pcloudy

The AI Quality Infrastructure Benchmark Report 2026

Surveyed 100+ QA leaders on building AI-native testing tools, models, context, and compliance challenges.

100+	40%	7
QA & Engineering Leaders	Director-Level or above	Industries covered



FOREWORD

Why We Built This

AI changed how fast QA teams can create tests years before it changed how fast they can trust, run, or govern them. That gap is the story this benchmark set out to measure.

Every QA leader we talk to is living the same contradiction. Their teams write more tests than ever, faster than ever, with AI doing most of the typing. And somehow, releases still feel like a gamble.

We kept hearing versions of the same story this year. A QA director who trusts her test suite completely, except she can only run 60% of it before a release ships. A security team that blocks every public AI API by policy, while half the engineering org quietly uses one anyway. A test engineer who spent more time this quarter fixing AI-written tests than a human would have spent writing them from scratch.

So instead of guessing how widespread it was, we asked. We talked to more than 100 QA and engineering leaders across banking, healthcare, gaming, media and technology, on four different continents, about what is actually happening inside their organizations right now. Not vendor pitches. Not analyst projections. The picture on the ground. This report is what they told us.



EXECUTIVE SUMMARY

Four Findings That Define 2026

More than 100 QA and engineering leaders — from test engineers to VPs of engineering across banking, healthcare, hospitality, media and logistics answered this survey, spanning India, the United States, Australia, Egypt, Thailand and the Middle East. Most QA organizations have adopted AI to help write tests. Few have restructured around it. Four findings stand out.

1

Test creation is hybrid, not autonomous yet.
59% of teams pair open-source frameworks with AI writing assistance. Only 10% run a dedicated commercial AI test generation tool in production. The AI tools teams reach for most are general-purpose coding assistants — Claude, ChatGPT, Cursor, Claude Code — wired into existing pipelines rather than built for testing.

2

Test execution is a bottleneck.
Only 14% of teams can execute more than 90% of their automated suite inside a standard release window. Almost half execute less than 70%. AI has made writing tests fast; running those generated tests at the pace releases demand is the problem teams haven't solved.

3

Test confidence is outrunning evidence.
93% of teams describe themselves as extremely or somewhat confident in release quality. That confidence holds even though more than half cannot execute the majority of their own test suite before shipping.

4

Compliance sets the agenda, but practice hasn't caught up.
67% cite data sovereignty and 58% cite security policy as reasons for their current AI deployment approach, both ahead of raw model performance at 54%. Yet 45% of teams send test data straight to public frontier model APIs, and 80% say their organization restricts or prohibits that practice to some degree.

WHERE THE NEXT BUDGET GOES
Agent orchestration and context management is the single largest planned investment for the next 12 months, cited by 69% of respondents — ahead of AI test generation, execution infrastructure, and observability. Teams are not asking for more AI. They're asking for AI that remembers, coordinates, and runs somewhere their compliance team will sign off on.

