aikido

State of AI in Pentesting

2026 REPORT

Contents

INTRODUCTION

Why this matters now

AI is accelerating software delivery and introducing new types of risks. But security testing was designed for a slower world. This report captures how this is impacting organizations and what leaders are doing about it.

Sapio Research, on behalf of Aikido Security, surveyed 200 security leaders (CISO or equivalent) and 200 senior engineering leaders (VPs of Engineering, CTOs, or equivalent).

All respondents work at cloud-native organizations and are actively involved in security decision-making or software delivery at a senior level.

Each participant has direct visibility into how their organization approaches security testing, risk, and engineering velocity.

Key findings

KEY FINDING #1

AI is already causing security problems

76%

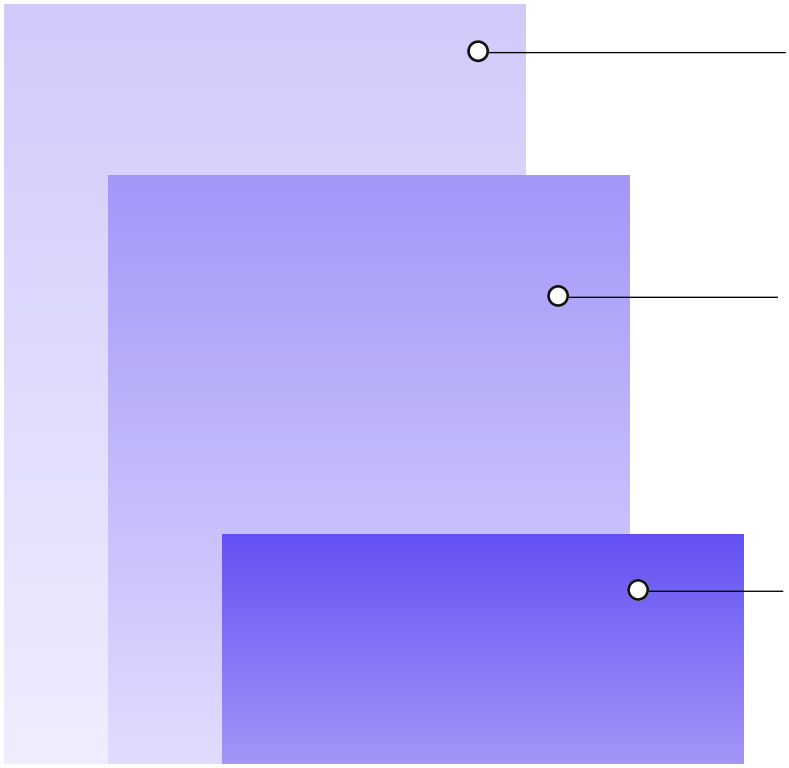
Have intervened to stop or restrict AI-driven behavior

71%

Say AI has made security incidents harder to detect, investigate, or fix

KEY FINDING #2

Security testing can't keep pace with software delivery



76%

Deploy significant changes weekly or faster

51%

Say testing delays releases or requires teams to ship while accepting known risk.

Only 21%

Validate security on every release

KEY FINDING #3

The time between tests create the biggest blind spot

79%

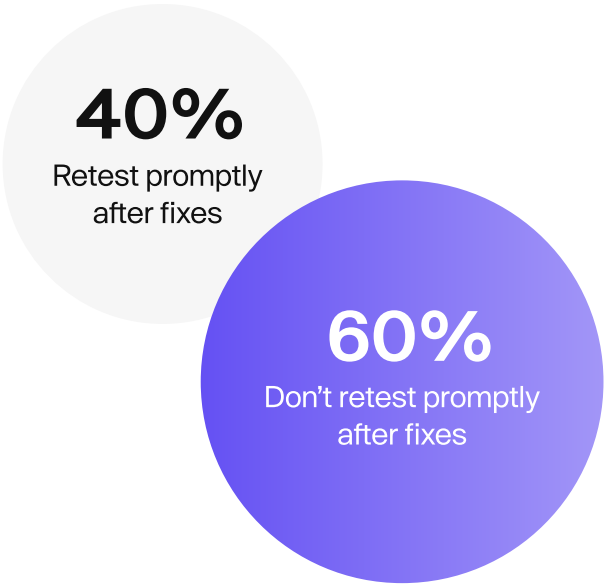
Are concerned about **missing vulnerabilities** introduced between scheduled tests

48%

Say findings are already **outdated** when they arrive

KEY FINDING #4

Most fixes aren't verified



KEY FINDING #5

Manual pentesting has hit its limit

51%

Say logic flaws, broken access controls, and multi-step vulnerabilities are missed always or often

92%

That number rises for teams shipping daily or faster

52%

Lack full visibility into what was actually tested

KEY FINDING #6

Trust breaks when findings are wrong, not when they're missing

TRUST BREAKERS

42% Say incorrect or hallucinated findings are the biggest trust breaker

Say missing vulnerabilities are the biggest trust breaker **32%**

AUDITS

40% Are more confident defending reproducible evidence in an audit

Are more confident defending provider reputation in an audit **26%**

BOTTLENECKS

40% Say the top bottleneck in acting on findings is validation

Say the top bottleneck in acting on findings is volume **14%**

KEY FINDING #7

Leaders want security testing to happen more often

69%

Would validate security on every release or at least quarterly if constraints disappeared

Speed **(39%)** and on-demand testing **(39%)** rank above cost **(27%)**

Explainability and deep logic-level testing both rank highly **(36%)**

Main report

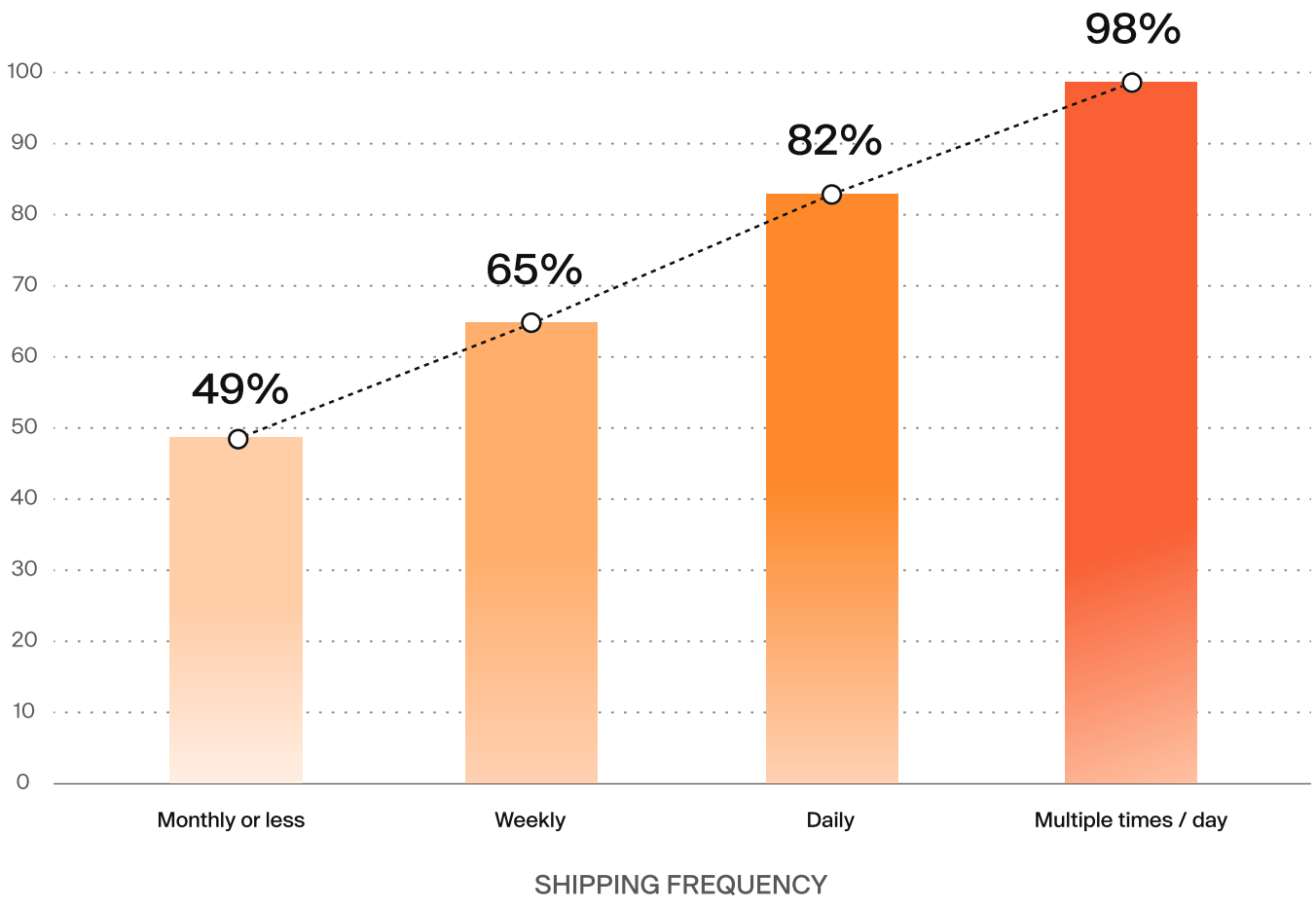
Impact of AI on security

76% of organizations have had to intervene to stop, restrict, or roll back AI-driven behavior in the past 12 months.

