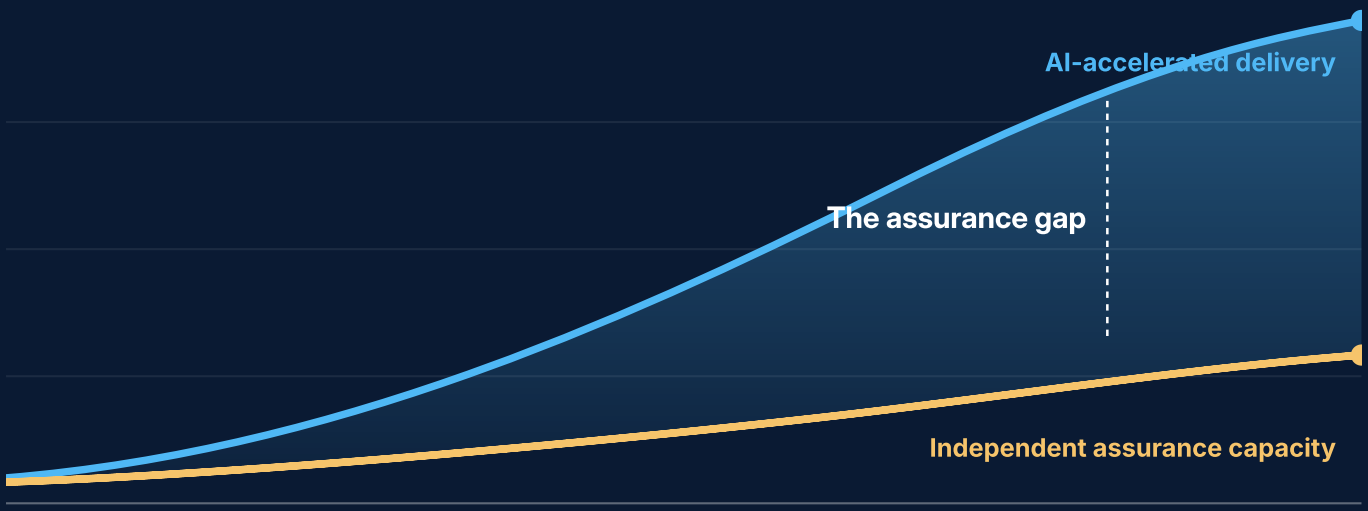
IQE LABS | RESEARCH WHITE PAPER 03

MARKET ISSUES & RESEARCH AGENDA

The AI Quality Assurance Gap

Why faster software delivery is creating new demands for testing, evidence and governance



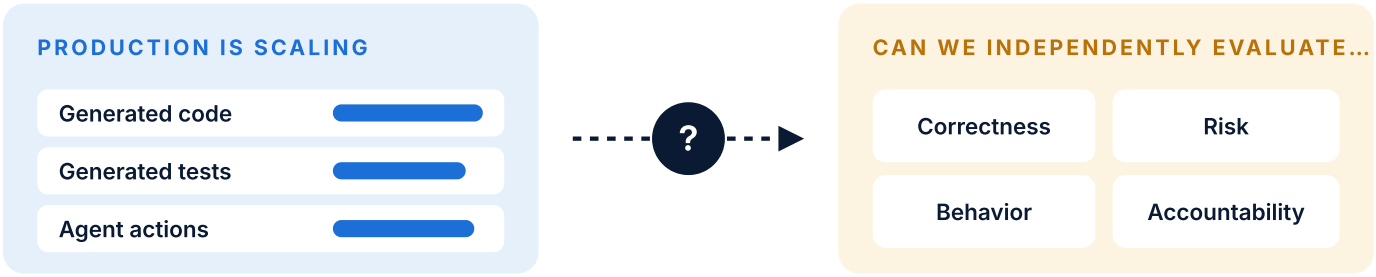
Conceptual illustration of the thesis — not measured data.

This paper presents IQE Labs research perspectives, proposed assurance methods and future experiments. It does not report original market survey results, validated client outcomes or measured forecasts.

EXECUTIVE SUMMARY

More artifacts are being generated. Can they be independently trusted?

AI-assisted coding and agentic workflows are accelerating software production, but assurance practices are not necessarily advancing at the same pace.



The central issue is not whether AI can generate more artifacts; it is whether teams can **independently evaluate correctness, risk, behavior and accountability** across the software lifecycle.

Our review highlights five connected market issues

01 Code generation outpacing independent verification

02 AI-generated tests inheriting defects or gaps

03 Agent behavior extending beyond final outputs

04 Evaluation and production monitoring remaining immature

05 Governance increasingly emphasizing evidence and oversight

IQE LABS POSITION

From test-execution volume → to risk-proportionate assurance.

Quality Engineering should evolve toward risk-proportionate assurance of **outputs, execution paths, changes, and operating conditions**.

This is a proposed engineering direction, not a demonstrated universal solution.

