsewhere trainable is a permit to try, not a promise. 9 of 80 refereed trainings beat the classical referee; 6 of the 9 are the product's default configuration on the labels it was designed for. Screening is a stop sign, not a promise: nothing that screened as concentrated won or reached parity with the kernel method, but three quarters of the trainable configurations did not win either. The one thing that predicts a win is a graded reading - how large the gradient is, or how far a 25-iteration probe gets - not the registered yes/no cutoff, which we set too coarse.

Q4 - What does the product do differently now? It screens every training run before spending the hours, rejects the measurement the study found dead at scale, and stops giving unmeasured advice. Five changes shipped: the global parity measurement is rejected from 12 qubits with the reason printed; every variational run prints a pre-training trainability screen with the study's calibration; the entanglement setting now applies to every ansatz (two template ansätze had silently ignored it); the warm-start advice states what the warm start actually buys; and the planner's folklore about entanglement and depth was replaced by the measured guidance.

The implications in four lines: **Measure locally** - the global observable is dead beyond a few qubits and nothing rescues it. **Trust the default start** - the product's initialization is the real protection; the warm start is a second attempt, not a guarantee. **Screen before training** - a concentrated screen is a stop sign; a trainable one is a permit to try. **Use the freedom** - entanglement pattern, depth and readout head are design choices, not risks, under the product's initialization.

Q5 - What does it cost to know before training? About a minute per configuration at 16 qubits against hours of training. The atlas: 7,488 screens in 75 processor-hours (58 seconds per 16-qubit screen). The verdicts: 80 trainings in 105 processor-hours, a median of 2.1 hours per 16-qubit arm before the referee even runs. Screening removes the dead configurations before that workflow starts; it does not replace it.

2. What was measured

Two layers again: the atlas layer measures three pre-training instruments on every configuration without training; the validation layer trains a pre-registered sample against the referee.

2.1 Factors

Factor	Levels
Ansatz × entanglement	hardware_efficient, real_amplitudes_like, custom_rx_ry_cz × {none, linear, circular, full}; basic_entangler and strongly_entangling with their built-in structure (14 combinations, see A.2)
Measurement	local Z (one observable per qubit), local Z+ZZ, global parity (one observable over all qubits)
Ansatz layers	1, 2, 4, 8
Register size	4, 8, 12, 16 qubits
Initialization	uniform random angles (the theory's regime); the product's default N(0, 0.1²); the identity-block warm start (at 12 and 16 qubits)
Readout head	linear on every cell; quadratic on the default-ansatz slice
Encoding slice	on the default ansatz: angle-Y, iqp, zz_like × 1, 2, 4 encoding repeats × bandwidth 0.25-2 × data re-uploading off/on
Replicates	3 datasets per size (fixed generator seeds)

168 main configurations × 3 replicates × 2 initializations = 4,032 records; 1,440 warm-start records; 288 quadratic-head records; 2,160 encoding-slice records: 7,488 in all.

2.2 The three instruments

All three are measured at initialization on 16 training rows, before any optimizer step, with four random draws of the starting angles. **I1** is the norm of the gradient of the product's training loss - the quantity the optimizer actually sees. **I1'** is the theory's quantity: the mean squared derivative of the first observable with respect to the first angle. **I2** is the variance of the first observable across rows - whether the output still depends on the input at all. A configuration is trainable by the registered rule when I1 ≥ 0.01 (one decade below the smallest healthy value in the island study) and I2 ≥ 0.001; otherwise concentrated. The validation layer adds **I3**, the validation score of a 25-iteration probe training.

