Pentest Vendor Selection Checklist

The Buyer's Guide to Choosing a Penetration Testing Vendor:
Scope, Qualify, Evaluate, Verify

Book a consultation

Version 1.0

CONTENTS

How to Use This Checklist

Who this is for. Heads of IT, CISOs, security engineers, and compliance leads who are evaluating penetration testing vendors. Whether you are buying your first pentest or refining a vendor relationship for the next cycle, this checklist covers the full evaluation lifecycle.

Five stages. The checklist follows the order you should work in: define scope, qualify vendors, ask questions during scoping calls, evaluate proposals, and verify the report after delivery. Each stage builds on the previous one.

How to use both documents. This workbook is your internal evaluation tool. The companion Vendor Questionnaire (unbranded, separate file) is what you fill in with your scope and send to each vendor; the vendor fills in their capabilities and returns it. Compare responses side by side using Module 04 criteria.

Priority levels. CRITICAL items are non-negotiable - failing them disqualifies the vendor or invalidates the test. HIGH items are strongly recommended; failure causes downstream problems. MEDIUM items are best practice and improve quality.

01

Define Your Scope

Map the attack surface before you can scope a test. Define which assets need testing, which compliance requirements apply, what test type fits your goals, and whether to test production or staging.

DEADLINE: COMPLETE BEFORE ISSUING ANY RFP

SCOPE

Module 01: Define Your Scope

1A: Asset Inventory

PRACTITIONER TIP

API testing is the most commonly under-scoped area. A web app test does not automatically cover its APIs unless scope says so explicitly - and APIs expose business logic that the UI never touches.

PRACTITIONER TIP

Run a subdomain enumeration before your first scoping call. Buyers who show up with a list get better scope coverage than buyers who say 'our website'.

LIST YOUR ASSETS HERE

- [ ] Identify all web applications (include admin panels, staging, API portals)
  CRITICAL
- [ ] Identify all externally accessible APIs (REST, GraphQL, SOAP) - treat as separate from web apps
  CRITICAL
- [ ] Identify all external network IP ranges and domains
  CRITICAL
  - [ ] All subdomains (run a subdomain enumeration before the scoping call)
  - [ ] Legacy portals and decommissioned-but-still-live systems
- [ ] Identify internal network scope (Active Directory, internal apps, file servers)
  HIGH
- [ ] Identify mobile applications (iOS and Android tested separately)
  HIGH
- [ ] Identify cloud infrastructure scope (IAM, storage, compute, serverless)
  HIGH
- [ ] Identify thick client / desktop applications
  MEDIUM

- [ ] Confirm which assets require written third-party authorisation before testing

  CRITICAL

  - [ ] Third-party SaaS (Shopify, Salesforce, Okta) - tenant config in scope, the platform itself is not
  - [ ] Partner-owned systems accessed via integration - written authorisation required for the partner side

1B: Compliance Coverage

FRAMEWORKS THIS TEST MUST SATISFY

- [ ] Which compliance requirements must this test satisfy? (PCI DSS / SOC 2 / ISO 27001 / NIS2 / FCA / none) - verify below that test scope and methodology cover what the standard expects. Full per-standard requirements: Compliance Matrix at the end of this document.

  CRITICAL

- [ ] If PCI DSS: confirm CDE boundary is documented and agreed

  CRITICAL

- [ ] If PCI DSS: confirm tester has organisational independence from CDE management

  CRITICAL

- [ ] If SOC 2: confirm test scope maps to the System Description in your SOC 2 report

  CRITICAL

1C: Test Type

Test type	What the tester gets	What you get back	Best for
White box	Source code, architecture, full credentials, all docs	Most findings per day. Code-level fix recommendations.	Maximum coverage. Same day rate as gray box often delivers more value because testers spend time finding bugs, not enumerating the stack.
Gray box	Authenticated user access, some docs, partial credentials	Real-world simulation of an attacker with valid credentials.	Standard engagements. Testing authenticated user attack paths.
Black box	Nothing. External view only.	External attack simulation. What an outsider can reach.	Testing your detection and response. Red team scenarios.

Black box does not give you "more realistic" coverage than gray box. It gives you less coverage in the same number of days - the tester spends time on recon that you could have given them upfront. Use black box only when testing detection is the goal.