IMPACT OF AI ON SECURITY

For teams shipping multiple times daily, 98% have had to stop, restrict, or roll back AI-driven behavior in the past year.

Intervention is a sign that AI is regularly causing issues for organizations, but that they are paying attention. The concern is what happens in the environments where they aren't.



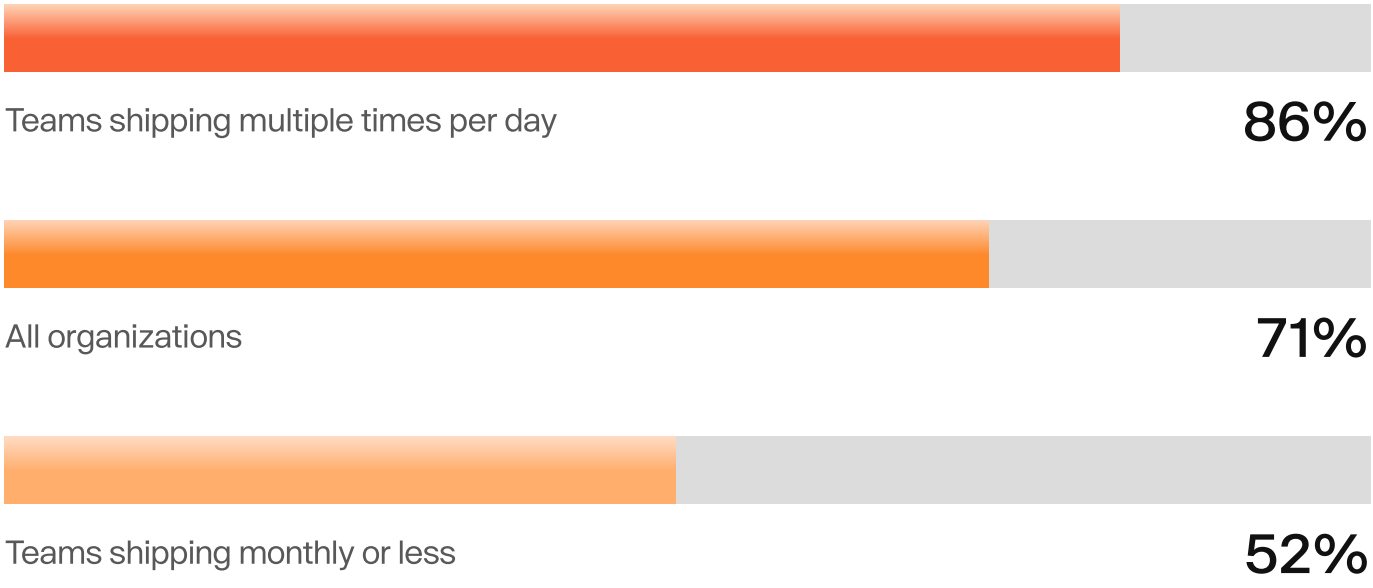
In the past 12 months, has your organization had to stop, restrict, or roll back AI-driven behavior due to security or safety concerns?

IMPACT OF AI ON SECURITY

AI is making security incidents harder to handle.

AI is moving faster into the software development lifecycle than the security practices around it. Teams are still building the muscle to detect when AI-driven behavior is the source of a problem, trace what it did, and understand how to contain it. The tooling, the processes, and the institutional knowledge are catching up. But they're not there yet.

Percentage who say they've experienced a security issue in the past year that was harder to detect, investigate, or remediate because of AI



In the past 12 months, has your organization experienced a security issue that was harder to detect, investigate, or fix because of AI or automation?

IMPACT OF AI ON SECURITY

Those who greenlight risky releases aren't the ones held accountable

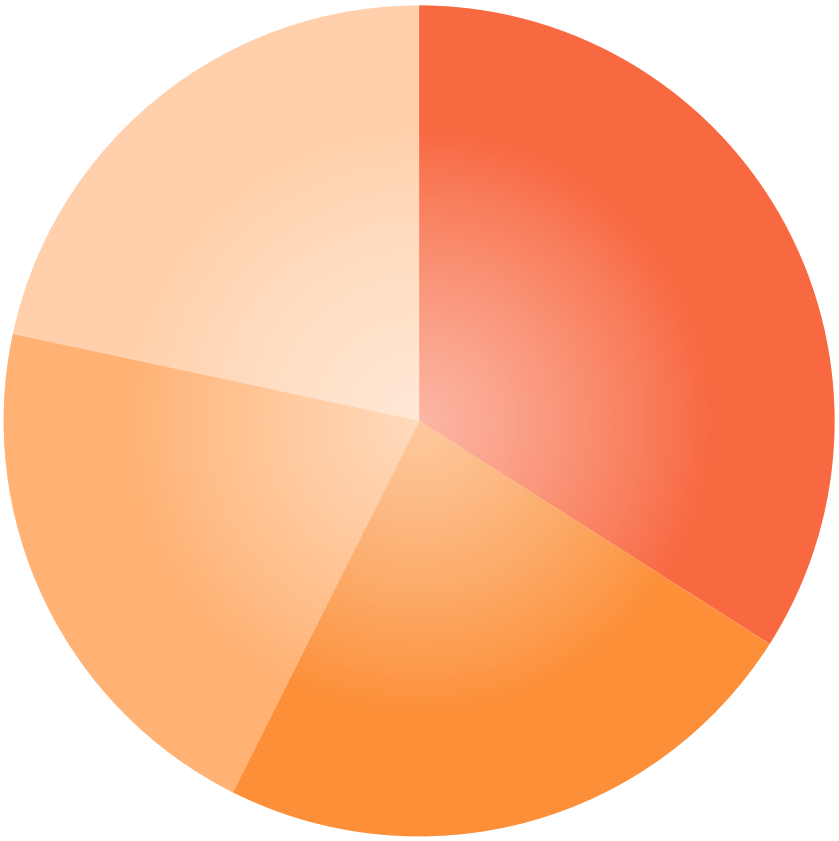
Security risk is accountable by teams that lack decision authority for releases, while the teams that own releases are not clearly accountable. **This suggests a governance gap: release authority and incident accountability often sit in different places.**

Decision ≠ Responsibility

IMPACT OF AI ON SECURITY

Misalignment is the norm, not the exception

Only a third of security teams have both the power to stop a release and the responsibility if it goes wrong. In other cases? They might receive the blame regardless of who made the decision.



34%
Security leadership made the call

23%
Executive leadership made the call

21%
Joint decision between security & engineering

22%
Someone else made the call

? In cases where the security team is accountable for AI issues, who made the call?

Where security testing stands now

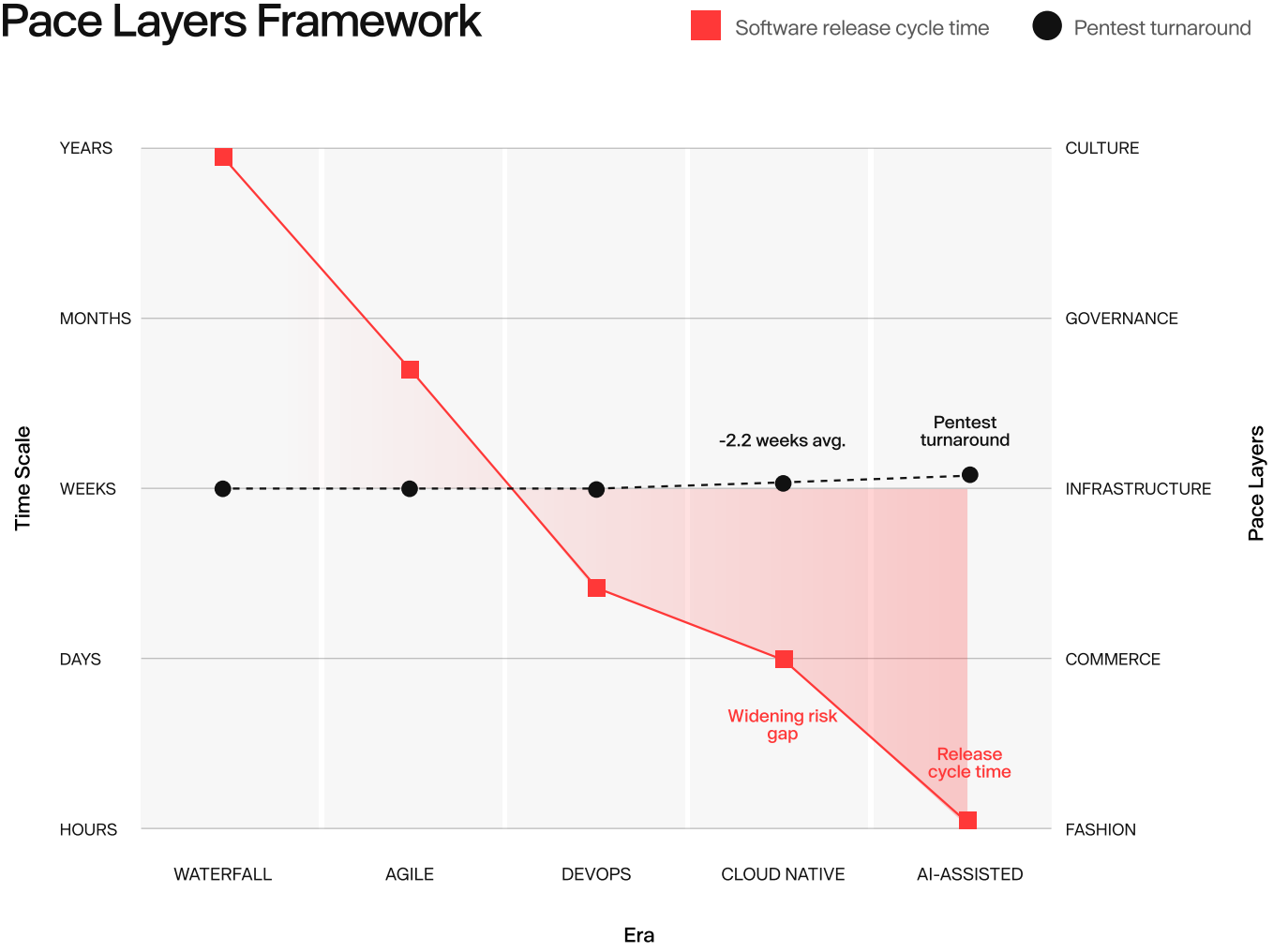
If AI is making systems harder to secure...is our safety net keeping up?

