

THE BUSINESS LEADERS SERIES

Penetration Testing for Business Leaders

WHAT TO TEST, WHAT TO ASK, AND HOW TO TURN
FINDINGS INTO ACTION

AN EXECUTIVE REFERENCE GUIDE

*Close the distance between what you believe
about your security and what is actually true.*

That distance is where breaches live.



THIS GUIDE · 1 OF 6 IN THE SERIES

J.W. WALKER · [JEREMYWALKER.US/RESOURCES](https://jeremywalker.us/resources)

Penetration testing is a business decision that most organizations make without the full picture. Most executives know they should care. Few have a clear picture of what penetration testing actually is, what it really costs when something goes wrong, or how to tell whether they are getting genuine value from it. This guide is written to close that gap.

WHAT YOU ACTUALLY CONTROL

Most of a penetration test is technical work you should delegate. But a handful of choices are business decisions, not technical ones, and they shape whether the test measures your real risk or just produces an artifact to get filed away.

Scope. What gets tested, and just as importantly, what does not. This is the decision that most determines the value of the engagement, and it is fundamentally a business judgment about what you can least afford to lose. See "Understanding the Variables" and "How a Penetration Test Works."

Depth. How far the engagement goes, from a focused test to a full red team. This is decided with input from your technical team, but the call on how much rigor is worth the cost is yours, and it should track your security maturity rather than a vendor's upsell. See "Understanding the Variables."

Cadence. How often you test, and what prompts a test outside the schedule. Every significant change made after a test is a change that has never been tested. See "Real-World Constraints" and "Best Practices."

Partner. Who does the work, and whether they can prove they are qualified. The value of a test is inseparable from the people running it. See "Choosing the Right Partner" and "Ten Questions to Ask Your Testing Firm."

WHAT'S IN THIS GUIDE

Introduction	The perspective behind this guide and who it is written for
Why You Should Care	The business case and what a breach actually costs
You May Not Have a Choice	Regulatory requirements, compliance obligations, and cyber insurance
What Is a Penetration Test?	Formal definitions and what testing is designed to do
What Penetration Testing Is Not	Key distinctions and common executive misconceptions
Who Needs to Be Involved	The full stakeholder map and what each group needs from the engagement
Understanding the Variables	The three decisions that define any test: knowledge, position, and model
What Are We Testing?	Target types, including network, web app, cloud, and social engineering
How a Penetration Test Works	The five-phase lifecycle from scoping through reporting
Real-World Constraints	Budget, time, disruption, third parties, and the compliance-checkbox problem
Choosing the Right Partner	Credentials worth trusting, credentials worth ignoring, and how to evaluate quality
Best Practices	Eight principles for getting genuine value from engagement
Ten Questions for Your Testing Firm	The vendor-selection questions that separate value from paper
Conclusion and Glossary	A closing framework and plain-language definitions of every key term

Introduction

Having spent a healthy portion of my career on the front lines of penetration testing in both commercial and government environments, owning a penetration testing company, and now serving as CTO of an information security firm, I have had the rare opportunity to view this field from both a technical and executive purview. I have sat in the tester's chair, run a business built on testing, and advised organizations of all shapes and sizes.

Over the course of those experiences, I have discussed penetration testing with people across every level of hundreds of organizations, from front-line administrators running the systems being tested to CEOs and board members trying to make sense of a report they just received. What I discovered is that **penetration testing means something fundamentally different to almost everyone**. That distinction is not trivial. The misalignment between what a penetration test is, what organizations believe they are getting, and what they actually need is one of the most persistent and costly problems in the field.

Today, penetration testing is written about almost exclusively by penetration testers, for penetration testers. The literature is technically dense, rich in methodology, and largely inaccessible to anyone who does not already live inside the field. That serves practitioners well, but it leaves a significant gap for the people making the business decisions around it.

Foundationally, penetration testing is conducted to serve the business and ultimately protect its customers, members, and other stakeholders.

What follows is a business-oriented reference: what penetration testing is, what it is not, why it matters, how it works, and how to get real value from it. It is intended to be both self-contained and a document you return to. Part article, part working manual for business leaders navigating the security landscape.

This guide intentionally avoids deep technical detail. The goal is to give decision-makers enough context to ask the right questions, set the right expectations, and make informed investments.

Why You Should Care

Unless your organization is a cybersecurity company, penetration testing will not show up as a line of revenue. You spend money on it because the cost of not spending it can be catastrophic when something goes wrong. Think of it the way you think about insurance, financial audits, or legal compliance: the value comes from what you avoid.