BY THE NUMBERS

The Report in Eight Numbers

59% of Organizations



Use Hybrid test creation

OPEN-SOURCE + AI ASSISTANCE

14% of Organizations



Clear 90% of the suite

INSIDE A RELEASE WINDOW

93% of Organizations



are Confident at ship time

DESPITE PARTIAL EXECUTION

80% of Organizations



Restrict public AI APIs

POLICY-LEVEL LIMITS

71% of Organizations



Hand-maintain prompts

NO PERSISTENT CONTEXT LAYER

54% of Organizations



Spend 10–20% on upkeep

OF EVERY RELEASE CYCLE

38% of Organizations



Run in-house device labs

ON-PREM OR HYBRID

69% of Organizations



Believe in Funding agent orchestration

NEXT 12 MONTHS

SOURCE: PCLOUDY STATE OF AI QA SURVEY 2026 · 100+ RESPONDENTS

Table of Content



About This Study 01

Test Creation Has Gone Hybrid, Not Autonomous Yet 02

The Execution Bottleneck Nobody's Pricing In 03

The Confidence Paradox 04

Compliance Sets The Agenda - Practice Hasn't Caught Up 05

Context Management Is The Unbuilt Layer 06

The Maintenance Tax 08

Execution Infrastructure Is Spliting Along Compliance Lines 08

Where The Money Is Going Next 09

Voices From The Field 10

What This Means For QA Leaders 11

METHOD

About This Study

This benchmark report is based on a survey of 100+ QA and engineering professionals from banking, fintech, healthcare, insurance, retail, eCommerce and adjacent sectors. Respondents were not selected for any particular point of view: they represent working QA organizations at varying stages of AI adoption, from teams still mostly manual to teams running multi-agent orchestration in production.

Who answered

Seniority skewed toward decision-makers: roughly two in five respondents hold director, VP, AVP, or founder-level titles, with the remainder split between QA managers and leads, and individual-contributor engineers.

Where they work

Respondents span banking and financial services, technology and IT services, healthcare, gaming and hospitality, media, security, and logistics. Organizations range from regional credit unions to global banks and Fortune-scale enterprises, based in India, the United States, Australia, Egypt, Thailand, and the Middle East.

How to read this report

Not every question was answered by every respondent; response counts are noted where relevant. Open-ended responses are quoted with light edits for clarity. Percentages are rounded and may not sum to 100 where a question allowed multiple answers.



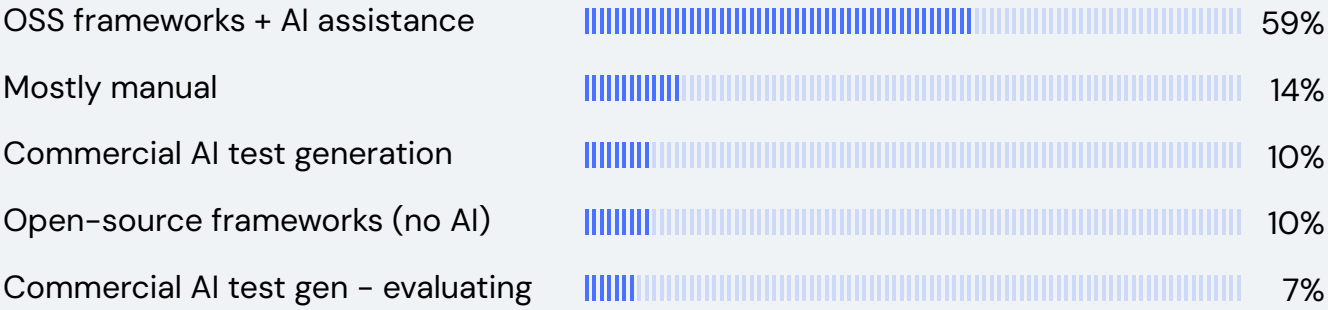
CHAPTER 1 OF 8 · TEST CREATION

Test Creation Has Gone Hybrid, Not Autonomous Yet

Manual test creation is nearly gone. Only 14% of teams describe their primary approach as mostly manual — but the tools that replaced it are hybrid, not autonomous.