WHERE SECURITY TESTING STANDS NOW

Security validation can't be slower than the software it protects (*and yet it is*)

Stewart Brand's Pace Layers framework shows that different parts of a system naturally move at different speeds. Software delivery has accelerated to the fastest layer (hours). Security validation is stuck three layers behind (weeks). Closing that gap requires a fundamentally different approach to security testing.

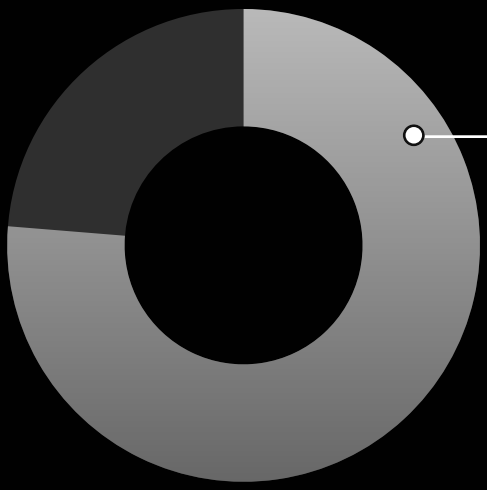
Stewart Brand's Pace Layers Framework



WHERE SECURITY TESTING STANDS NOW

Software changes continuously. Security validation doesn't.

Pentesting was built as a point-in-time exercise, designed to validate a system at a specific moment. Modern software doesn't stay still long enough for that model to work reliably anymore.



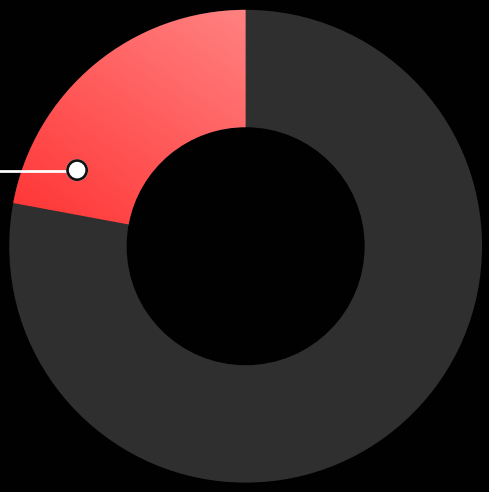
76%

deploy significant production changes weekly or faster

ONLY

21%

validate security on every release



Lovable

“When your users deploy multiple times a day, you can't treat security validation as a periodic checkpoint. The window between tests is where real risk accumulates.”

Igor Andriushchenko
CISO, Lovable

WHERE SECURITY TESTING STANDS NOW

The faster you ship, the more testing slows you down

Nearly two-thirds (64%) of all organizations say security testing timelines negatively impact release decisions by delaying releases or requiring teams to ship while accepting known risk . But for the fastest teams, those shipping multiple times per day, it's near-universal (92%).

The teams with the most to gain from speed are the ones most constrained by testing. As release cadences accelerate, this bottleneck gets worse.

“When you're managing security across multiple regions and teams, you can't scale pentesting by adding more testers. The model breaks down as soon as the organization grows faster than your ability to schedule engagements."

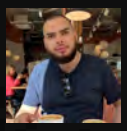
believe. | Yolanda Amorim
Application Security Manager, Believe

Percentage reporting security testing timelines have a larger or very large negative impact on release decisions

92%	80%	54%	45%
Multiple times per day	Daily	Weekly	Monthly or less

Cabinet Office

“By the time a 20-page pentest report lands on a developer's desk, they've already moved on. The code's changed, the priorities have shifted, and now you're asking them to go back and fix something in a system that looks completely different. Nobody wants that conversation."



Walid Mahmoud
DevSecOps Lead, UK Public Sector

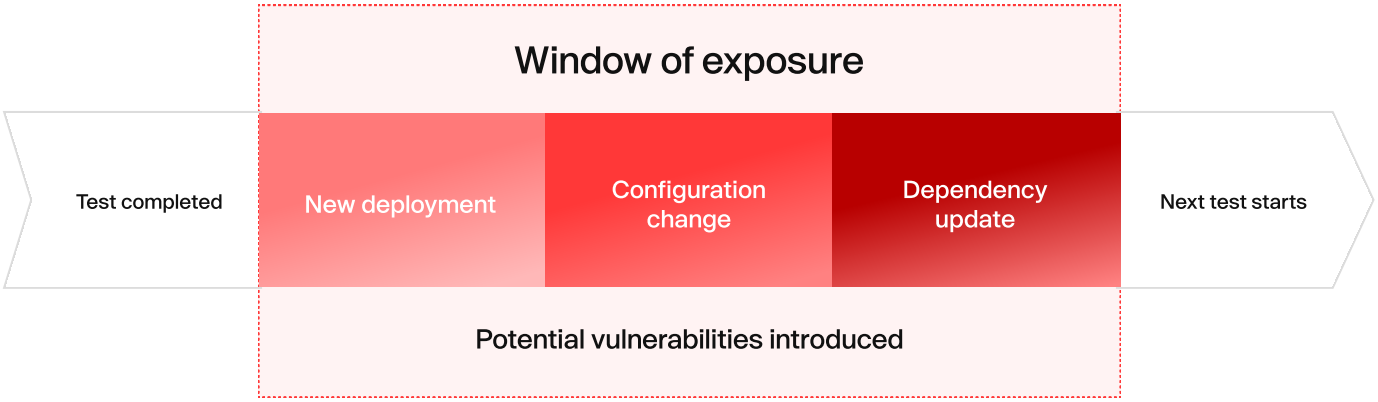
WHERE SECURITY TESTING STANDS NOW

79% are concerned about missing vulnerabilities

Only 4% of companies have no concern that their current testing approach will miss vulnerabilities introduced between scheduled tests.

Modern engineering teams deploy continuously, and every change between tests should be validated. For those organizations, the window between tests is where vulnerabilities can go undetected and increase risk of an attack.

The blind spot between tests



> "Most teams are shipping constantly between pentests. Everyone knows things are going out with problems, but you don't find out until the next engagement comes around. That gap is where the real risk sits."

Omnea | Gavin Williams, Engineering Manager, Omnea

? How concerned are you that your current security testing approach will not detect issues introduced between scheduled tests?

WHERE SECURITY TESTING STANDS NOW

Half of organizations say findings are already outdated when they arrive

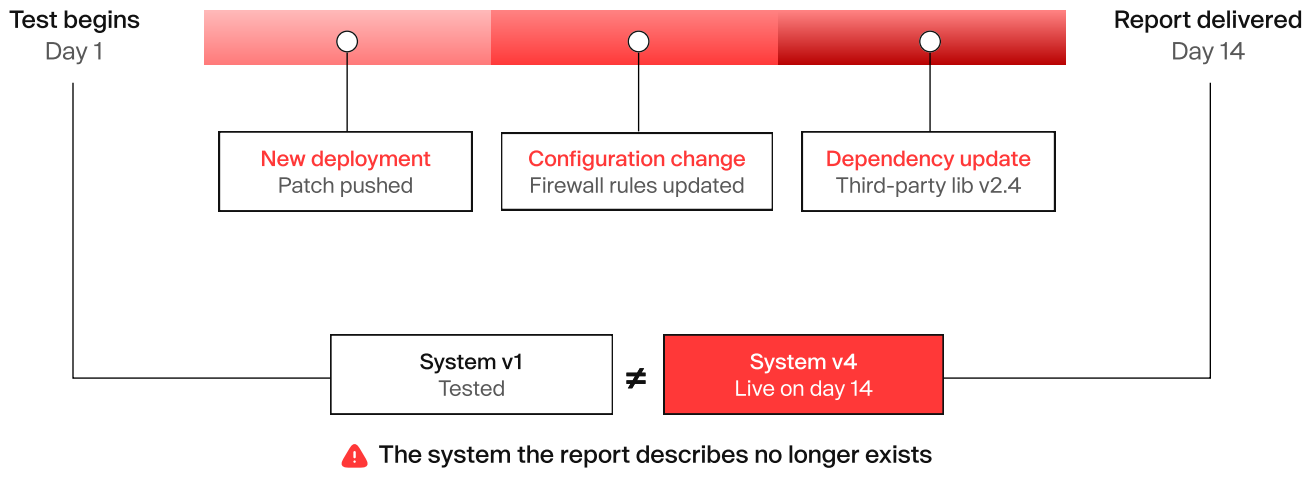
48% of teams say that pentesting findings are always or often delayed by the time they get them.