That calculation has only grown more urgent. Attackers now use AI to find and exploit weaknesses faster than ever, compressing the time between when a vulnerability appears and when it is used against you. The window for finding your own gaps first is narrower than it was even a year ago.

What a Breach Actually Costs

Organizations that have not experienced a significant breach tend to underestimate what a breach actually involves. **A breach is rarely a single, contained event. It is a cascade, and the costs accumulate quickly:**

- Regulatory fines and penalties. GDPR, HIPAA, and PCI-DSS violations can run into the millions, and significantly more in serious cases.
- Incident response and forensic investigation. Understanding what happened, how far it went, and what was accessed requires specialized expertise and time.
- Customer notification obligations. Breach notification is widely required across jurisdictions; credit monitoring is sometimes offered or expected but is not universally required. Both are operationally complex to execute.
- Litigation and regulatory proceedings from both regulators and affected customers or partners.
- Reputational damage. Breaches make headlines. Customer trust, once lost, is difficult to rebuild. The speed and transparency of the response shapes public perception more than the breach itself.
- Lost and deferred revenue. Customers and partners increasingly require evidence of security maturity before signing contracts. An organization that cannot provide it loses business it may never know it lost.

A Note on Right-Sizing

Not every organization faces the same level of risk, and no organization has unlimited resources. The appropriate investment in penetration testing is proportionate to the sensitivity of what you hold, the regulatory environment you operate in, and the realistic threat landscape for your industry and size. The goal is to understand your risk clearly enough to make informed decisions about it. **Elimination of all risk is not achievable, and pretending otherwise wastes resources.**

You May Not Have a Choice

For some organizations, penetration testing is explicitly required. For others, regulators and standards require risk-based security testing, with penetration testing serving as one method of validating key controls. Either way, testing requirements increasingly shape how organizations design their security programs.

Two governance concepts show up repeatedly across security programs:

- **Due care** refers to acting responsibly and implementing reasonable protective measures. It is the organizational equivalent of locking your doors.
- **Due diligence** goes further: actively investigating your vulnerabilities, testing your defenses, and being able to demonstrate that you have done so.

Penetration testing is an essential part of an effective security program. By finding exposed services, weak configurations, and lateral movement paths that static controls miss, it is one of the clearest ways to satisfy both your security and your compliance obligations at once.

Financial Institutions: NCUA, FFIEC, and GLBA

For banks and credit unions, security testing is not optional and examiners actively seek proof of completion. The framework is layered. The Gramm-Leach-Bliley Act (GLBA) requires financial institutions to protect customer information. For federally insured credit unions, NCUA implements this through 12 CFR Part 748 and its Appendix A. NCUA guidance calls for regular testing of key controls, with scope and frequency driven by risk. Testing should be conducted or reviewed by an independent third party or staff independent of those maintaining the security program. Management should report the overall status of the information security program to the board at least annually, including material testing results.

The Federal Financial Institutions Examination Council (FFIEC), the interagency body that includes the NCUA, FDIC, Federal Reserve, OCC, and CFPB, publishes the IT Examination Handbook that examiners work from. Its Information Security booklet calls for testing by qualified, independent parties as a component of an effective program. FFIEC guidance does not name penetration testing as a rigid standalone mandate, but examiners treat independent testing, scaled to the institution's risk profile, as a clear expectation. NCUA's own control framework lists penetration testing explicitly as a distinct control.

NCUA examines credit unions through its Information Security Examination (ISE) program, which scales expectations to asset size: a streamlined track for the smallest institutions, a standard track above roughly \$50 million in assets, and an optional higher-rigor track that places stronger emphasis on penetration testing for larger or higher-risk institutions. Examiners consistently look for testing tailored to the institution's specific risk profile, documented remediation, and evidence that findings were actually closed rather than simply reported.

One time-sensitive note for financial-institution leaders: the FFIEC retired its Cybersecurity Assessment Tool in August 2025. It did not prescribe a single successor. Institutions may use other recognized government and industry resources to assess cybersecurity preparedness, such as the NIST Cybersecurity Framework 2.0, the CIS Controls, CISA Cybersecurity Performance Goals, or the Cyber Risk Institute Profile.

Other Frameworks You May Encounter

Depending on your activities, several other frameworks impose testing expectations. A financial institution that processes card payments is subject to PCI-DSS in addition to its prudential supervision.