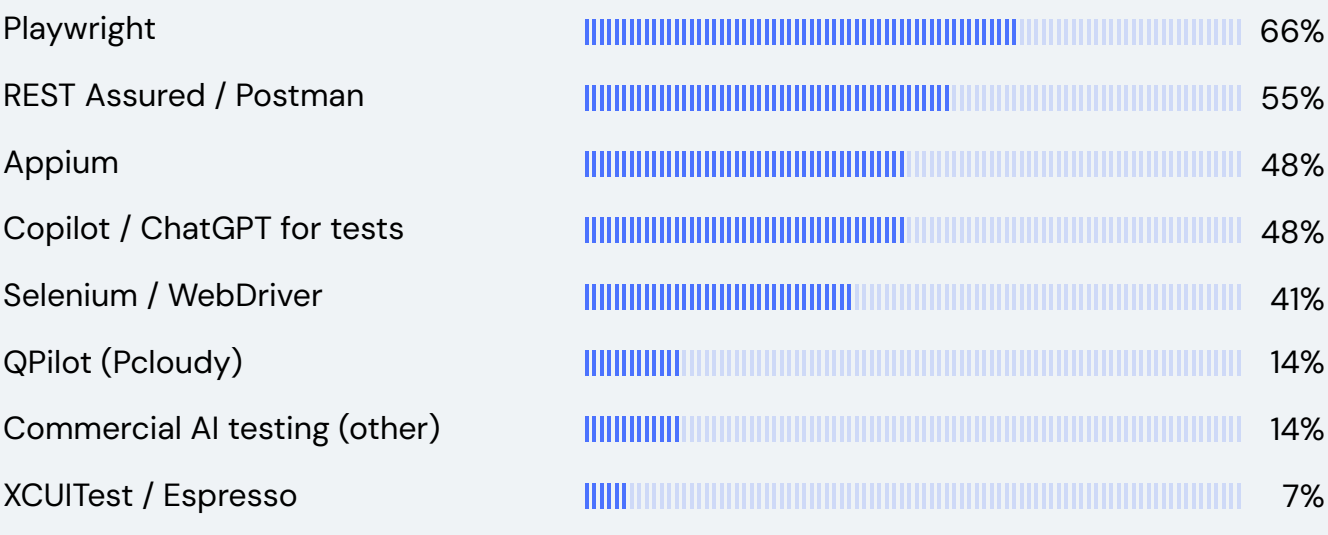
59% of teams use open-source automation frameworks with AI writing assistance layered on top, and just 10% have moved to commercial AI test generation in production, with another 7% still evaluating it.

Q: Which best describes your team's primary test creation approach today?



The tool list confirms the pattern. Playwright leads at 66% adoption, followed by REST Assured/Postman at 55%, Appium at 48% and Selenium at 41%. General-purpose AI coding assistants like GitHub Copilot and ChatGPT are already used for test writing by 48% of teams, matching Appium's reach. Purpose-built AI testing tools — QPilot and other commercial AI testing tools — sit at 14% each.

Q: Which automation frameworks or AI testing tools does your team currently use? Select all that apply.



FOOD FOR THOUGHT

The AI test-creation layer is being assembled, not bought. Any platform decision now has to co-exist with Playwright, Postman and a general-purpose coding assistant already in the pipeline.

The 'other' responses tell a more interesting story - Claude, Claude Code, Cursor and Thumba.ai all appear alongside a hand-built hybrid framework and a UiPath - plus - Playwright - agents combination. Teams are not waiting for a dedicated AI testing platform. They are assembling one out of whatever general-purpose AI tool is already approved for developers and stitching it into frameworks they already run.

CHAPTER 2 OF 8 · EXECUTION

The Execution Bottleneck Nobody's Pricing In

Even at modest test volumes, fewer than one in five teams can execute more than 90% of their automated suite inside a standard release window.

Test volume is modest for most respondents: 45% run between 500 and 2,000 test cases per release cycle, and only 7% run more than 10,000. That makes what happens next more striking.

Q: What proportion of your automated tests can your team execute within a release window?



Combine the bottom two rows and 24% of teams execute less than 40% of their own automated suite before a release ships. This is not a test-writing problem — AI has already made test authoring fast. It is a throughput problem: parallel device and browser capacity, queueing, and infrastructure that cannot clear the backlog AI-assisted authoring now produces faster than ever.

FOOD FOR THOUGHT

Execution capacity, not authoring speed, is now the binding constraint on release confidence. Every hour saved writing tests is returned to the queue unless parallel capacity grows with it.

CHAPTER 3 OF 8 · CONFIDENCE

The Confidence Paradox

93% of respondents describe themselves as extremely or somewhat confident in release quality. Not one selected 'somewhat unconfident' or 'not confident'.

Put the execution numbers next to how teams feel about release quality and a gap opens: 45% are extremely confident and 48% somewhat confident at the point code ships.

Q: How confident is your team in release quality at the point it goes to production?