OUR TEST TYPE

WHY

- [ ] Test type chosen by goal, not budget

HIGH

- [ ] If web app: agreed on which user roles to test (unauthenticated, user, admin, API keys)

HIGH

- [ ] If gray or white box: credentials and accounts ready before test starts

CRITICAL

- [ ] Decision documented for procurement audit trail

MEDIUM

1D: Production or Staging

PRACTITIONER TIP

Production vs staging is rarely an either/or. Test production for read-only and external assets where parity matters; test staging for destructive payloads and internal logic. Most engagements are a mix.

Test in	What you get	What to watch for
Production	Real config, real data, real users. Findings reflect actual attack surface.	Exclude DoS-style payloads. Agree on test windows. Some compliance standards require this (PCI DSS CDE).
Staging	Safe to use destructive payloads. Can test internal logic deeply.	Only useful if staging mirrors production (code, patch level, config). Document any divergence and what it hides.

OUR ENVIRONMENT

PRODUCTION TEST WINDOWS (IF APPLICABLE)

STAGING DIVERGENCES FROM PROD (IF APPLICABLE)

- [ ] Decided where the test runs

CRITICAL

- [ ] If production: DoS-style payloads excluded in writing
  CRITICAL
- [ ] If production: test windows agreed in rules of engagement
  CRITICAL
- [ ] If staging: confirmed it mirrors production OR documented gaps
  CRITICAL
- [ ] If compliance applies: confirmed framework allows your choice
  HIGH

1E: Technical Scope

PRACTITIONER TIP

If you cannot answer these, vendors cannot quote accurately - they will guess, then renegotiate when reality differs from the guess. The cheapest part of an engagement is the hour you spend writing this down before the scoping call.

- [ ] Documented business goal of the test (what success looks like, what risk you are buying down)
  CRITICAL
- [ ] Documented test approach with rationale (white / gray / black, depth, why this fits the goal)
  HIGH
- [ ] Documented architecture summary - modules, microservices, data flows, third-party integrations
  CRITICAL
- [ ] Documented hosting model (cloud provider and region, on-prem, hybrid)
  HIGH
- [ ] Documented technology stack per app or repository (languages, frameworks, major libraries)
  CRITICAL
- [ ] Have CLOC output for each in-scope repository (one report per repo if multi-repo)
  HIGH
- [ ] Documented CI/CD pipeline summary (build, test, deploy, gating, who can push to prod)
  HIGH
- [ ] Counted application entry points - REST / GraphQL / SOAP endpoints, message queues, scheduled jobs
  CRITICAL
- [ ] Counted forms in scope (HTML form and input elements in the user-facing app)
  MEDIUM
- [ ] Listed all application roles in scope (every distinct role the app supports)
  CRITICAL

- [ ] Listed all permission groups or capabilities in scope

  CRITICAL

- [ ] Decided which roles and permissions are in test scope vs out of scope (if not all, name the subset)

  CRITICAL

02

Qualify the Vendor

Use pass/fail checks before engaging any vendor deeply. These are qualifying criteria, not scoring items. Any FAIL disqualifies the vendor from shortlist consideration.

DEADLINE: BEFORE ISSUING SHORTLIST INVITATIONS

QUALIFY

Module 02: Qualify the Vendor

2A: Certifications

PRACTITIONER TIP

When calling references, ask whether the testers who showed up matched the proposal. Named-tester swaps are the most common bait-and-switch in this market.

PRACTITIONER TIP

Verify certifications for the specific testers on your engagement, not the company's overall list. Company-level credentials tell you nothing about who will actually do the work.

- [ ] Confirm the vendor can name specific testers assigned to your engagement before contract signature

  CRITICAL

- [ ] Verify hands-on offensive cert on every named tester (OSCP minimum, or OSWE / OSCE / OSEP / OSED / GPEN / GWAPT / GXPN / eWPTX). For red team scope, CRTO and CRTE add value. CISSP / CISA are governance certs, not proof of hands-on skill.

  CRITICAL

- [ ] For UK regulated sectors only (HMG suppliers, NCSC CHECK scope): verify the vendor holds NCSC CHECK status with UK CSC Security Testing Professional Titles (Practitioner minimum) on the named testers. OSCP alone does not satisfy this - verify at ncsc.gov.uk.