PCI-DSS (Payment Processing). The most explicit of the frameworks. Requires internal and external penetration testing at least annually and after any significant infrastructure change, with defined coverage of the cardholder data environment. This applies to any organization that stores, processes, or transmits card data, credit unions included.

HIPAA (Healthcare). The Security Rule requires periodic technical and non-technical evaluation of security controls. It does not name penetration testing; testing can form part of that evaluation process when appropriate to the organization's risk.

GDPR (Personal Data, EU). Article 32 requires a process for regularly testing and evaluating the effectiveness of security measures protecting the personal data of EU residents. Penetration testing is one method organizations may use to support that obligation.

SOX (Public Companies). Financial reporting systems fall under Sarbanes-Oxley controls, and penetration testing is routinely used to validate the security of those environments.

SOC 2 (Technology and Service Providers). SOC 2 does not require a penetration test, although testing results are commonly used as supporting evidence of a mature security posture during the audit.

CMMC (Defense Contractors). The Cybersecurity Maturity Model Certification framework includes penetration testing at Level 3, requiring it at least annually or when significant security changes occur. As of July 2026, however, CMMC implementation is paused in Phase I while the Department of Defense reviews later implementation phases, with Phase I focused on Level 1 and Level 2 self-assessment. Organizations should confirm the requirements applicable to their specific contracts and the current program phase.

Cyber Insurance as a De Facto Regulator

Beyond formal regulation, cyber insurance has become an increasingly influential factor in security expectations for mid-market organizations. During underwriting, insurers may request or consider evidence of penetration testing, documented remediation, and specific technical controls when evaluating coverage and pricing. Requirements vary by insurer and policy, so treat testing as something that strengthens your position rather than a guaranteed universal mandate.

Organizations without a mature testing program may face higher premiums, coverage exclusions, or outright denial of coverage. In the event of an incident, insurers may also examine testing history and remediation records when evaluating claims. A well-documented testing program can meaningfully influence both the availability and cost of coverage. In a market where cyber coverage has become both more expensive and harder to secure, that is a material business consideration.

Even without direct regulatory pressure, many customers and partners require evidence of penetration testing as a condition of doing business with you. This contractual pressure is accelerating across industries.

Compliance-driven testing and security-driven testing work together. A well-executed penetration test satisfies the auditor and genuinely improves your security posture at the same time. The failure mode is treating compliance as the ceiling rather than the floor.

What Is a Penetration Test?

A penetration test is an authorized, simulated attack on an information system, conducted by skilled professionals, for the purpose of identifying security vulnerabilities before a malicious actor can exploit them. The key word is *authorized*. Penetration testing is legal, contracted, and bound by a formal written agreement. Without that authorization, it is simply hacking.

Formal Definitions

An authorized simulated attack performed on a computer system to evaluate its security.

Common industry definition.

Penetration testing is security testing in which assessors mimic real-world attacks to identify methods for circumventing the security features of an application, system, or network. NIST SP 800-115.

A specialized type of assessment conducted on information systems to identify vulnerabilities that could be exploited by adversaries. NIST SP 800-53.

What these definitions share is a focus on simulation, control, and purpose. A penetration test is a methodical, bounded process designed to produce specific, actionable information for the organization that commissioned it.

Limitations matter as well. A penetration test is a time-bound assessment of a defined scope, and it does not certify permanent security. Environments change, and a test conducted today reflects today's posture. It also does not substitute for a broader security program.

What Penetration Testing Is Not

The terms "penetration testing," "ethical hacking," "red teaming," "purple teaming," "vulnerability assessment," and "network scanning" are routinely confused as interchangeable. They are not. Understanding the distinctions lets you have an informed conversation with any security vendor about what you are actually purchasing.

Ethical Hacking

Ethical hacking is a broad umbrella term for using offensive security techniques in an authorized, controlled, and defensive context. It is best understood as a collection of testing methods. Depending on the objective and scope, ethical hacking may include vulnerability scanning, penetration testing, social engineering, application testing, wireless testing, red teaming, or other adversarial techniques. The goal is to identify weaknesses, validate security controls, and help an organization understand how an attacker might target its systems, people, or processes.

Vulnerability Scan

A vulnerability scan is an automated examination of systems, applications, or network devices for known security weaknesses, missing patches, exposed services, and insecure configurations. It compares observed conditions against databases of known vulnerabilities and produces a list of potential findings. A scan identifies where weaknesses may exist. It does not confirm exploitability, determine business impact, or demonstrate what an attacker could accomplish. Its results require review and validation.