2.3 The referee and the validation layer

80 arms: 12 anchors (the product's default configuration with the island study's fair recipe on every label family × size, so the known territory is re-measured), 60 arms drawn by a pre-registered stratified sampler (trainability class × family × size, seed fixed before the atlas finished), and 8 rescue arms - concentrated configurations re-trained with the warm start. Every arm: 600 iterations of Adam, early stopping, threshold-tuned on validation, scored against ten classical variants plus a tuned kernel SVM and a neural network; a win requires the whole 95% confidence interval of the paired-bootstrap Quantum Advantage Factor to clear 1.0. Every score is reported under both the product's tuned threshold and a plain 0.5 threshold (the island studies' collapse disclosure, designed in this time).

2.4 Registration

Design, hypotheses, decision rules and the sampler were locked and tagged in the public repository (**vqc-config-atlas-design-v1**) before the first record. From this study onward the series' registration grade is the public design tag; the previous study's independent deposit was made once and is not continued (methodology paper, revision 1.2). Appendix A lists every deviation and every analysis that was not pre-registered.

3. The atlas: which configurations can be trained

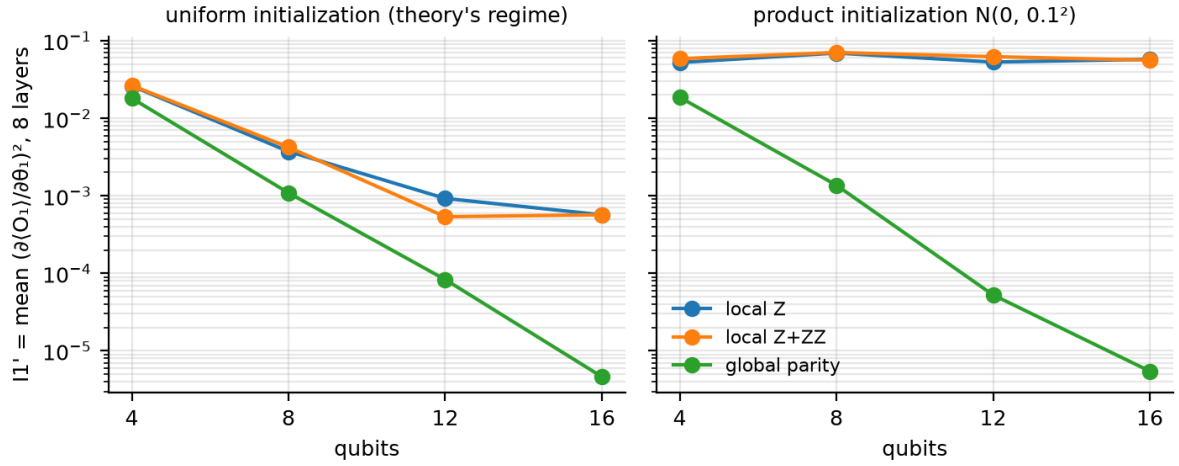

Figure 1 - The theory's quantity I1' against register size at 8 layers, by measurement, under uniform random initialization (left) and the product's default initialization (right). Log scale; medians over all ansätze, entanglement patterns and replicates.

S6-H1 - measurement locality (supported under the product's initialization; not as registered under uniform). The global observable's gradient falls exponentially with size: -0.83 decades of $\log_2$ I1' per qubit under the product's initialization, -1.10 under uniform - a halving per qubit, the textbook barren plateau [1]. The registered test asks whether the global slope is steeper than the local one in at least 80% of the 14 ansatz × entanglement combinations. Under the product's initialization it is (13 of 14); under uniform initialization only 9 of 14, because there the local gradients decay too and the difference vanishes in custom rx ry cz / full, hardware efficient / circular, hardware efficient / full, real amplitudes like / circular, strongly entangling / template. The finding behind the numbers: at 16 qubits the local Z gradient is about 100 times larger under the product's small-angle start than under uniform angles, and about 11,000 times larger than the global observable's under the same start.

S6-H2 - entanglement concentrates (refuted). The registered rule asks for a monotone decrease none → linear → circular → full in at least 75% of cells; 17 of 72 are monotone (24%, Wilson interval 15%-35%). The endpoints do differ - full is below none at 16 qubits and 8 layers by 0.096 (CI [-0.116, -0.071]) - but the pattern in between is not ordered. Which qubits are connected matters less than that the measurement is local.