This problem gets worse for fast moving teams. 84% of teams shipping multiple times per day say pentest findings are already outdated very often or always — vs 26% for teams shipping monthly or less.

"A pentest report that's outdated before the remediation meeting is a compliance artifact, not a security control. Most modern companies should be shipping weekly or faster. The annual model tells you where you were, not where you are."

PSG | Adam Glick, CISO, PSG

What changes during a test



In your experience, how often are security findings already outdated by the time you receive the final report? Select one.

Lovable

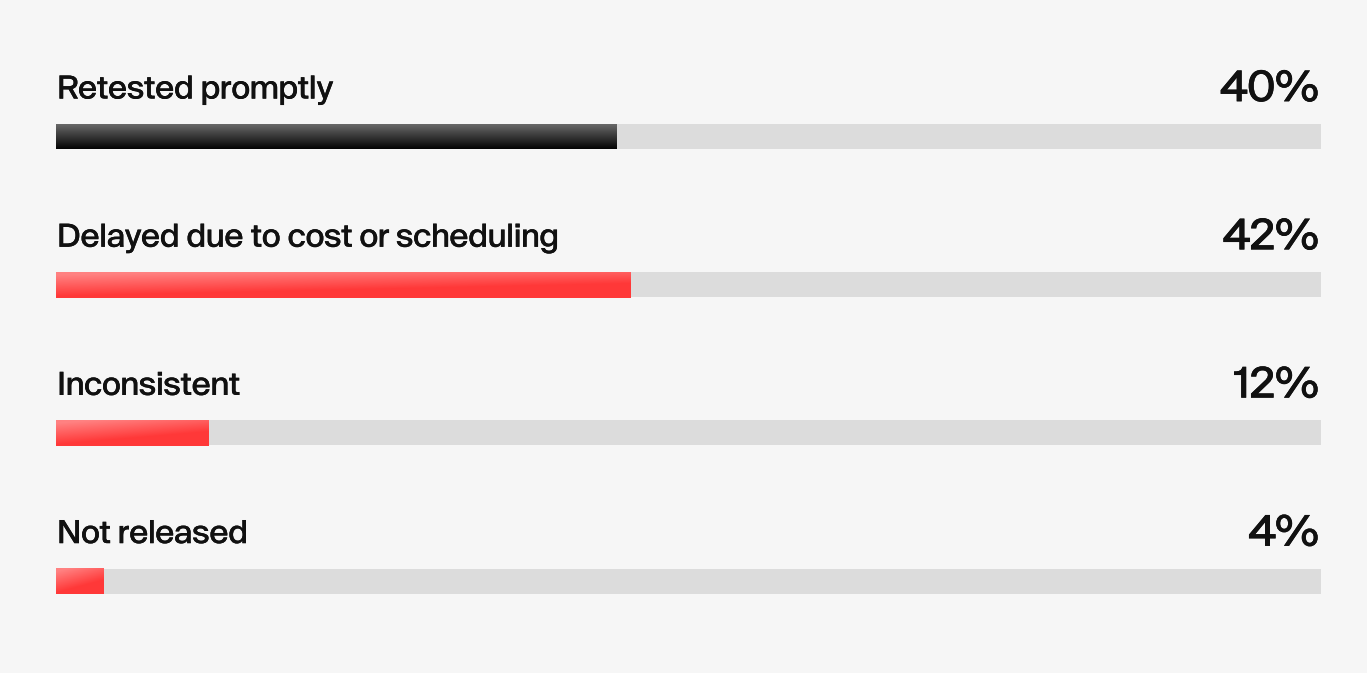
"We see teams go from idea to production in hours, so when security testing takes weeks to return results, you're testing a system that no longer exists."

Anton Osika
CEO, Lovable

WHERE SECURITY TESTING STANDS NOW

Most vulnerabilities aren't promptly re-tested after fixes

Only 40% retest promptly as part of the same engagement. The rest delay, do it inconsistently, or skip it entirely. A fix that isn't verified is a fix you're taking on trust.



Retesting behavior is strongly associated with delivery speed and whether teams actually know what their pentest covered.

72%
have clear coverage among those who retest promptly

VS

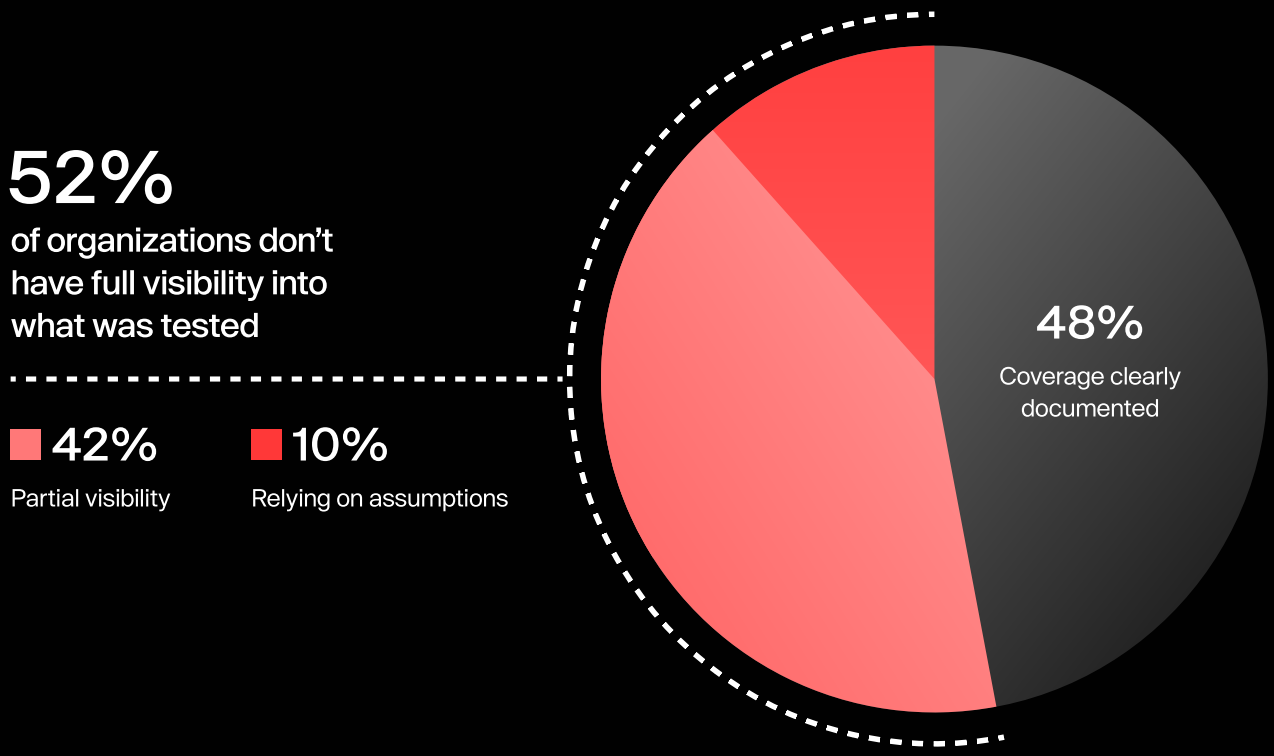
33%
have clear coverage among those who delay or skip

? After a vulnerability is fixed, how soon does your team retest it?

WHERE SECURITY TESTING STANDS NOW

Teams lack full visibility into what was tested

52% of organizations don't have full visibility into what was tested, with 10% relying entirely on assumptions. Many don't have full clarity on what they paid for. Completing a manual pentest doesn't guarantee understanding.



Problems with reproducibility make the uncertainty worse. 27% of organizations say pentest findings are sometimes difficult to reproduce. A finding you can't reproduce is a finding teams can't act on or explain to an auditor with confidence that it was remediated or fixed.