Vulnerability Assessment

A vulnerability assessment is a natural extension of a vulnerability scan. It adds a layer of analysis by reviewing, validating, and prioritizing identified weaknesses based on factors such as severity, exposure, asset criticality, existing controls, and potential business impact. The goal is to separate meaningful risk from scanner noise and determine which findings require remediation first. A vulnerability assessment provides a more accurate picture of the organization's security posture, though it generally stops short of active exploitation and does not demonstrate how multiple weaknesses could be combined by an attacker.

Purple Team

In a purple team exercise, specific tactics, techniques, and procedures (TTPs) are agreed upon prior to testing. These TTPs are executed by the offensive team, and the defensive team records what is observed. The two teams work together in real time through transparent, interactive protocols. The primary goal is to better understand the capabilities of defensive controls and uncover gaps in visibility. Purple teaming focuses on improving the defensive team's ability to detect and respond to attacks. It is primarily a maturity-building exercise.

Red Team

Red team operations are advanced exercises for mature environments. They simulate sophisticated, long-running adversaries and test the entire security program including people, processes, and detection capabilities. These engagements often take place over weeks or months, frequently without the internal security team's knowledge. Red teaming answers a different question. Instead of asking what vulnerabilities exist, it asks whether the organization would know if someone made their way inside, and what they could accomplish once there. Red team engagements are expensive and are generally only appropriate for organizations with mature security programs that have already addressed foundational vulnerabilities.

Bug Bounty

A bug bounty program is an open, standing invitation for independent researchers to find and report vulnerabilities in exchange for a reward. It can surface issues continuously and draw on a wide range of talent, which makes it a useful complement to structured testing. But it is not a substitute for a penetration test. A bug bounty can have a tightly defined scope, but it does not guarantee systematic coverage of that scope. Researchers choose where to spend their effort and report what they happen to find, so there is no assurance that every critical asset or attack path was examined. Organizations sometimes ask whether a bounty can replace a pentest. It cannot. The two answer different questions.

Common Misconceptions

Myth: Penetration testing is the same as hacking. Penetration testing is legal, authorized, and bounded by a formal written agreement. The techniques may overlap; the authorization, intent, and structure are entirely different.

Myth: Penetration testing will disrupt operations. Professional penetration tests are carefully scoped to minimize operational impact. Interruptions are possible; however, most activities are non-destructive by design, and any action with real disruption potential requires explicit authorization. Scheduling, exclusions, and escalation protocols are established before testing begins.

Myth: Pentesters are rogue hackers. Most professional penetration testers are methodical engineers and researchers who follow structured frameworks, document everything, and produce formal written deliverables they must defend to both technical and non-technical audiences.

Myth: All security testing is the same. A vulnerability scan, a penetration test, a red team engagement, and a bug bounty are distinct activities with different methodologies, different outputs, and different appropriate use cases. Treating them as interchangeable creates coverage gaps and misaligned expectations.

Who Needs to Be Involved

Penetration testing touches multiple stakeholders across the organization, each with different interests and different relationships to the findings. Organizations that treat it as an IT-only function consistently get less value from it than those that engage the right people throughout the process.

The Board. The board is the key governance stakeholder. Directors have a fiduciary responsibility to ensure the organization is managing risk appropriately, and cybersecurity risk sits squarely within that duty. The board does not need technical detail. It needs to understand overall risk posture, the business implications of key findings, and whether management is investing appropriately to address them.

Executive Leadership. The CEO, CFO, COO, and their peers translate board expectations into resource decisions. They need to understand the business implications of findings, what investment is required to address them, and how the testing program supports strategic objectives. The executive summary of the report should be written with this audience in mind. If it reads like a technical manual, ask for a different firm.

Internal Audit. May independently sponsor, commission, or review testing as part of its assurance role. To preserve independence, it should not own remediation or the management of the security program it is expected to assess.

Risk. Uses testing results to challenge assumptions, inform enterprise risk reporting, and track material exposure over time.

Security (CISO). Typically coordinates security testing and drives remediation. Reporting structure varies, but the function needs enough authority to challenge the technology organization whose controls it evaluates.

IT Teams. IT provides the access, context, and technical support needed to run the engagement, and is generally responsible for implementing many of the technical remediation recommendations. They need detailed findings, specific guidance on addressing each

vulnerability, and clarity on prioritization. A good penetration test should produce a work list the internal team can begin acting on immediately.