S6-H3 - expressivity concentrates (supported, weakly). Ordering the five ansätze by parameters per layer, the rank correlation with the median gradient is negative at 12 and 16 qubits for every measurement (Spearman -0.63 to -0.16). With five ansätze the intervals are coarse; the direction is consistent.

S6-H4 - the registered bet: the warm start rescues everything (lost). Of 521 configurations concentrated under uniform initialization at 12-16 qubits, the identity-block warm start lifts 149 above the cutoff - 29% (Wilson 25%-33%) against the registered 90%. The product's own default start rescues 30%. The anatomy: 336 of the 521 are global-parity configurations, and the warm start rescues 20 of them - after a near-identity circuit the global observable's output barely depends on the input; the local-measurement configurations it rescues in 129 of 185 cases, the shallow ones (one or two layers) in 18 of 216. A warm start makes a deep local circuit trainable; it does not make a dead observable alive. The reading that matters for users: the product's default start already does most of what the warm start was meant to do.

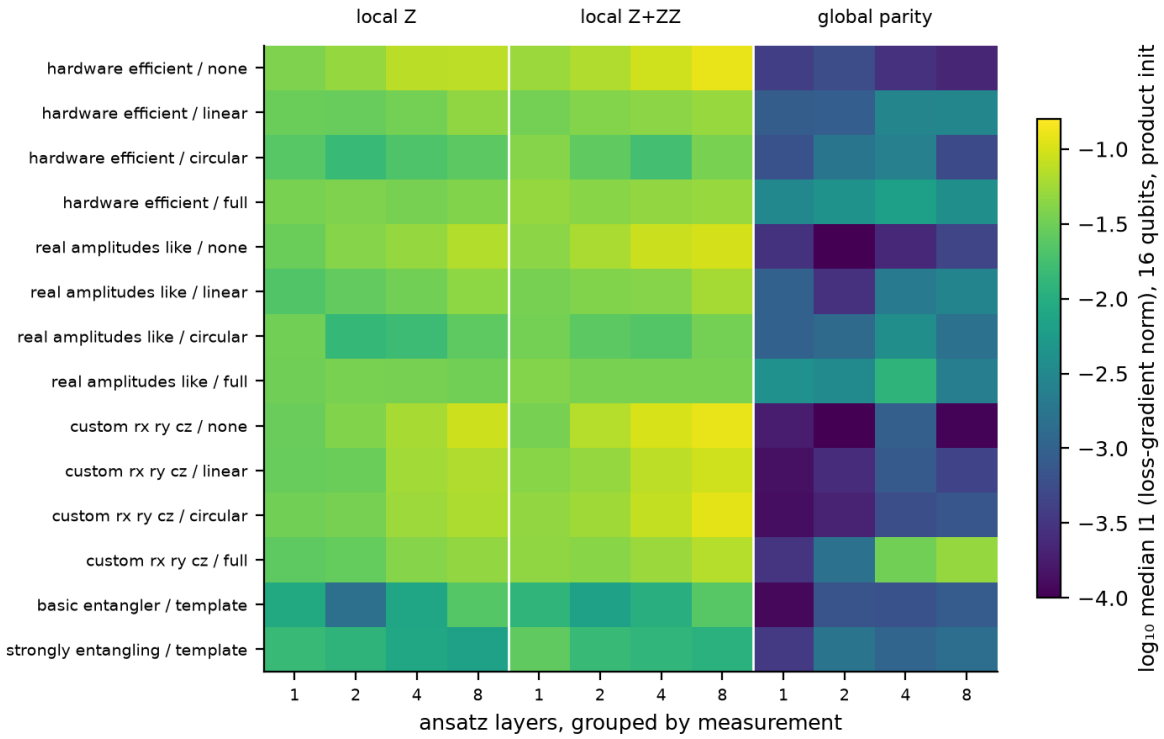

Figure 2 - The 16-qubit regime map under the product's initialization: log10 of the median loss-gradient norm I1 per configuration cell; rows = ansatz and entanglement, columns = layers, grouped by measurement. The global parity block is dark at every depth; local measurements stay bright and, if anything, brighten with depth.