WHERE SECURITY TESTING STANDS NOW

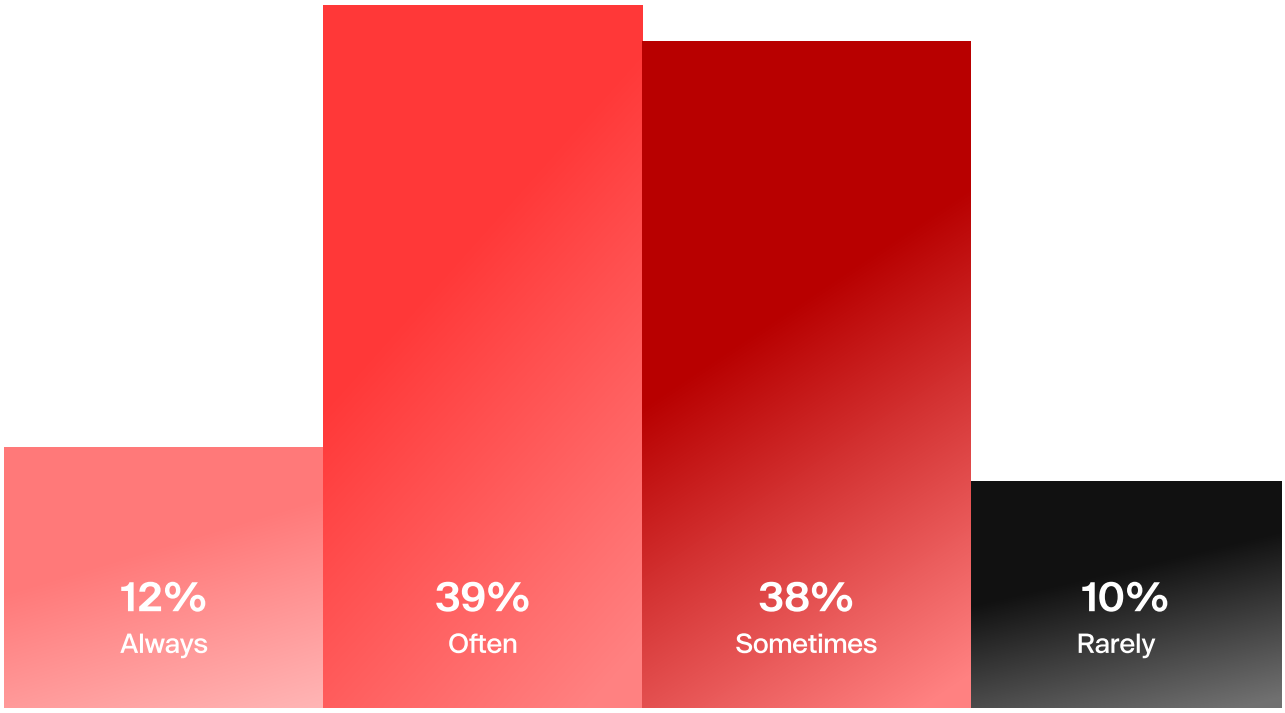
Deeper vulnerabilities are frequently missed in manual testing

Deeper vulnerabilities like logic flaws and multi-step attack paths are regularly missed in manual testing. Over half of respondents say this happens often or always.

Testing is time-boxed, so testers cannot achieve full context, which makes consistent depth difficult to achieve as applications and services are changing more frequently.

Teams shipping multiple times per day report this more strongly, so the gap widens as development speed increases.

How often do you believe human testers miss issues such as logic flaws, broken access controls, or multi-step vulnerabilities?



OWASP

"By the time many assessments are completed, the application has already changed. New code has been introduced, behaviors have shifted, dependencies have updated, and the results may no longer accurately represent what is running in production. That creates operational friction for both security and engineering teams — uncertainty about coverage, missed vulnerabilities"

Andrew van der Stock
Executive Director, OWASP

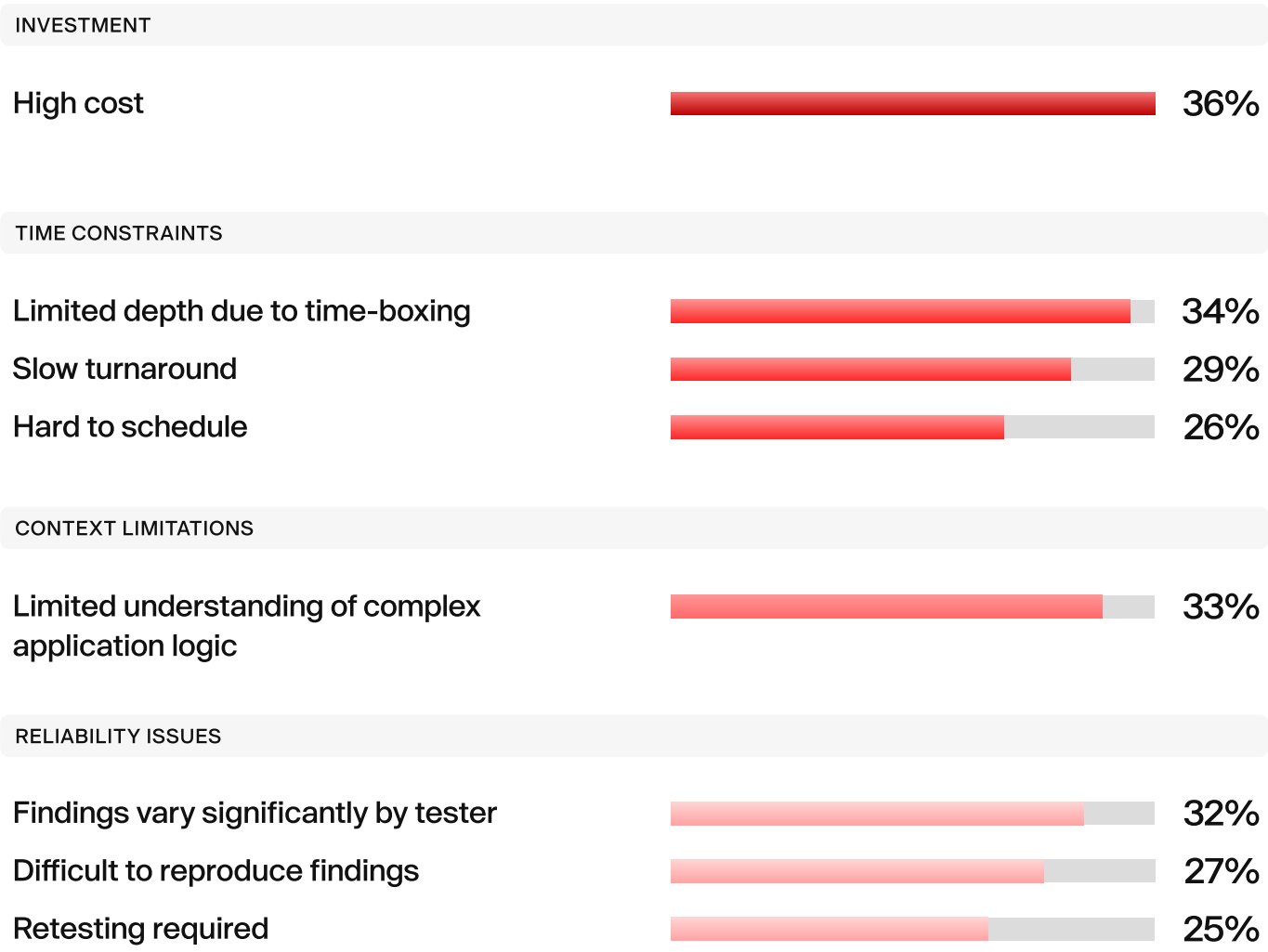
WHERE SECURITY TESTING STANDS NOW

Manual testing is limited by time, context and consistency

97% of leaders report challenges with manual penetration testing.

The challenges of manual testing compound. Time-boxed engagements limit depth, which means testers miss complex logic flaws, which produces findings that vary by tester and are hard to reproduce. The result is expensive testing that still leaves teams uncertain about what was actually covered.

High cost is the most cited issue (36%), but the deeper problems are structural.



glasswall

“A manual pentest is tightly scoped and time-boxed. In that window the obvious things get found: misconfigurations, missing headers, the odd injection. What rarely surfaces are what matter most: logic flaws that only show up when two or three workflows interact, or assumptions buried so deep in the code that you'd need the whole system in your head to spot them. No single tester can carry that much context."

Chris Holman
DevSecOps Engineer Team Lead, Glasswall

balena

“As applications become more specialized, security testing increasingly depends on how quickly testers can build enough context around custom APIs, identity models, and business workflows. Too often, engagements still become exhausting onboarding exercises or drift toward compliance-focused checklists instead of the deep, system-specific analysis modern platforms actually require."

Harald Fischer
Security Aspect Lead, balena

WHERE SECURITY TESTING STANDS NOW

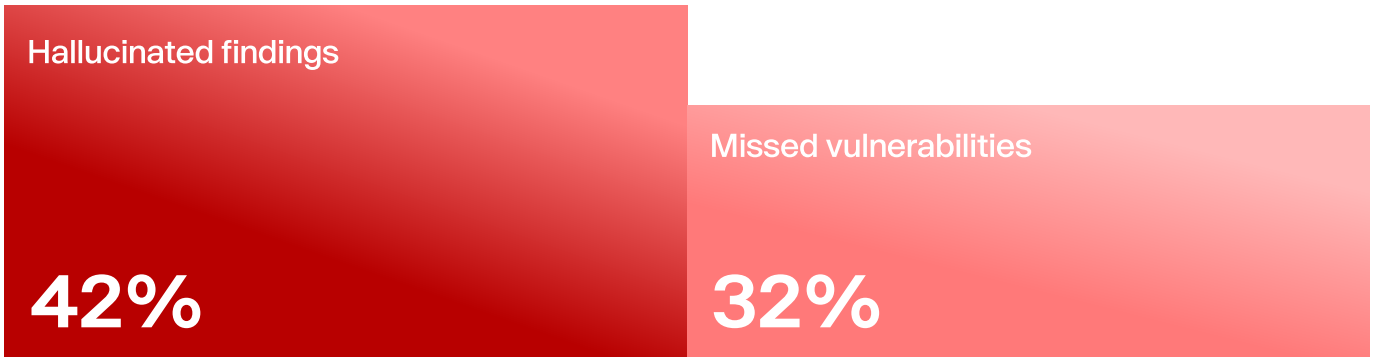
For AI, getting it wrong breaks trust faster than missing vulnerabilities

Accuracy is a bigger trust breaker than operational disruption. False positives damage trust faster than false negatives.