Compliance. Penetration testing can provide important evidence that security controls are being independently tested, particularly where applicable standards or regulatory expectations call for it. Compliance should be involved in interpreting findings against applicable obligations and in documenting the engagement for examinations, audits, and customer inquiries.

Legal and Privacy. Legal counsel should review the rules of engagement before testing begins. If the test reveals evidence of a past or ongoing breach, legal needs to be involved immediately, as most jurisdictions have breach notification requirements with tight timelines. Legal should also review data handling provisions in the engagement agreement with the testing firm.

Marketing and Business Development. Security posture is increasingly a commercial differentiator. Evidence of regular penetration testing and documented remediation supports customer acquisition, contract negotiations, and vendor qualification, particularly in sectors where customers are themselves regulated or security-conscious.

Vendors and Third Parties. If your organization relies on external service providers whose platforms fall within or adjacent to the testing scope, their coordination and explicit written permission may be required. Knowing which components cannot be tested, and what that leaves unexamined, is an important output of the scoping process.

Customers and Partners. While not directly involved in the test, many customers and partners will ask for evidence of security testing as part of their vendor risk management process. A recent penetration test with documented remediation is a meaningful and increasingly expected response to those requests.

If you are the executive sponsor of a penetration testing engagement, your most important contributions are at the beginning: ensuring the right people are in the room for scoping, and ensuring the findings are treated as a prioritized action plan rather than a document that gets filed.

Understanding the Variables

Before any penetration test begins, three fundamental questions need to be answered. How you answer them shapes everything about what the test will and will not reveal. These are business

decisions as much as technical ones, and business leaders should be actively involved in making them.

Variable 1: How Much Will You Tell the Tester?

The amount of information provided to the testing team before they begin is one of the most significant design decisions in any engagement. It is commonly described using the analogy of box colors.

Black Box. No prior knowledge. The tester is given nothing but the name of the target. They approach the engagement exactly as an external attacker would, starting from zero. Provides the most realistic simulation of an outside threat, but is the most time-intensive, least budget-efficient, and highest-risk approach for causing unintended disruption. Best for validating external exposure and perimeter defenses.

White Box. Full knowledge. The tester is given everything: architecture diagrams, source code, configuration details, and credentials. Allows testers to focus directly on vulnerabilities. Thorough and efficient. Best for in-depth application testing, internal threat simulation, and code review.

Gray Box. Partial knowledge. The most common real-world approach. The tester is given some context (a user account, a network diagram, or knowledge of the technology stack) without full access. Simulates a threat actor who has gained a foothold, and offers strong cost effectiveness by focusing tester time where it produces the most value. Best for balanced assessments that prioritize efficiency without sacrificing realism.

There is no universally correct choice. Organizations with mature programs often run a combination over time: external black box testing to simulate outside threats, and white or gray box assessments of critical internal systems.

Variable 2: Where Do They Start, and Who Knows?

Two additional dimensions work alongside the knowledge variable and are equally important.

External vs. Internal. An external test begins outside your network perimeter, approaching from the internet as a remote attacker would. An internal test begins from a position inside the perimeter, simulating a malicious insider, a compromised vendor, or an attacker who has already established a foothold. It may be conducted as an assumed-breach exercise, where initial compromise is treated as a given and the test focuses on lateral movement, privilege

escalation, and access to critical assets. These answer different questions. External tests ask whether an attacker can gain initial access; internal tests ask how far they could go and what they could reach once they were in. The most common failure is organizations that test only their external perimeter and assume that is sufficient.

Overt vs. Covert. An overt test is conducted with the knowledge of the internal IT and security team. A covert test is conducted with knowledge limited to a select group of individuals on a need-to-know basis. Sometimes this deliberately includes one or two security staff members so that alarms, escalations, and executive fire drills do not disrupt the engagement, though the broader security team is not informed. Covert testing answers a fundamentally different question. Beyond whether vulnerabilities exist, it asks whether your own team would detect and respond to an actual attack. This is a business and organizational judgment, not a technical one.

Variable 3: What Engagement Model Is Appropriate?

Beyond knowledge and positioning, the overall structure and depth of the engagement matters. These models represent a spectrum from targeted and efficient to broad and comprehensive.

Focused Penetration Test. Scoped, time-boxed, formal report with severity-rated findings and remediation guidance. The standard engagement for most organizations.

Comprehensive Penetration Test. Broader scope, multiple target types, extended timeframe. For organizations seeking a fuller picture across multiple systems and attack surfaces.