4. Structure: readout head, re-uploading, encoding

S6-H7 - the readout head is irrelevant (null rejected). The quadratic head raises I1 by 0.0053 on average (CI [0.0051, 0.0056]) - 22% of the linear head's median, 288 paired cells. The circuit's gradient (I1') is head-independent by construction; what changes is the scale the loss puts on it. Real, small, disclosed.

S6-H8 - data re-uploading raises the output variance (refuted). Re-applying the encoding before every layer lowers I2 at every size: by 0.049, 0.048, 0.083 and 0.062 at 4, 8, 12 and 16 qubits, all intervals excluding zero. Richer function classes on paper [5]; a flatter output at initialization in practice. The knob stays in the product with that caveat printed.

5. Trainable is not sufficient: the validation layer

Of 80 refereed trainings, 9 beat the classical referee under the registered rule (9 under the plain threshold); 6 of the 9 are the product's default configuration. Every winner measures locally, has at most four layers and starts from the product's default initialization. 2 arms collapsed to a single class under the tuned threshold; verdict directions agree between the two threshold rules in 91% of arms.

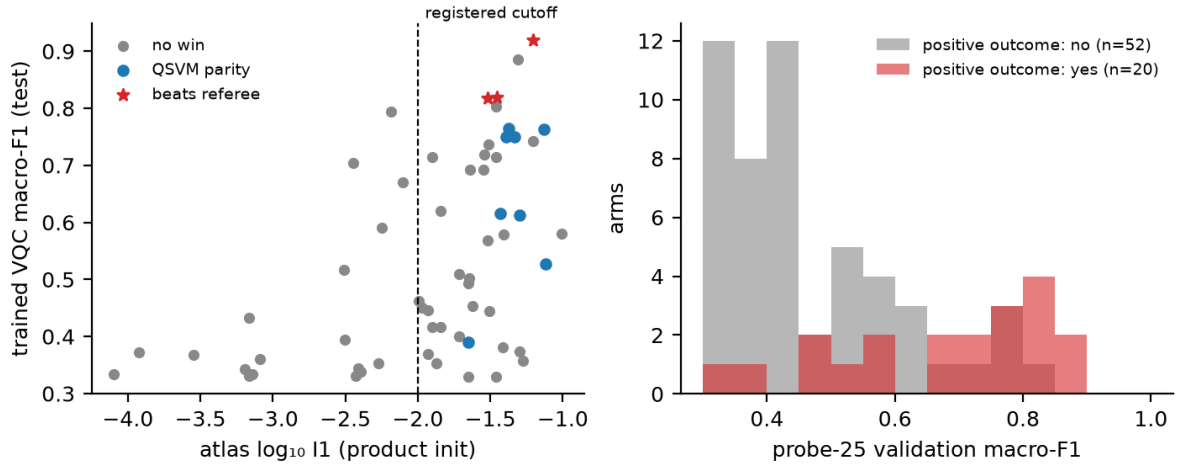

Figure 3 - Left: the 60 sampled arms, trained macro-F1 against the atlas gradient reading; stars beat the referee, blue reach parity with the kernel method, the dashed line is the registered cutoff. Right: the 25-iteration probe score for arms with and without a positive outcome.