  HIGH

- [ ] Ask which certs the testers on your engagement personally hold - vendors often list aggregate company certs that no single tester possesses

  HIGH

2B: Team

- [ ] Are testers employees or freelancers?

  HIGH

- [ ] How many concurrent engagements per tester during yours? More than 3 is a yellow flag.

  HIGH

- [ ] Do senior testers do the hands-on work, or supervise juniors? Confirm the model upfront.

  CRITICAL

2C: CVE Research History

- [ ] Search the vendor's CVE history on NVD (nvd.nist.gov) or CVE.org. Published CVEs in enterprise software signal genuine research capability.

  HIGH

2D: Methodology

- [ ] Does the vendor ask detailed scoping questions before quoting? (No scoping questions = no customisation)

  CRITICAL

- [ ] Can the vendor name the methodology per test type without prompting? Web: OWASP WSTG. Network/infra: NIST SP 800-115 + MITRE ATT&CK. Red team: MITRE ATT&CK.

  CRITICAL

- [ ] How do they prioritise findings - CVSS base score only, or adjusted by business context in your environment?

  HIGH

2E: References

PRACTITIONER TIP

Ask for a sample report from a similar engagement. If NDAs prevent sharing, a good vendor offers a call walkthrough or a redacted excerpt. A flat 'no' with no alternative is a disqualifier.

- [ ] Ask for three references from similar engagements. Contact them.

  HIGH

- [ ] Ask for a sample report. If NDAs prevent sharing, a walkthrough on a call or a redacted excerpt is the alternative.

  CRITICAL

03

Questions to Ask Every Vendor

Twenty-four questions for methodology, team, tooling, reporting, retest policy, data handling, and insurance. Document vendor answers; use the answers to populate Module 04 scoring.

DEADLINE: DURING SCOPING CALLS AND PROPOSAL EVALUATION

QUESTIONS

Module 03: Questions to Ask Every Vendor

3A: Methodology and Tooling

PRACTITIONER TIP

No vendor can guarantee zero disruption - they can manage the risk. Snapshot procedures, agreed test windows, and a defined kill switch should be in the rules of engagement before testing starts.

- [ ] 'Which specific OWASP WSTG categories will you cover? Which will you skip and why?' - forces vendor to commit to coverage

  CRITICAL

- [ ] 'Show me a vulnerability chain you exploited end-to-end on a recent engagement (anonymised)' - separates testers who chain from those who list independent findings

  HIGH

- [ ] 'How do you handle out-of-scope discoveries - tell us, document, or silently exploit?' Good vendors document and notify; they do not silently exploit.

  CRITICAL

- [ ] 'How much testing time on this scope is manual vs tool-assisted? Show me a finding scanners would miss.'

  HIGH

- [ ] 'What commercial and open-source tools do you use for this scope type?'

  HIGH

3B: Reporting and Retest

- [ ] 'Does your report map findings to compliance controls (SOC 2, PCI DSS, ISO 27001)?'

  HIGH

- [ ] 'What does a finding look like when remediation requires an architecture change rather than a patch?'

  MEDIUM

- [ ] 'Is retest of Critical and High findings included in this scope?' Inclusion is standard for typical engagements; very large or red team scopes may scope retest separately, which is fine when stated upfront.

  CRITICAL

- [ ] 'How long after report delivery is retest available?' 30-90 days is typical; longer windows are a strong signal.

  HIGH

3C: Data Handling, NDA, Insurance

PRACTITIONER TIP

Pushback on specific NDA clauses is normal. Good vendors flag clauses that conflict with how testing works - blanket bans on retaining methodology notes, or unrealistic destruction timelines. A vendor who refuses to sign at all, or signs without reading, is a problem.

- [ ] 'Data retention policy for test artifacts - screenshots, logs, exploit code, captured credentials?'
  CRITICAL
- [ ] 'How do you store and destroy credentials and sensitive data collected during testing?'
  CRITICAL
- [ ] 'Will you sign our NDA? Standard clauses you typically negotiate?' Good vendors flag conflicts and propose mutual edits.
  HIGH
- [ ] 'Where are testers located? Any offshore handoff or freelancers without disclosure?'
  HIGH
- [ ] 'Professional indemnity / E&O insurance carried? Coverage limit?' Insurers should be named and the limit should be stated in writing.
  CRITICAL
- [ ] 'Policy if the test causes service disruption? Kill switch procedure documented in the rules of engagement?'
  CRITICAL

Red Flags: Stop the Evaluation

Two desk references for shortlist review. The left column lists vendor signals that should pause or end your evaluation. The right column lists proposal signals that require documented justification before proceeding. Multiple flags = disqualify.