1 IQE LABS RESEARCH PERSPECTIVE

Emerging AI Quality Engineering challenges

IQE Labs identifies five engineering challenges for organizations adopting AI-assisted development and agentic systems. These are qualitative research observations, not findings from an original IQE Labs market survey.

01

AI-generated code reliability

Faster generation can increase verification demand. Independent checks remain necessary for correctness and business-critical risks.

02

AI-generated test relevance

More generated tests do not necessarily mean stronger risk coverage. Human review should assess omissions, invalid cases and assumptions.

03

Agent action-path assurance

Final-response testing may miss unsafe tool calls, permission misuse and side effects. Evaluation should include intermediate actions.

04

Lifecycle evaluation and monitoring

Changes to models, prompts, data and tools can alter system behavior. Teams need repeatable evaluations and risk-based monitoring.

05

Evidence-led governance

Policies alone cannot demonstrate quality. Traceable evaluations, exceptions, approvals and accountable decisions strengthen assurance.

IQE LABS RESEARCH PERSPECTIVE | NOVEMBER 2026

These observations inform proposed assurance research; experimental validation and measured results remain future work.

2 MARKET ISSUES

Five market issues requiring QE attention

Each issue is a place where generation capacity can grow faster than the ability to verify it.

01 Verification debt

Generated code and tests can increase faster than review, test design and traceability capacity. A passing generated test may reflect a shared misunderstanding between generator and verifier. Independent, risk-based checks are needed.

02 Test quantity versus test relevance

AI-generated test suites may expand quickly while omitting business invariants, negative scenarios, authorization boundaries or non-functional risks. Teams need measurable coverage of risks, not merely counts of generated tests.

03 Agent action-path risk

Agentic systems can retrieve data, call tools, alter records and trigger downstream actions. Checking only the final response can miss unsafe intermediate behavior, unauthorized tool use or unintended side effects.

Where output-only testing looks — and where risk can hide

Illustrative agent run



04 Evaluation and monitoring gaps

Pre-release benchmarks alone cannot capture changes in prompts, models, data, tools or user behavior. Deployed systems need monitoring, periodic re-evaluation and documented exceptions.

05 Governance without actionable evidence

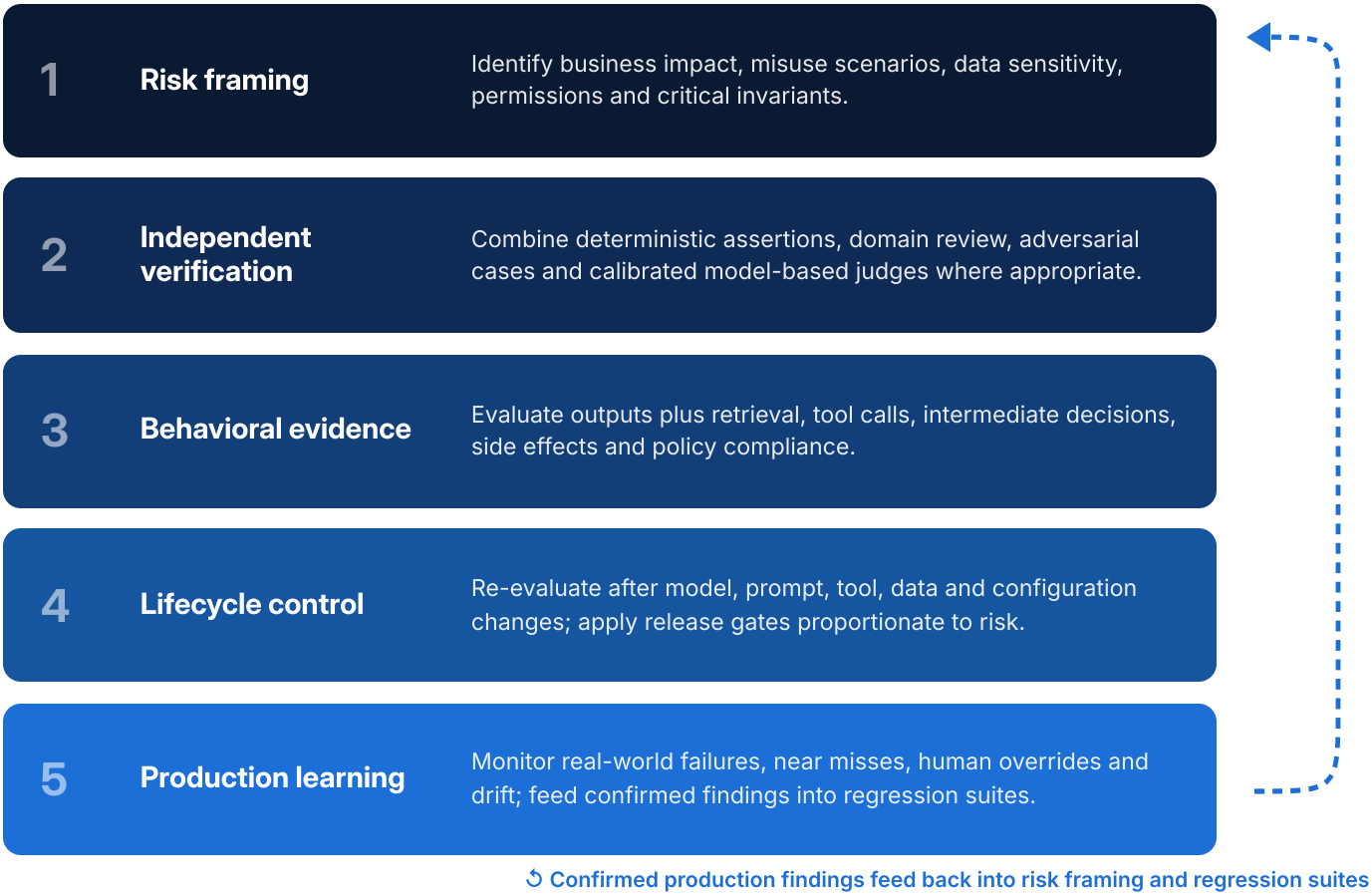
Policy statements alone cannot demonstrate operational control. Teams need versioned evaluation criteria, action logs, human review triggers, exception records and release decisions tied to risk.