Red Team Engagement. Full-scope adversarial simulation. Long duration, typically covert, and requires a substantial budget. Tests detection and response as well as vulnerability presence. Appropriate for mature security programs.

Purple Team Exercise. Collaborative offensive and defensive exercise. Builds detection capabilities in real time. A maturity tool.

Bug Bounty Program. Open invitation to independent researchers. A useful complement to structured testing. Does not guarantee systematic coverage, a coordinated testing timeline, or the formal reporting and remediation process of a penetration test.

For most organizations, particularly those in the small-to-mid-size range, a well-scoped penetration test is the right starting point. Red team engagements are an appropriate evolution once a baseline of security maturity has been established.

What Are We Testing?

With the knowledge variable and engagement model established, the remaining core question is target selection: which parts of your environment are you testing? Penetration testing has evolved into a collection of specialized disciplines, each focused on a distinct part of the technology stack. Understanding these categories allows you to map your testing program to your actual risk profile.

Network Penetration Testing

Network testing examines the security of the infrastructure connecting your organization: routers, switches, firewalls, servers, remote access gateways and VPN appliances, and the connections between them. It divides naturally into external network testing, which approaches from the internet and assesses public exposure, and internal network testing, which assesses what an attacker can accomplish from inside the perimeter. External testing evaluates whether an attacker can gain an initial foothold; internal testing evaluates the potential extent to which an attacker could compromise the organization once inside. Internet-facing remote access infrastructure deserves particular attention, since a single unpatched perimeter appliance can expose the entire environment.

Web Application Penetration Testing

Web applications, meaning the sites, portals, and platforms your organization operates, are among the most commonly exploited attack vectors in modern breaches. They are publicly accessible by design, handle sensitive data, and are frequently built under time and resource pressure that creates security gaps. Web application testing examines authentication, session management, data validation, access controls, and backend interactions. If your organization runs or hosts any customer-facing web application that handles sensitive information, this should be a regular testing priority.

An important distinction: this applies to web applications your organization operates. Do not independently test a third-party SaaS provider (online banking portals, payroll systems, CRM platforms) without explicit authorization; some vendors permit certain customer testing, but many do not. In most cases your role is to obtain assurance through vendor risk management rather than to test the vendor's environment yourself.

API Penetration Testing

APIs are the connective tissue of modern software. They allow your applications to communicate with each other, with third-party services, and with your customers' systems. As organizations have moved toward cloud-based architectures and third-party integrations, APIs have become a major and frequently under-tested attack surface. API testing examines authentication, authorization, input handling, and the sensitivity of data exposed through these interfaces.

Cloud Security Testing

Misconfiguration remains a common source of cloud exposure. Storage buckets left publicly accessible, identity and access management policies that are far too permissive, and inter-service connections that violate the principle of least privilege are widespread and routinely undiscovered. Cloud security testing examines these configurations and evaluates how far an attacker could move laterally within the environment.

Cloud security testing is often described loosely as "cloud penetration testing," though the term deserves qualification. Major cloud providers (AWS, Azure, Google Cloud) impose specific rules of engagement on what testing customers are permitted to perform against their infrastructure. Much of what is valuable in a cloud security assessment is configuration review, identity and access analysis, and controlled testing of customer-built workloads, rather than the broad, unconstrained probing implied by traditional penetration testing. The shared responsibility model, where the provider secures the infrastructure and the customer secures what they build on it, means many organizations are unclear on what their obligations actually are. A good cloud security assessment clarifies that.

AI and LLM System Testing

As organizations adopt AI features, chatbots, document summarizers, and models that inform lending, fraud, or hiring decisions, those systems become a new attack surface with failure

modes traditional testing was not built for. Prompt injection, training-data leakage, and manipulation of model outputs are real and largely untested risks. This matters even when the AI is supplied by a vendor: if a compromised or manipulated model touches member data or drives a decision, the accountability is still yours. Testing here is an emerging discipline, but any organization deploying AI in a sensitive workflow should be asking whether, and how, it gets tested.

Social Engineering

Not all attacks target technology. Social engineering tests the human element by simulating techniques attackers use to manipulate employees into taking actions that compromise the organization: clicking phishing links, providing credentials over the phone, or granting physical access to unauthorized individuals. Phishing simulations are the most common form. Social engineering also includes vishing (voice phishing), smishing (SMS-based attacks), and pretexting scenarios. Particularly valuable for evaluating the effectiveness of security awareness training.