Vendor signals - stop evaluation

- ✕ Cannot name specific testers before contract signature
- ✕ Cannot say which certifications the assigned testers personally hold
- ✕ Promises zero disruption to production (impossible)
- ✕ No questions asked about your tech stack or scope
- ✕ Methodology described in acronyms only ('we follow all major frameworks' or equivalent non-answer)
- ✕ Offshore team or freelancers not disclosed before contract
- ✕ Refuses to negotiate any NDA clause - rubber-stamp or walk away
- ✕ No sample report and no offer to walk through one on a call

Proposal signals - require justification

- ✕ Report promised in 24-48 hours after testing
- ✕ No OSCP-equivalent hands-on cert on any named tester
- ✕ Methodology section copy-pasted from the vendor's website
- ✕ No escalation procedure for critical findings mid-test
- ✕ All sample findings rated High or Critical (severity inflation)
- ✕ Retest excluded with no explanation

04

Evaluate Proposals

Score proposals against weighted criteria. Identify red flags that should pause or end the evaluation. Compare the scope section against your requirements from Module 01.

DEADLINE: WITHIN 5 BUSINESS DAYS OF PROPOSAL RECEIPT

EVALUATE

Module 04: Evaluate Proposals

4A: Proposal Quality

PRACTITIONER TIP

CVSS base scores are not the same as business risk. A CVSS 9.8 on an isolated dev server may matter less than a CVSS 6.5 on your payment API. Vendors who weight findings by business context give you a remediation queue you can actually work; vendors who hand you a CVSS-sorted list make you do that work yourself.

PRACTITIONER TIP

Ask: 'What does a finding look like when remediation requires an architecture change rather than a patch?' A senior tester has seen this. A junior with a scanner hasn't. The answer quality tells you which you are getting.

- [ ] Scope section restates your requirements in the vendor's own language - proof they read the RFP
  CRITICAL
- [ ] Names specific testers with personal certifications - not an aggregate company list
  CRITICAL
- [ ] Methodology specified per test type with named OWASP WSTG categories or MITRE ATT&CK techniques
  CRITICAL
- [ ] Sample report provided OR walkthrough offered when client NDAs prevent sharing
  CRITICAL
- [ ] Retest policy explicitly stated: which severities included, which separate, timeframe
  HIGH
- [ ] Insurance, NDA position, escalation, and data handling addressed in writing
  HIGH
- [ ] Deliverables list complete: exec summary, technical report with PoC, attestation, retest
  HIGH

4B: Red Flag Checklist

- [ ] Report delivery promised within 24-48 hours of end of testing - manual analysis takes days
  CRITICAL
- [ ] No hands-on offensive cert (OSCP, OSWE, OSCE, CRT, GPEN, eWPTX or equivalent) on any named tester
  CRITICAL

- [ ] Senior tester named in proposal with no contractual commitment they will be on your engagement
  CRITICAL
- [ ] No documented escalation procedure for Critical findings discovered mid-test
  CRITICAL
- [ ] No scoping questions asked before the proposal was sent
  HIGH
- [ ] Methodology described as 'we follow all major frameworks' or 'industry best practices'
  HIGH
- [ ] Offshore delivery team or freelancers not disclosed before contract
  HIGH
- [ ] Retest excluded entirely with no explanation
  HIGH
- [ ] No data retention/destruction policy in writing
  HIGH

4C: Required Deliverables

- [ ] Executive summary: non-technical, 2 pages max, suitable for board/CEO
  CRITICAL
- [ ] Technical findings: CVSS-scored, affected assets, exploitation steps
  CRITICAL
- [ ] PoC evidence per finding: screenshots + request/response OR terminal output
  CRITICAL
- [ ] Remediation: specific (library version, config, code fix), not 'apply vendor patch'
  CRITICAL
- [ ] Attestation letter: confirms date, scope, and methodology
  CRITICAL
- [ ] If compliance-driven: findings mapped to SOC 2 / PCI DSS / ISO 27001 controls
  HIGH

05

Post-Engagement Quality Control

Evaluate the report you received. Establish remediation tracking. Execute retest. Decide whether to retain the vendor for the next cycle.