Somewhat confident	48%
Extremely confident	45%
Uncertain	7%
Somewhat unconfident / not	0%

Read this against Chapter 2 — where 52% of teams execute less than 70% of their automated suite before release — and confidence starts to look decoupled from coverage. One respondent put a name to the gap directly:

> "AI helped us solve 'more tests, faster' well before it helped us solve 'trust the results, faster'. Closing that gap is where we're focused next."



Sankaranandhan Sundaresan
Digital Engineering Lead Engineer | MGM Resorts International

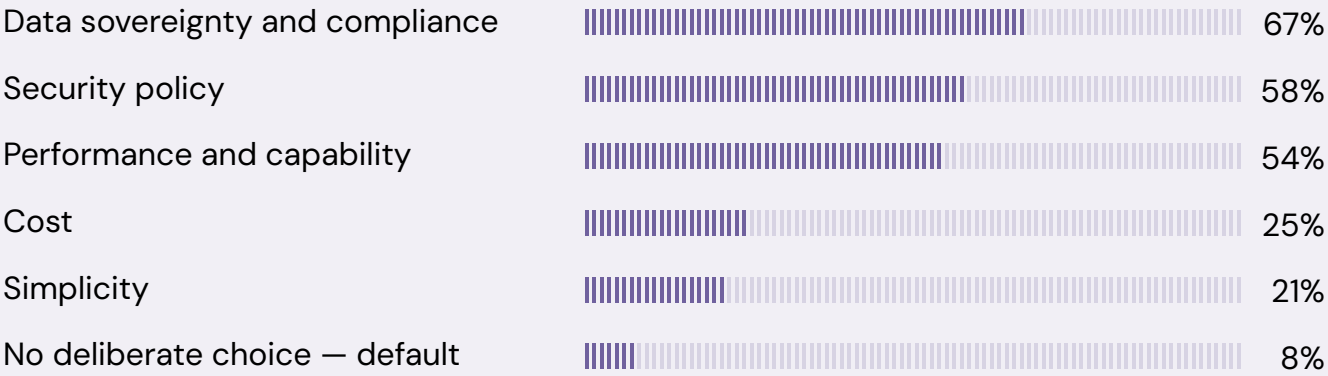
This is not a claim that release confidence is unfounded. Manual checks, staged rollouts and production monitoring all sit outside this survey's scope. But it is a signal worth taking seriously: teams are shipping on confidence built from partial execution, and few have instrumented the difference between the two.

CHAPTER 4 OF 8 · COMPLIANCE

Compliance Sets the Agenda - Practice Hasn't Caught Up

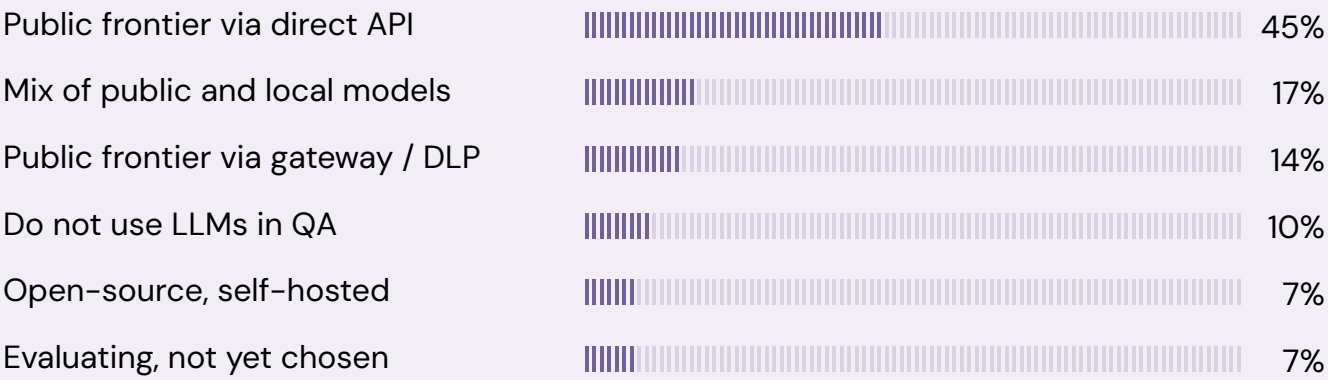
67% cite data sovereignty and 58% cite security policy as the reason for their AI deployment approach — both ahead of raw model capability at 54%.