Hallucinated findings force human testers to chase fake results and retest, wasting time and raising unnecessary alarm, ultimately compounding costs.

AI security tools must be provably correct before they can be autonomous and trusted.

What leaders say will damage their trust in AI security testing



By failure category (total mentions)

Accuracy
- Hallucinations
- Missed issues
- Irreproducibility

Operational
- Scope creep
- Outages

Transparency
- Unexplained findings

2% say trust would remain regardless of mistakes

This is what a broken model looks like

76%
deploy changes weekly or faster

64%
Say testing timelines impact release decisions

79%
are concerned about missing vulnerabilities between tests

52%
Lack full visibility into what was tested

48%
Say findings are outdated when they arrive

60%
Do not reliably retest vulnerabilities after fixes

What teams actually want now

Teams want speed, frequency, guardrails and validation

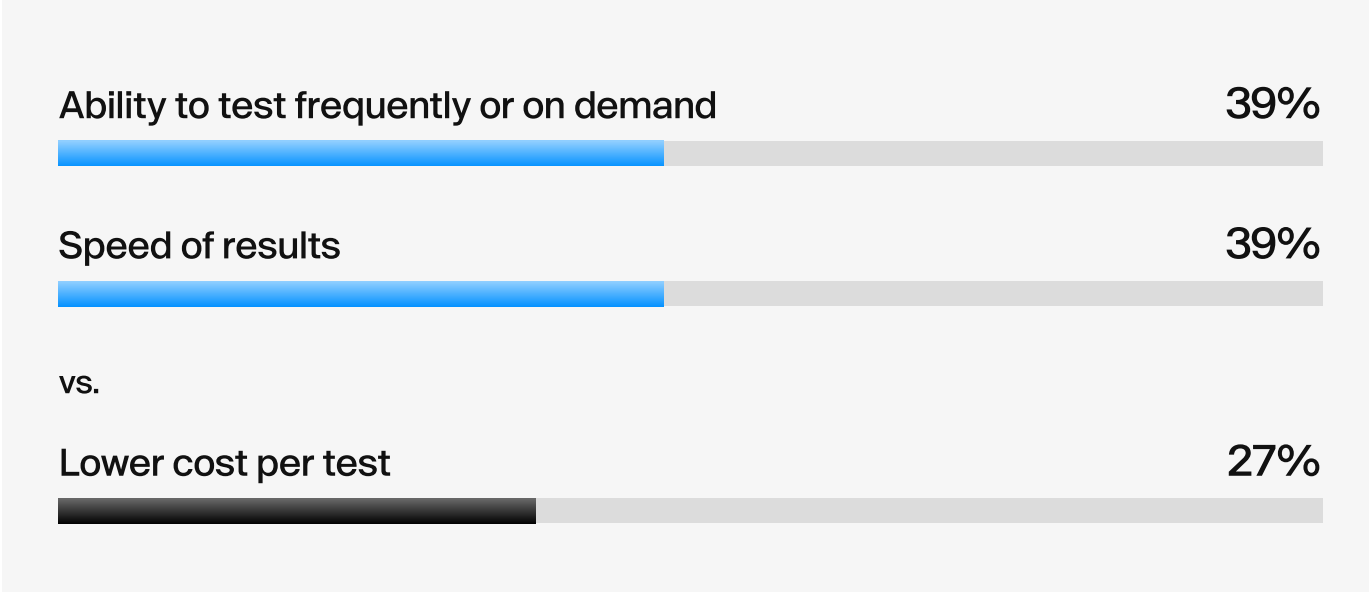
WHAT TEAMS ACTUALLY WANT NOW

Leaders want speed, not savings

On-demand testing and speed of results are the joint top criteria, both at 39%. Cost ranks last at 27%. Leaders aren't looking for a cheaper pentest. They want one that keeps up with how they already ship software.

“AI goes a long way to democratize the dream of continuous pentesting for teams. No longer forced into the costly tradeoffs of scoping full engagements, they're able to run meaningful testing as fast as they're shipping code."

Latio. | James Berthoty, Founder & Analyst, Latio



OTHER CRITERIA

Ability to explain failures to stakeholders	36%	Depth of logic and workflow analysis	36%
Consistency of coverage	35%	Ease of audit or compliance reporting	34%
Reproducible, evidence-backed findings	33%	Automatic fixes	28%

? Which characteristics matter most when considering new approaches to security testing? Select all that apply.

FROST & SULLIVAN

"Organizations may increasingly evaluate security testing solutions not only based on coverage breadth, but also on the reliability, validation, and consistency of the outputs generated as AI becomes more embedded into testing workflows."

Vivien Pua,
Industry Principal, Cybersecurity, Frost & Sullivan

WHAT TEAMS ACTUALLY WANT NOW

Security controls must be built into AI systems, not added later

Which technical controls must be enforced by the system itself before you'd consider AI-driven pentesting acceptable?

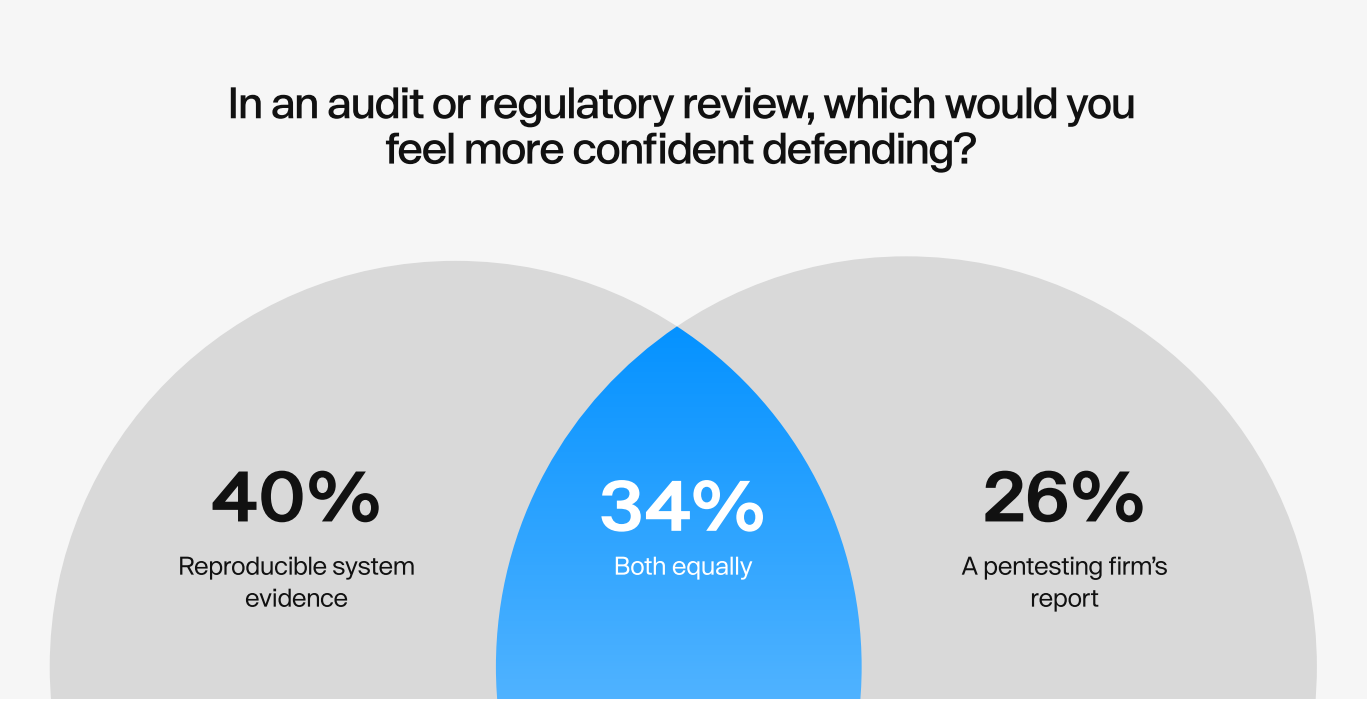
3 controls selected on average per respondent. These are deliberate priorities, not a wish list.

The top two are both purely technical: terminate and data residency. Human oversight ranks third. The instinct is to build hard boundaries into the system rather than rely on a human in the loop for every step.