S6-H5 - screening predicts verdicts (registered class: not supported; graded readings: supported). Over the 72 anchor and sampled arms, the registered yes/no class predicts "beats the referee or reaches parity with the kernel method" with AUC 0.68 (CI [0.62, 0.75]), below the registered 0.75. The reason is asymmetry, not noise: 0 of 18 concentrated arms had a positive outcome, but so did only 11 of 42 trainable ones - the class is a necessary condition, the same shape as the kernel atlas's "resolvable is not sufficient". The graded readings do better: $\log_{10}$ I1 predicts the outcome with AUC 0.88 (CI [0.78, 0.96]) and the 25-iteration probe with AUC 0.82 (CI [0.69, 0.93]; 0.77 under the plain threshold). The cutoff we registered from the island study's healthy values was too coarse; the instrument is not, and the product now prints the graded readings rather than the cutoff alone.

S6-H6 - native is specific; no entanglement is not quantum (supported). On the labels it was designed for, the default configuration outscores every other trainable configuration at the same size in 14 of 14 comparisons. No configuration without ansatz entanglement beat the referee (0 after the classical twin, 0 flagged untwinned; A.3).

S6-H9 - the rescue arms (refuted). 2 of 8 concentrated configurations re-trained with the warm start improved by the registered margin. Six of the eight are global-parity configurations, and their trained scores stay at the one-class floor. The warm start does in training what it did in the screen: little for the global observable.

6. What the product ships now

- **Global parity limited to 11 qubits**: rejected from 12 with the study's numbers printed, advisory at 8-11.
- **Pre-training trainability screen in every variational run**: I1, I1' and I2 at initialization with the registered cutoffs and the calibration of section 5; a concentrated screen prints a warning with the 0-of-18 fact.
- **Entanglement setting wired into every ansatz**: the two template ansätze had silently ignored it (A.2); **circular** now reproduces their built-in structure gate for gate, **none** is rejected for them.
- **Warm-start advice measured**: the guide and the planner state what the identity-block start rescues and what it does not.
- **Folklore retired**: the planner no longer claims that full entanglement or depth drives a circuit into a barren plateau under the product's initialization; it points to the screen.
- **Engineering by-products** shipped during the study: a memory-lean gradient (one backprop graph per row at 16 qubits, exact) and automatic simulator selection (the adjoint simulator for the global observable and for kernels, backprop for local observables).

7. What it cost

The atlas: 7,488 records in 75 processor-hours (4.8 seconds per record at 4 qubits, 58 at 16). The validation layer: 80 trainings in 105 processor-hours (118 with referee and probe), a median of 20 minutes per 4-qubit arm and 2.1 hours per 16-qubit arm. Per configuration the screen is about 127 times cheaper than the training it precedes at 16 qubits, before the referee, the probe and a human reading the result. With the caveat of section 5, which is the point: screening tells you what is dead, not what will win.

8. Where the potential lies

Three of the results point forward rather than closing a door. **The 16-qubit win on native labels** is the largest register at which any model in this series has beaten the classical referee; the natural next step is the noise-ladder treatment the QAOA studies received - the same trained circuit under shot noise and on hardware. **The graded screen** was calibrated on synthetic label families; calibrating it on customer data inside the product would turn a study instrument into a routine decision aid, and the 25-iteration probe suggests a cheap two-stage training protocol (probe every candidate, train the promising ones). **Data re-uploading** flattened the output at initialization, but the construction's promise is a richer function class; a per-round trainable input scaling - learned frequencies rather than a fixed bandwidth - is the version worth measuring next, and the knob now exists in the product to build on. Mixing different encodings per round was considered and set aside: the encodings the product offers share the same generator spectrum, so alternating them adds combinations, not frequencies. And since neither the entanglement pattern nor depth hurt local measurements under the product's initialization, deeper and more connected circuits are an open design space rather than a hazard.