DEADLINE: WITHIN 5 BUSINESS DAYS OF REPORT RECEIPT

VERIFY

Report Quality Anatomy

Use this to triage the report you receive. A finding that lacks the elements on the left is not a finding - it is a scanner result. The right column lists the failure patterns that almost always signal an automated scan packaged as a pentest.

Good finding includes

- → Title with vulnerability type
- → Severity (Critical/High/Medium/Low/Informational)
- → CVSS v3.1 or v4.0 score with vector string
- → Affected URL / asset / component
- → Proof-of-concept: screenshot + request/response or terminal output
- → Exploitation steps (what an attacker could do with this)
- → Business impact in plain language
- → Remediation: specific, actionable (version, config, code change)

Red flags in a finding

- × Findings sourced from scanner output, no exploitation evidence
- × 'Refer to vendor documentation for remediation'
- × All findings rated High or Critical (severity inflation)
- × Generic descriptions that could appear in any pentest report
- × No PoC screenshots, or blurred/illegible ones
- × No reproduction steps a developer could follow

Module 05: Post-Engagement Quality Control

5A: Report Quality

PRACTITIONER TIP

Pick three Critical/High findings from the report at random and try to reproduce them from the report alone. If you can't, the vendor cut corners on documentation.

- [ ] Executive summary: a non-technical exec can read it in one sitting and know risk level plus top priorities

  CRITICAL

- [ ] Every finding has CVSS v3.1 or v4.0 score with vector string

  CRITICAL

- [ ] Every Critical/High has reproduction evidence: screenshot plus HTTP request/response or terminal output

  CRITICAL

- [ ] Remediation specific: library version, config parameter, code example - not 'apply vendor patch'

  CRITICAL

- [ ] Findings prioritised by business impact in your environment, not just CVSS

  HIGH

5B: Remediation Tracking

- [ ] Assign an owner to every Critical/High finding within 48 hours of report delivery

  CRITICAL

- [ ] Load all findings into your issue tracking system (Jira, ServiceNow, Linear)

  HIGH

- [ ] Set remediation SLAs per severity: Critical 30 days, High 90 days, Medium next release cycle, Low/Informational backlog

  HIGH

- [ ] Architectural findings: escalate to engineering leadership with a timeline

  HIGH

- [ ] Do not close a finding without evidence of fix and successful retest

  CRITICAL

5C: Retest Execution

PRACTITIONER TIP

Check for systemic patterns after retest. If 8 out of 12 High findings share the same root cause, that is a process or architecture problem - not 8 separate bugs. A good vendor names this pattern in the debrief.

- [ ] Retest scope confirmed: only findings from the original engagement, not new features or assets
  CRITICAL
- [ ] Retest scheduled within the vendor's post-delivery window (typically 30-90 days)
  HIGH
- [ ] Addendum includes finding ID, original severity, fix applied, Pass/Fail, updated evidence
  CRITICAL
- [ ] New vulnerability introduced by a fix? Treat as a new finding with its own tracking.
  HIGH

5D: Vendor Retention Decision

PRACTITIONER TIP

In most organisations, the second engagement with the same team uncovers deeper issues - testers spend less time on recon and more on novel paths. A vendor with consistent findings across two consecutive cycles may be finding the same risk tier each time. Switch for a quality reason, not procurement optics.

- [ ] Decide whether to retain. Key questions: did the testers match the proposal? Were findings novel or low-hanging fruit? Was remediation specific? Was the vendor responsive during and after?
  HIGH
- [ ] Document the retention decision with rationale - procurement audit trail
  MEDIUM

RETENTION DECISION AND RATIONALE

ABOUT THE PUBLISHER

AFINE

AFINE is a Polish offensive security firm operating since 2015. We deliver penetration testing, security audits, red team operations, and TLPT/DORA engagements for banks, fintechs, healthcare providers, and SaaS scale-ups across Europe.

Our research team has published over 150 CVEs in SAP, Microsoft, IBM, CyberArk, Check Point, F5, Rapid7, Palo Alto Networks, BMC, and Adobe. Active research programs keep testers current on exploitation techniques - the same skills that surface issues on client engagements.