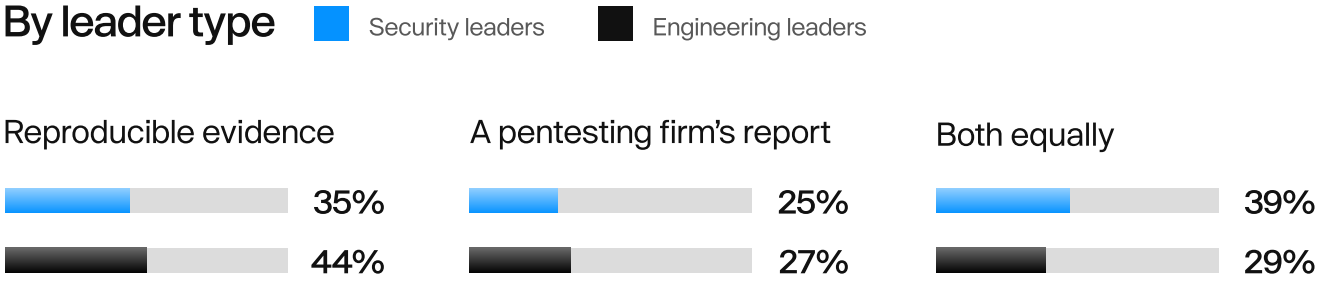
AI is more useful when it's taking work away from people, not adding more.

WHAT TEAMS ACTUALLY WANT NOW

In audits, evidence matters more than reputation



By leader type



Trust in security testing comes from what can be demonstrated, not who delivered the finding.

When auditors raise questions, leaders are more confident defending reproducible system evidence than relying on a provider's reputation (40% vs 26%).

Engineering leaders lean hardest toward evidence (44%), while security leaders are more likely to value both equally (39%).

WHAT TEAMS ACTUALLY WANT NOW

Teams get stuck trying to understand findings

What is the biggest bottleneck in turning externally reported vulnerabilities into fixes?

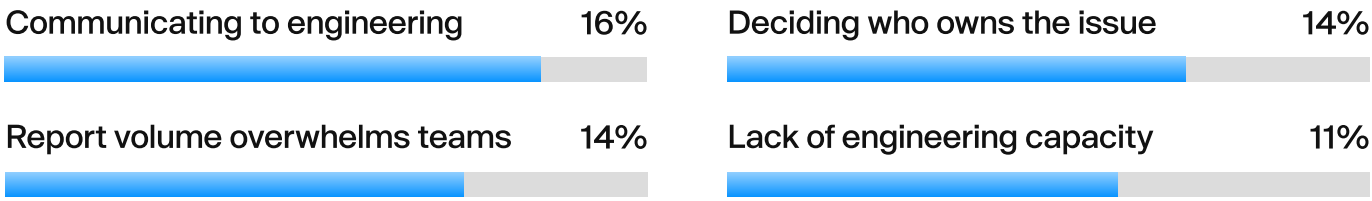
21% Determining true impact and severity

19% Verifying whether the issue is real

40%

say they need more clarity on reported vulnerabilities.

OTHER BOTTLENECKS



Before a finding ever reaches an engineer, someone has to confirm it's real and figure out how much it actually matters. For most teams, that triage step is where progress stalls.

Even once a finding is understood, getting the right people to act on it adds more friction. Teams are not overwhelmed by the number of findings, but feel uncertain about the ones they have been assigned.

WHAT TEAMS ACTUALLY WANT NOW

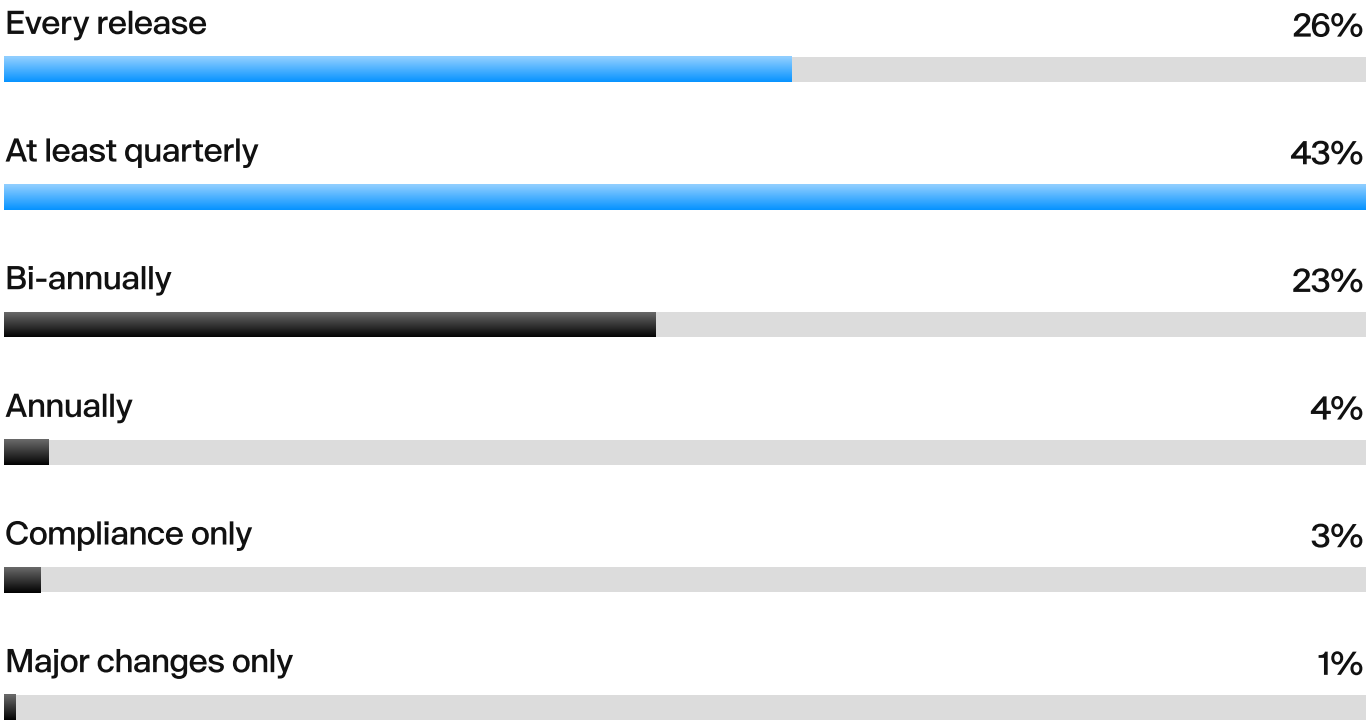
Security validation isn't happening as often as teams want

Cost, staffing and scheduling are all constraints with existing security testing. In an ideal world, teams would run security validation on every release.

Leaders most worried about between-test exposure are also the most likely to want frequent validation. 84% prefer at least quarterly, compared to 65% who are comfortable with quarterly or bi-annual validation.

Security leaders are more likely than engineering leaders to favor validation at every release. Engineering doesn't want to slow deployment, which indicates tradeoff with the way testing is currently done.

69% want validation every release or at least quarterly



? In an ideal world, how often would your team run security validation?

IDC

"Traditional penetration testing was built around periodic assessments, but modern software delivery no longer operates on a periodic schedule. As development velocity increases, especially with AI-assisted software creation, security testing will increasingly need to move closer to the release cycle itself."

Katie Norton
Senior Research Manager, IDC

WHAT TEAMS ACTUALLY WANT NOW

Where does this leave us?

Security testing hasn't kept pace with how software is built, and AI is making that gap more obvious. Teams are shipping changes constantly, but testing still happens at fixed points in time.

By the time results come back, the risks and vulnerabilities in the code have already changed. New code has been deployed, behaviour has changed, and the findings don't always reflect what's actually running in production.

That's where the friction shows up. Teams are forced to deal with missed issues, uncertainty about what was actually tested, and delays when trying to release.

The current model is broken. As software becomes faster and more complex, testing needs to keep up in a way it hasn't before.

Willem Delbare
CEO & Co-founder, Aikido

Data points

DATA POINTS

What does the average pentest look like?

2.6 weeks

average duration of a pentest

~ €25,000

average cost per test

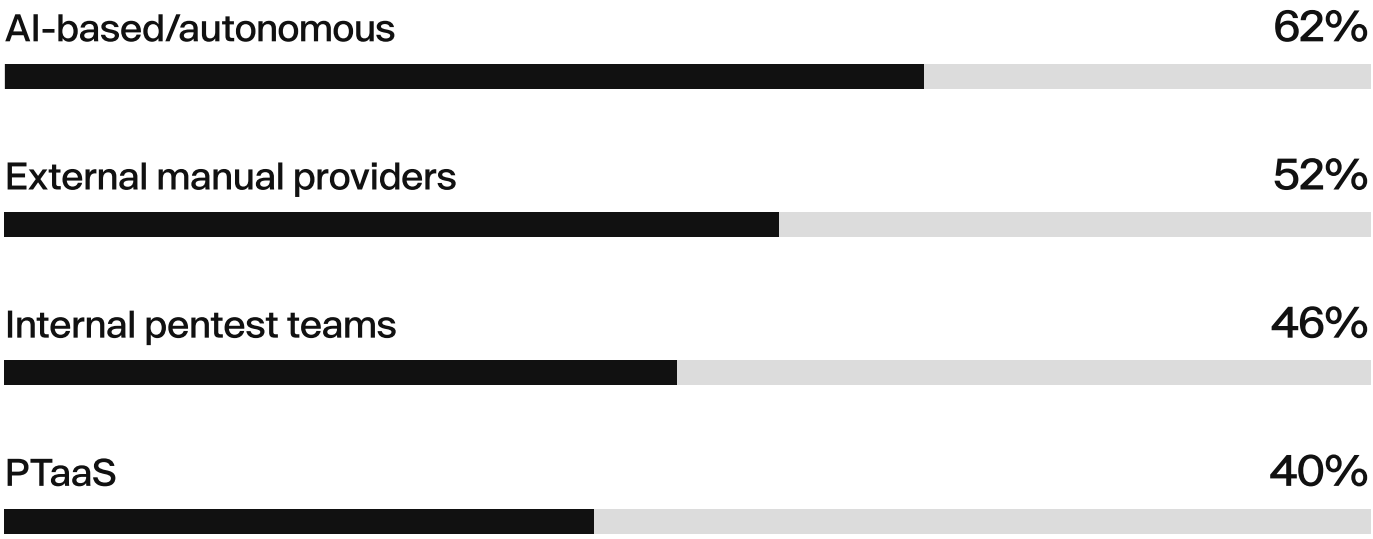
€10k - €50k

what most organizations spend per test

Longer tests mean higher cost. Testing isn't continuous because at these price points, running a pentest on every release isn't realistic for most teams.