Q: Why did your team choose its current AI model deployment approach? (multi-select)



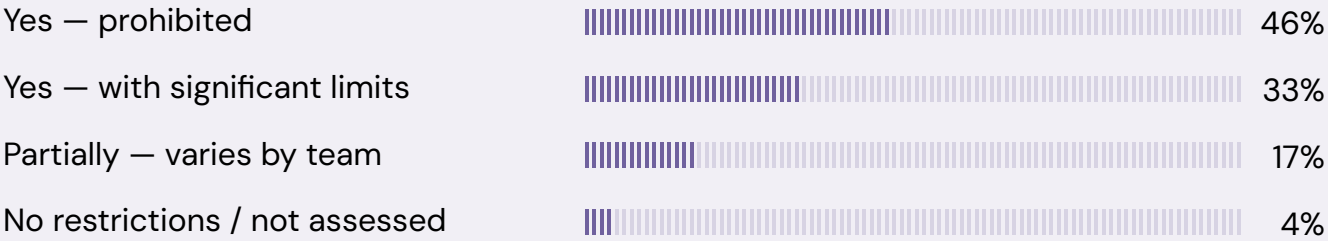
Yet the deployment mix doesn't fully reflect that stated priority. 45% of teams call public frontier models — GPT-4/o, Claude, Gemini — directly via the vendor's API, with no gateway or DLP layer in between. 14% route through a managed gateway, only 7% self-host open-source models, and a further 17% mix public and local models by use case.

Q: Which AI model(s) does your team team use within your QA or testing workflow?



The tension sharpens on direct policy: almost 80% of teams say their organization restricts sending test data, application flows or business logic to a public AI API.

Q: Does policy restrict sending test data or business logic to a public AI API?



One QA leader with regional banking oversight named the specific friction point directly: “Approval by Central Bank” was cited, unprompted, as the single biggest unresolved challenge in using AI effectively for testing. This is not teams ignoring policy. It is a market where the AI tooling that meets the compliance bar — private deployment, on-prem execution, auditable data handling — has not yet matched the convenience of the public API teams already know how to use.

FOOD FOR THOUGHT

Policy says one thing; the daily workflow says another. 80% restrict public AI APIs, and 45% use one directly. That gap closes only when private and on-premise options are as convenient as the public API.

CHAPTER 5 OF 8 · CONTEXT

Context Management Is the Unbuilt Layer

71% of teams still hand-maintain prompts. Only 25% have a structured memory agent that persists context across sessions.

Getting an AI model to reason accurately about a specific application requires feeding it context — requirements, business rules, prior test history. Most teams still do this by hand. Retrieval-Augmented Generation, which pulls context automatically at inference time, is used by 50%.

Q: How does your team manage context and knowledge for AI agents? (multi-select)

Prompt engineering (manual)	71%
Retrieval-Augmented Generation	50%
Fine-tuned / domain-adapted	33%
Structured memory agent	25%
None — immediate input only	4%

That manual-first posture shows up again in how teams structure AI agents. 34% use AI tools but not in an agentic pattern at all. Only 29% have reached multi-agent orchestration, where agents coordinate and share state as a system.

Q: How does your team currently use AI agents in the testing workflow?

AI tools used, but not agentic	34%
Multi-agent orchestration	29%
Single agent, isolated task	29%
Sequential agent handoff	8%
No AI agents in testing	0%

Domain and business context, not model capability, is what respondents named most often as their biggest unresolved challenge: “getting business context and real application usage”, “domain knowledge and product knowledge”, “lack of proper, full descriptive inputs”. Better prompting only goes so far against that gap. What most teams describe wanting is closer to what the leader below described already living inside:

“AI now spans our whole QA loop — authoring the plan, driving the browser, healing broken locators, publishing results and keeping the suites green in CI.”