9. Limitations

- **Synthetic label families** inherited from the island studies: best-case envelopes for the native family, not real-world data.
- **One training recipe** (the island study's fair recipe), not tuned per configuration; a configuration that fails under it might train under another.
- **Exact simulation only**; no shot noise in training, no hardware.
- **Pre-training instruments**: late-training plateaus are outside the atlas; the arms' gradient histories are the only view on them.
- **The registered cutoff** proved too coarse; the graded readings that work were pre-registered as instruments but not as the decision rule (section 5).
- **Eight registered hypotheses fall.** They are printed with the same prominence as the supported findings; most of them retire community advice rather than a capability of the product.

Appendix A - Registration, deviations and post-hoc analyses

A.1 Registration. Design tag **vqc-config-atlas-design-v1** pushed to the public repository on 2026-09-07 before the first record (same day, after the tag). Registration grade: public design tag.

A.2 Product observations recorded before the lock. The entanglement setting applied to the three hand-built ansätze only; the two template ansätze carried their built-in structure and ignored it - the grid measured them as their own rows. The product's **random** initialization draws small angles, not uniform ones; both were measured.

A.3 Deviations after the tag. (1) The classical twin for **entanglement = none** was registered as a product of single-qubit circuits; that holds for the angle encoding only, since the atlas's zz_like encoding entangles by itself - the twin was applied to angle-encoded cells, other **none** cells were flagged, not twinned. (2) I1 carries a readout-width bias: with one global observable the loss gradient is smaller than with 16 or 31 local ones at equal circuit health, which the registered cutoff does not correct; a width-normalized reading is reported as a sensitivity variant (it raises the global measurement's trainable share at 16 qubits from 5% to about a third; the majority stays dead). (3) After 5,483 records

and 3 arms the simulator backend was made selectable per job (adjoint for the global observable and forward-only work, backprop otherwise); both simulate exactly, agreement 10^{-15}, no measured quantity changed.

A.4 Run mechanics. The atlas and the validation layer ran on cloud workers after the first 5,483 records; records are append-only per shard and deduplicated by key; 80 of 80 arms present.

A.5 Post-hoc analyses (not pre-registered, labelled where used): the rescue anatomy by measurement and depth (section 3); the concentrated/trainable outcome cross-table and the graded-reading AUCs beyond the registered class test (section 5); the width-normalized I1 sensitivity (A.3).

A.6 Product changes made after the measurements, none touching the study's code path: global-parity limit, trainability screen, entanglement wiring, advice texts, simulator selection, memory-lean gradient.

References

1. M. Cerezo, A. Sone, T. Volkoff, L. Cincio, P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications 12, 1791 (2021); arXiv:2001.00550 - local versus global cost functions.
2. J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications 9, 4812 (2018); arXiv:1803.11173 - the original result for random circuits.
3. Z. Holmes, K. Sharma, M. Cerezo, P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum 3, 010313 (2022); arXiv:2101.02138 - expressivity.
4. C. Ortiz Marrero, M. Kieferová, N. Wiebe, Entanglement-induced barren plateaus, arXiv:2010.15968 (2020) - entanglement.
5. A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum 4, 226 (2020); arXiv:1907.02085 - data re-uploading.
6. E. Grant, L. Wossnig, M. Ostaszewski, M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum 3, 214 (2019); arXiv:1903.05076 - the identity-block warm start.
7. qubit-lab.ch, The Advantage Island of Variational Classifiers (August 2026) - the fair recipe, the label families and the instruments this study calibrates.
8. qubit-lab.ch, The Configuration Atlas of Quantum Kernels (September 2026) - the companion atlas for the kernel method.
9. qubit-lab.ch, The RQP Evidence Engine (August 2026, revision 1.2) - the registration grade.

Disclaimer

This document reports a method study on synthetic benchmark data with a rapid quantum prototyping tool by qubit-lab.ch. Its figures are unedited outputs of the described runs, intended for technical review, education and structured discussion. The configurations found trainable or concentrated are properties of the tested circuits on the tested label families under one training recipe; they are not financial advice, not a performance claim for real-world datasets, and not a hardware result. qubit-lab.ch is not affiliated with the authors of the cited references.