3 PROPOSED MODEL

The IQE Labs assurance model

Proposed research model | IQE Labs Research | November 2026

Five assurance layers proposed for evaluation and future experimentation. These are working research concepts, not a new certification or replacement for established security and risk standards.



What changes in practice

ASSURANCE OF...	PRIMARILY ADDRESSED BY LAYER
Outputs	2 · Independent verification
Execution paths	3 · Behavioral evidence
Changes	4 · Lifecycle control
Operating conditions	5 · Production learning — all scoped by 1 · Risk framing

4 RESEARCH WORKSTREAMS

IQE Labs Research Priorities: 2026–2027

Proposed research workstreams — not claims that the experiments have been completed or independently validated.

AI–human assurance

Extend the AI-Human Quality Assurance Protocol (AHQAP) with review rubrics for generated test cases, code and evaluation artifacts.

Agentic QE evaluation

Compare output-only testing with output-plus-action-path assessment using reproducible scenarios, policy constraints and side-effect checks.

RAG quality evaluation

Separate retrieval relevance, groundedness, answer usefulness and unsupported-claim rates; include a human adjudication sample.

QE maturity assessment

Examine how organizations assess automation reliability, quality evidence, AI readiness and operational governance across nine capability dimensions.

Evidence architecture

Explore traceable release-readiness records linking risks, evaluations, model/tool versions, exceptions and accountable decisions.

5 EXPERIMENTAL AGENDA

Proposed 90-day plan

DAYS 1–30

Define & baseline

- Select a bounded public or synthetic use case
- Define risks and success criteria
- Establish a small golden evaluation set
- Document model and tool versions
- Measure a baseline

DAYS 31–60

Compare assurance methods

- Deterministic checks
- LLM-assisted grading with human calibration
- Output-only vs. action-path checks
- Track false positives, missed failures, repeatability, review effort, latency and cost

DAYS 61–90

Review & publish

- Replicate runs and assess limitations
- Publish methodology and negative findings alongside positive findings
- Separate evidence from hypotheses

Publication rule: no claims of improvement should be published without measured results.

6 PRACTICAL ACTIONS

Six steps for QE leaders

Actions teams can start now, independent of tooling choice.

1. **Inventory** AI-generated artifacts and agent capabilities, including their permissions and data access.
2. **Select** a small set of business-critical risks and specify independent pass/fail evidence.
3. **Introduce review checkpoints** for generated tests, code changes and high-impact agent actions.
4. **Treat** prompt, model, retrieval source, tool and policy changes as testable configuration changes.
5. **Add** adversarial and negative tests; monitor production behavior and human overrides.
6. **Track** quality outcomes and evidence sufficiency rather than test volume alone.

7 LIMITATIONS & CLAIM DISCIPLINE

How to read this paper

- These are qualitative IQE Labs research perspectives, not results from a representative market survey or completed experimental study.
- Proposed controls require context-specific evaluation and independent validation.
- Regulatory applicability depends on sector, geography, system classification and evolving rules.
- IQE Labs has not yet published experimental results validating the proposed assurance layers.

This paper is a research perspective, not legal, security, compliance or investment advice. IQE Labs does not claim to:

- ✕ Eliminate hallucinations
- ✕ Guarantee AI trustworthiness
- ✕ Certify regulatory compliance



About IQE LABS

IQE Labs is an emerging Quality Engineering research and consulting practice exploring Intelligent QE, Agentic QE, AI-system evaluation, automation engineering, and quality evidence. Our approach is to distinguish practitioner insight and proposed methods from independently measured results.

Our research combines practitioner reasoning, proposed engineering methods and transparent hypotheses. We distinguish research directions from measured results.

Website: iqelabs.ca | Series: Research White Papers | Version 1.0 | Published 6 November 2026