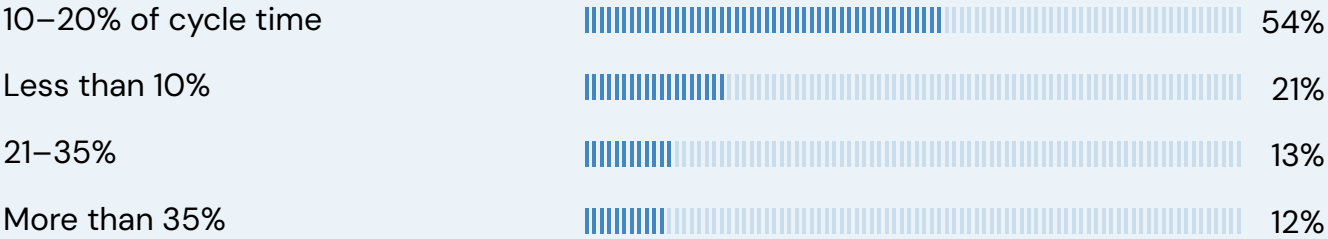
Archana Bangar
Lead QA Automation Engineer | Blend

CHAPTER 6 OF 8 · MAINTENANCE

The Maintenance Tax

54% of teams spend 10–20% of every release cycle keeping existing automated tests alive rather than extending coverage.

Q: What share of your QA team's time goes toward maintaining or fixing existing tests?



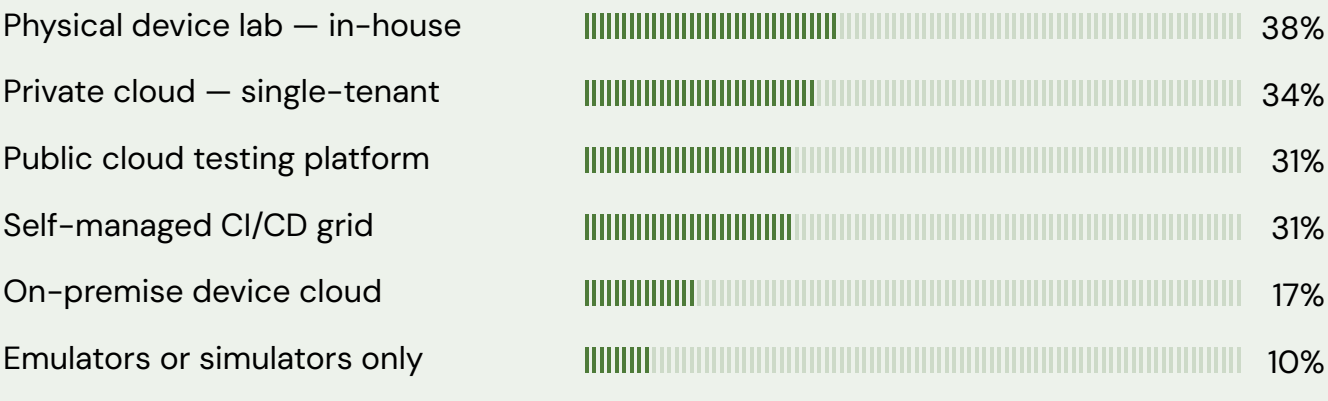
For the 12% burning over a third of every cycle on maintenance, AI-assisted authoring adds new tests faster than the team can keep old ones alive — a treadmill effect that shows up directly in the open-ended responses: "Script development is challenged for complex scenarios; sometimes it changes the existing working script as well."

CHAPTER 7 OF 8 · INFRASTRUCTURE

Execution Infrastructure Is Splitting Along Compliance Lines

No single execution model dominates. In-house device labs lead at 38%, private single-tenant cloud follows at 34%, public cloud at 31%.

Q: What execution infrastructure does your team primarily use? (multi-select)



Combined, private single-tenant and on-premise options outweigh public cloud platforms — a split that tracks directly with the compliance pressure documented in Chapter 4. On-premise device clouds give organizations full cloud capability inside their own firewall. Execution infrastructure decisions increasingly follow the same sovereignty logic as model deployment decisions: where the data can legally sit, not just what runs fastest.