DATA POINTS

How pentesting works today



74%
want validation every release or at least quarterly

2/3
favour white-box or high-context testing

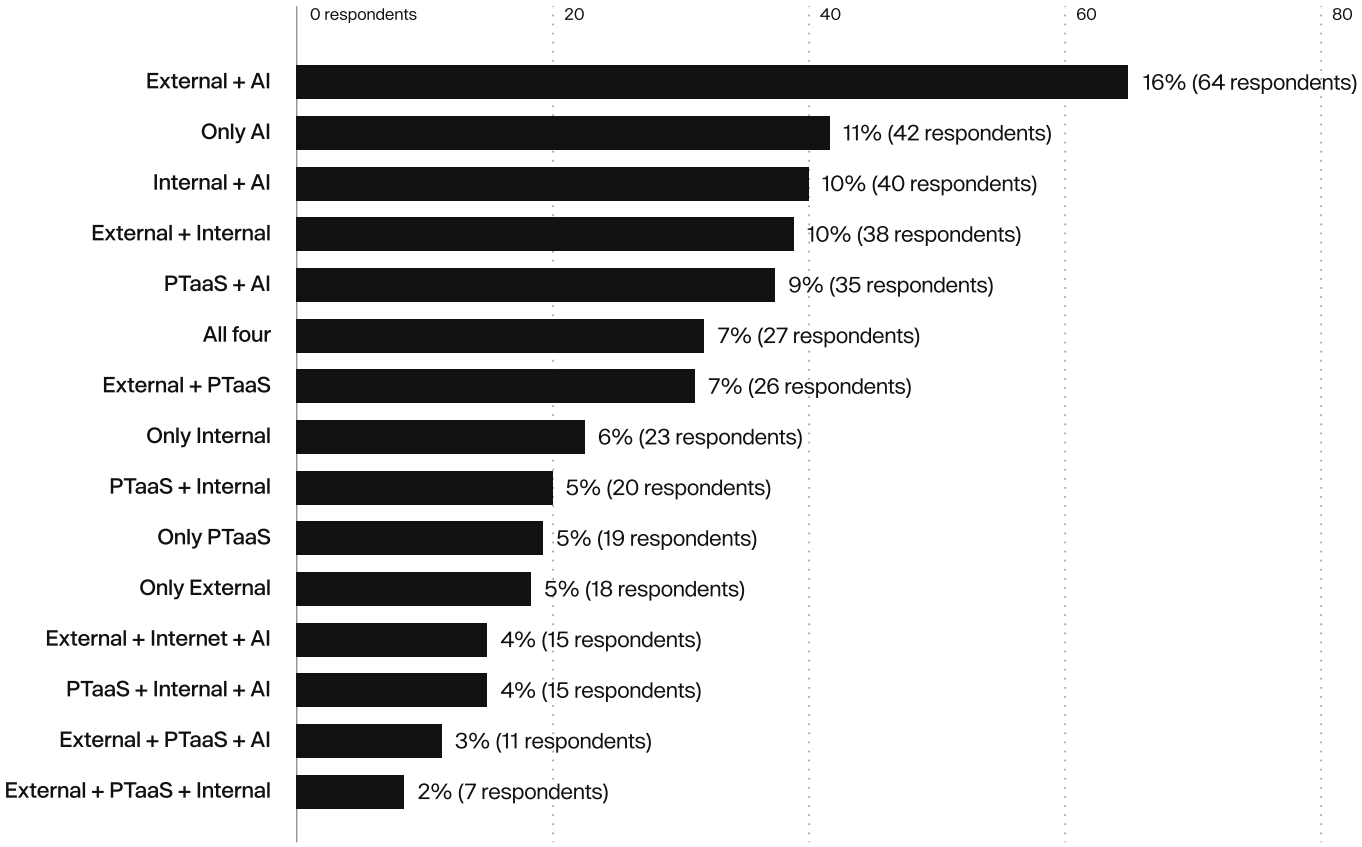
Top drivers
risk reduction and compliance

The testing stack is getting wider, not simpler. Organizations are layering AI on top of existing approaches rather than replacing them. The totals add up to well over 100% because most teams use more than one method.

DATA POINTS

The most common approach is to use four different pentesting methods in parallel

Q3 COMBINATIONS



No single operating model seems to dominate The most common resourcing combination is external providers + AI (16%), but the remaining 84% is spread across many different setups - signalling a fragmented, mixed-model landscape rather than a standardised approach.

Single-source reliance is relatively uncommon. Only 6% rely solely on an internal team, 5% solely on PTaaS, and 5% solely on external providers - reinforcing that most organisations are combining methods rather than betting on one channel.

AI is being layered into hybrid stacks - and sometimes used alone. AI features across many of the top combinations, and 11% report AI-only, suggesting automation is not just additive for some teams, but a standalone approach for a meaningful minority.

Overall takeaway: Pentesting resourcing is trending toward blended models, with AI emerging as a key component both alongside established methods and, in some cases, as the only approach.

DATA POINTS

Who we surveyed

400 full-time employees working in cloud-native environments

COUNTRY OF RESIDENCE

USA	140
Belgium	65
France	65
Germany	65
United Kingdom	65
Total participants	400

JOB ROLE

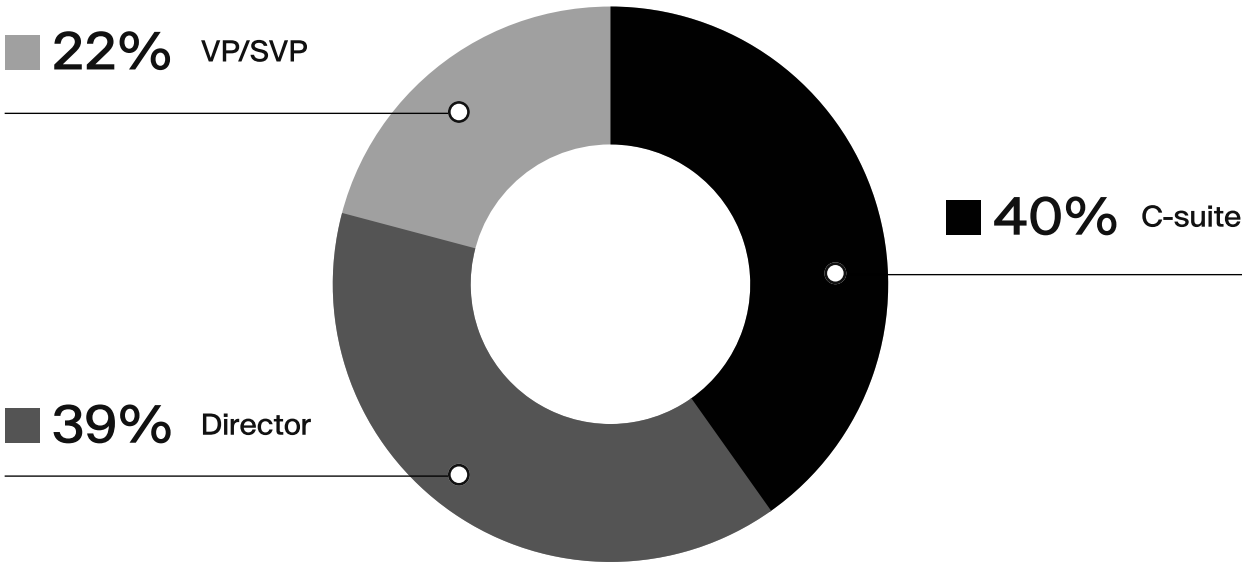
50% (200 employees)

Engineering leaders
CTOs, VP/head of engineering

50% (200 employees)

Security leaders
CISO, head/VP of security

SENIORITY



Credits

Research
Sapio Research, Inc.

Editorial
Sooraj Shah
Dania Durnas
Madeline Lawrence

Design
Alex Moore

Comms & PR
Sooraj Shah
Brittany Cattuci


aikido