Physical Penetration Testing

Physical testing evaluates whether an unauthorized person can reach secure areas: server rooms, workstations, network equipment, or sensitive files. Organizations focused on cyber threats routinely overlook it, but physical access bypasses many technical controls entirely. A tester who walks into your server room and plugs in a device does in minutes what a remote attacker might spend weeks attempting.

Wireless Penetration Testing

Wireless testing examines Wi-Fi infrastructure for rogue access points, weak or outdated encryption, and the risk of an attacker in physical proximity gaining network access. In shared spaces, open campuses, or facilities near public areas, someone in the parking lot can be as dangerous as someone inside the building.

Mobile Application Testing

Mobile apps are now a primary interface for customers and employees alike, and a frequently under-tested one. Mobile testing examines Android and iOS applications for insecure data storage, weak authentication, improper session handling, and unsafe communication with backend services.

Open Source Intelligence (OSINT)

OSINT testing measures how much information about your organization and its people is publicly available, and how an attacker could use it. Job postings, LinkedIn profiles, press releases, and domain records expose more than most organizations realize, and those findings often become the foundation for targeted social engineering and credential attacks.

A Note on Specialized Environments

Beyond the disciplines above, testing extends into specialized environments: operational technology and industrial control systems, embedded devices and IoT networks, mainframes, and more. These demand domain expertise distinct from conventional IT security, and testing them incorrectly can cause operational disruption or physical risk. If you operate any such environment, vendor selection matters even more and the rules of engagement require additional care.

How a Penetration Test Works

While the specific approach varies by provider, engagement type, and scope, most penetration tests follow a structured lifecycle. Understanding that lifecycle helps you set expectations, participate meaningfully in scoping, and evaluate whether the engagement is being run well.

Phase 1: Scoping and Rules of Engagement

No reputable penetration test begins without a clearly defined scope. Scoping is the process of establishing exactly what will be tested, under what conditions, and within what boundaries. This is where business leader involvement matters most. Scope decisions determine what risk you are measuring and, critically, what risk you are leaving unexamined.

Scope decisions include which systems, applications, or networks are included and excluded; the testing window and timeframe; whether testing will be overt or covert; time-of-day constraints; and what constitutes acceptable risk during the engagement. The rules of engagement document formalizes all of this. It is a legal instrument and an operational guide, and both your legal team and IT team should review it before testing begins.

Phase 2: Reconnaissance and Discovery

Before any active testing begins, the team collects information about the target environment. In a black box engagement, this involves significant open-source intelligence gathering: public records, domain registrations, job postings, technology stack signals, and organizational information visible from outside. In white or gray box engagements, some information is provided directly. Discovery also includes active network scanning to identify live systems, open ports, and running services.

Phase 3: Enumeration

Enumeration goes deeper than discovery. Where discovery identifies that a system exists, enumeration determines what it is running and what specific vulnerabilities or misconfigurations may be present: software versions, configuration details, user account structures, and service-level information that can be used to map potential attack paths.

Phase 4: Exploitation and Validation

This is the phase most people picture. Exploitation is where the tester actively attempts to leverage identified vulnerabilities to gain unauthorized access, escalate privileges, move laterally, or access sensitive data. Not every identified vulnerability will be exploited. A reputable tester follows a measured approach, attempting only what is authorized, and flagging situations where further exploitation would create unacceptable risk to production. The goal is to validate that vulnerabilities are genuinely exploitable and understand real-world impact.

Phase 5: Reporting

The report is the deliverable, and its quality is one of the clearest signals of the quality of the firm that produced it. A good report communicates findings at both the technical and executive level. At minimum it should include an executive summary a non-technical leader can read and act on; severity-rated findings with evidence; specific actionable remediation guidance; and a clear distinction between confirmed exploitable vulnerabilities and those identified but not tested.

The report is the beginning of remediation. A penetration test that produces a report nobody acts on has delivered no security value. Treat it as a prioritized work plan.

Real-World Constraints

Penetration testing in practice rarely happens in ideal conditions. Organizations have operational realities that shape what is possible, and the value extracted depends on navigating those constraints deliberately.

Budget

Testing costs vary considerably based on type, scope, and depth. A focused test of a single application or defined external perimeter sits at the lower end of the spectrum. A comprehensive network penetration test with multiple target types and internal coverage sits meaningfully higher. Red team engagements, with their extended duration and specialized skill requirements, are the most expensive of the three by a significant margin. The variation reflects real differences in expertise required, time invested, and depth of findings.