150+	10	ISO 27001
PUBLISHED CVES	YEARS OPERATING	CERTIFIED

> *"AFINE keeps finding critical issues where other pentesters have not found them."*
>
> ISABEL GROUP

SERVICES

- Penetration testing
- Red team operations
- TLPT / DORA engagements
- NIS2 compliance assessments
- CVE research

REFERENCE CLIENTS

- PKO BP
- ING Bank
- BGK
- Medicover
- Qualtrics
- CPBil, BIK, Tpay, BPS, Altira

TEAM CERTIFICATIONS

- OSCP (min. per researcher)
- OSCE, OSWE, OSED, OSEP
- CRTO, CSSA
- CISSP, CISA
- BSCP

Ready to scope your next engagement?

afine.com | pwoyke@afine.com

Book a consultation

Compliance Matrix

A side-by-side view of how the major compliance standards treat penetration testing. Verify the requirements that apply to your organisation before scoping the engagement.

Framework	Frequency	Tester independence	Methodology required	Retest required	Notes
PCI DSS v4.0	Annually + after major change	Independent from CDE management	Exploitation attempts required (Req 11.4); 'manual' appears in guidance, not normative text	Required for CDE	ASV scanning (11.3) is separate
SOC 2 Type II	Commonly expected annually; not mandated	External vendor recommended	OWASP WSTG, NIST 800-115	Auditors commonly accept it	Scope = system description; CC4.1 monitoring activities
ISO 27001:2022	Risk-based (not mandated)	External vendor	OWASP, NIST 800-115	Best practice	Control A.8.8 (Management of technical vulnerabilities)
NIS2 (Art. 21(2)(f))	Risk-based (regulator interpretation)	Independent recommended (regulatory interpretation, not directive text)	Recognised methodology	Best practice	Effectiveness assessment of measures
FCA	Risk-based (annually typical)	Sector experience expected	Industry-accepted methodology	Best practice	Operational resilience; CBEST for supervisory scope

Ready to choose the right penetration testing partner?

Book a consultation

AFINE sp. z o.o.
pwoyke@afine.com

Notices and Disclaimer

Please read before relying on any guidance in this document.

GENERAL INFORMATION ONLY

This document is published by AFINE sp. z o.o. for general informational purposes. It reflects penetration testing procurement practice as of the publication date and is intended as a buyer-side reference for organisations evaluating third-party security testing vendors.

NOT LEGAL, REGULATORY, OR PROFESSIONAL ADVICE

Nothing in this document constitutes legal, regulatory, compliance, financial, or technical advice. Every engagement carries unique contractual, jurisdictional, and risk considerations. Readers should obtain qualified advice based on their own circumstances before acting on any guidance contained here.

NO WARRANTY

This document is provided "as is" without representation or warranty of any kind, express or implied, as to accuracy, completeness, or fitness for any particular purpose. Compliance frameworks, threat landscapes, and procurement practices evolve; readers are responsible for verifying that referenced standards (including but not limited to PCI DSS, ISO 27001, DORA, NIS2, SOC 2, OWASP, MITRE ATT&CK) and their interpretations remain current at the time of use.

LIMITATION OF LIABILITY

To the fullest extent permitted by applicable law, AFINE sp. z o.o. and its officers, employees, and contributors accept no liability for any direct, indirect, incidental, consequential, or special loss or damage arising from use of, or reliance on, this document. Vendor selection decisions remain the sole responsibility of the reader.

AUTHOR'S INTEREST

AFINE is itself a provider of penetration testing services. This document is offered as a community resource and is not vendor-neutral marketing. The criteria here should not be treated as endorsement of, or qualification for, any specific vendor including AFINE.

THIRD-PARTY MARKS AND REFERENCES

All third-party trademarks, framework names, regulatory references, and reference clients mentioned remain the property of their respective owners and are used here for descriptive purposes only. No affiliation, sponsorship, or endorsement is implied unless expressly stated.

PERMITTED USE

You may use, share internally, and adapt this document for non-commercial procurement purposes within your own organisation, provided this notice remains intact. Republication, resale, or commercial redistribution requires prior written permission from AFINE sp. z o.o.