CHAPTER 8 OF 8 · INVESTMENT

Where the Money Is Going Next

Agent orchestration and context management ranks first for the next 12 months, by a wide margin, at 69%.

Q: Which areas of QA infrastructure will you invest in over the next 12 months? (multi-select)



This lines up with every gap this report has surfaced. Chapter 5 showed context management is still mostly manual. Chapter 3 showed confidence isn't matched by execution evidence — which is what observability and failure analysis exist to fix. Chapter 2 showed execution capacity is the binding constraint on how much of the suite actually runs. Investment plans are pointed, almost exactly, at the weakest layer in the stack teams described building over the previous seven chapters.

INDUSTRY INSIGHTS

Voices from the Field

"We treat Generative AI as a hyper-efficient assistant that does 80% of the heavy lifting in seconds, while relying on human critical thinking for the remaining 20% that actually guarantees quality."



Jayprakash Mourya
AVP | DBS

"Get the QA process fundamentals right first. AI can accelerate a good process, but it cannot fix a broken one."



George Ukkuru
Practice Director | Speridian Technologies

"AI can generate tests, but transforming undocumented legacy back-end applications into reliable, maintainable regression suites with realistic test data remains a significant challenge."



Mohamed Vaquas Umair
VP, Head of Engineering | Velocix India Solutions

"AI industry focus is more on scripting, while analysis and validation lags a lot."



Mohamed Abdelkader
Senior QA Manager | MBC Group

"We didn't wait for the right approach — we explored, experimented, and put it into our regular workflow."



Deepankumar
QA Manager | SecureW2 India Private Limited

"Never use AI for autonomous execution, always keep human in the loop to make sure humans are in control. We never know what automation could lead to otherwise."



Sandeep Garud
CEO | SpurQLabs

"We're redefining the STLC through agentic AI, accelerating releases while making QA smarter, more efficient, and more impactful."



Priyanka Ajith
QA Manager | Tricog Health India

"AI is helping improve test coverage and accelerate debugging. We are looking forward to leveraging agent orchestration to autonomously execute tests, reducing manual effort and minimizing the need for human intervention."



Sarala Ponnaganti
Lead Software Quality Assurance Engineer | Suncoast Credit Union

"Looking to improve Leaps and Bounds in using AI-powered QA demonstrating cost reduction by improving on release quality and cutting down re-work."



Giridharan Kalpathi Sukavaneswaran
Head Digital QA and Banking, ME & KSA | Horváth Partners

CONCLUSION

What This Means for QA Leaders

An industry past the AI pilot stage, but short of a stable operating model.
Test creation is AI-assisted almost everywhere. What's missing is the layer that turns AI-assisted authoring into something a compliance officer, a release manager and an engineering VP can all trust at the same time. Three requirements come through consistently across the data:

Execution capacity that matches AI-accelerated authoring.
Writing tests faster than the team can run them just moves the bottleneck downstream — Chapter 2 shows almost a quarter of teams already can't clear 40% of their suite in a release window.

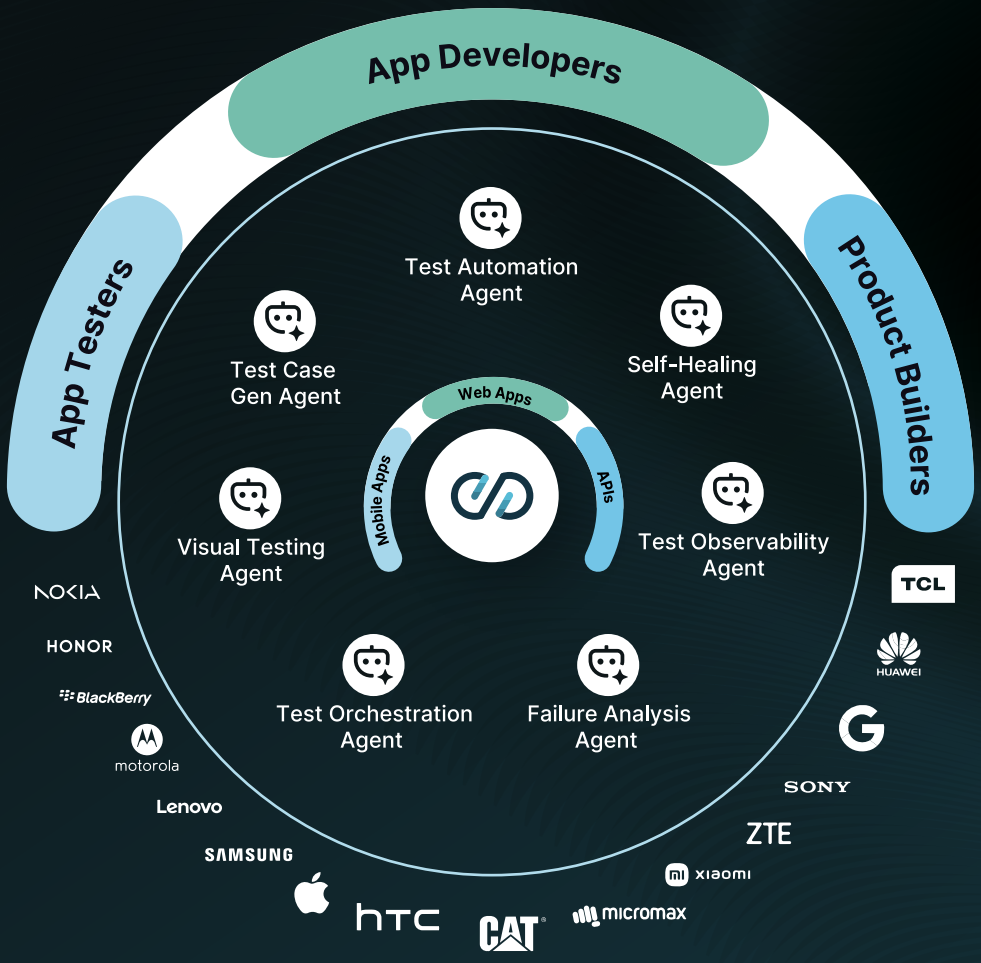
Data handling that satisfies compliance without a manual workaround.
The gap between stated policy and actual API usage will not close by asking teams to try harder; it closes when private and on-premise AI and execution options are as convenient as the public API.

Persistent context and agent orchestration that is engineered, not improvised.
This is both the least mature layer today and the top investment priority for the year ahead. This is the specific gap that purpose-built AI-native QA platforms are built to close: combining AI test generation, real-device execution at scale, and agent orchestration with context that persists — deployable in the private or on-premise configurations compliance teams require. Pcloudy's QPilot and Device Cloud, including on-premise deployment through Lab-in-a-Box, are built to suit this requirement set: AI agents that retain context across a testing session, execution infrastructure that can run inside a customer's own environment, and reporting built to answer confidence with evidence rather than sentiment. Teams planning their next 12 months are describing exactly this combination. The open question is whether they have yet found a single platform that delivers it.

CHECK HOW RELEASE READY IS YOUR ORGANIZATION
Benchmark your stack in 30 minutes
A read-out of this benchmark against your own execution coverage, model deployment posture and context maturity. — pcloudy.com

pcloudy

The Real Device Cloud with Intelligence



Most enterprise QA teams are stuck choosing between speed and security. Public clouds get flagged by InfoSec. Building an in-house lab takes months and costs a fortune. And testing on too few devices means bugs reach production before your team catches them.

Pcloudy fixes that. Test mobile apps, web apps and APIs on 5000+ real devices and browsers — with 60+ real-device performance metrics built in.

Enterprise security. Zero compromise. Public cloud, private cloud, on-premise, or hybrid — air-gap compatible, data residency compliant, and deployed inside your environment in weeks. Not months.

And with AI Agents handling everything from test authoring to failure analysis, your team stops firefighting and starts shipping with confidence. Trusted by 500+ enterprises across industries globally.

pcloudy

+91-762 499 9076 | info@pcloudy.com

Visit our Website

Contact us