A useful analogy: a general practitioner handles a wide range of health concerns competently and affordably. A cardiac surgeon costs more because the expertise is more specialized and the stakes are higher. Organizations with constrained budgets should prioritize testing around their most critical systems. A thorough test of what matters most delivers more value than a shallow test of everything.

Time

Penetration testing is a point-in-time exercise. It reflects the security posture of the environment at the moment it was conducted, and the gap between tests is a window of unknown risk. Testing frequency should track your rate of change and your risk profile, a point covered further under Best Practices.

Disruption

As noted earlier, well-executed testing rarely disrupts production. Where disruption is a genuine concern, standard mitigations keep it contained: testing during maintenance windows, excluding specific production systems from active exploitation, and staging higher-risk activities in non-production environments.

Staff Availability

Penetration tests require coordination. Someone on the internal team needs to be available to provide access, answer questions from the testing team, and respond if something unexpected

occurs. Organizations that underestimate this coordination burden often end up with tests that are slower, shallower, or less useful than they should be. Assign a clear internal point of contact before the engagement begins.

Third-Party Dependencies

Most organizations rely on SaaS vendors, managed service providers, cloud platforms, and third-party integrations that sit within or adjacent to the attack surface. Testing these components requires explicit written permission from the third party, which can be time-consuming or impossible to obtain. Knowing which parts of your environment cannot be tested, and what that leaves unexamined, is an important part of interpreting your results.

The Compliance Checkbox Problem

One of the most common failure modes is organizations that conduct penetration testing solely to satisfy an audit requirement, with an implicit goal of producing a result that clears the compliance hurdle without surfacing anything requiring significant remediation. Tests scoped this way are deliberately narrow, avoid aggressive techniques, and leave meaningful risk undiscovered. Compliance-driven testing and security-driven testing should be the same engagement. When they are not, someone made a deliberate choice to know less than they could.

Scope concessions are sometimes necessary, though they should be made deliberately, with a clear understanding of what risk they leave unexamined. A test with a highly constrained scope tells you something. You should know exactly what it does not tell you.

Choosing the Right Partner

Good penetration testers are, at their core, methodical engineers and careful documentarians. They follow structured processes, research extensively, and produce written work product that has to hold up to scrutiny from both technical and non-technical audiences. The value of a penetration test is inseparable from the quality of the people conducting it.

Technical Credentials: What to Look For

Hands-on credentials are evidence of capability, not a gate. Practical, performance-based certifications help confirm real skill, but none substitutes for relevant experience, sound

methodology, references, and sample work product. The certifications that carry the most weight require demonstrated practical skill rather than a passing score on a written exam. Treat the list below as useful signals, not pass-fail requirements.

- **OSCP / OSCP+ (Offensive Security).** A widely respected practical credential. Candidates conduct a live, timed penetration test against a real network and submit a professional report. Strong evidence of hands-on capability, though not a substitute for experience.
- **OSEP / OSED / OSWE.** Advanced Offensive Security certifications requiring 48-hour practical exams. Highly respected for depth and hands-on focus. Signal specialized expertise in evasion, exploit development, and web application security respectively.
- **GPEN / GWAPT (GIAC Certifications).** Well-regarded certifications, now including hands-on performance testing, from the SANS Institute covering network penetration testing and web application security. Reflect both training rigor and practical assessment.
- **PNPT (Practical Network Penetration Tester).** A newer but well-respected certification emphasizing practical testing skills and professional report writing.

Technical Credentials: What NOT to Be Fooled By

Not every widely-recognized security certification signals penetration testing capability. Some are excellent credentials for what they cover, though they are not evidence that the holder can conduct a competent penetration test. If a firm leads with any of these as its primary technical qualification, dig deeper.

- **CISSP (Certified Information Systems Security Professional).** A highly respected and widely held certification for information security management. It is a broad governance and program-management credential, not evidence that someone can perform a penetration test. The credential is legitimate; it simply is not a testing qualification.
- **CEH (Certified Ethical Hacker).** Widely marketed and widely held, though the practical requirements are far less demanding than the certifications listed above. A useful baseline signal but not sufficient on its own as evidence of deep testing capability.
- **Vendor-specific product certifications.** Certifications tied to specific security products (firewalls, SIEM platforms, EDR tools) demonstrate operational competence with those products. They are not evidence of offensive testing skill.

Certifications are a starting point. Experience matters enormously. Ask specifically about the types of environments tested, industries worked in, and the seniority of the individuals who will actually conduct your engagement, rather than the firm's credentials at a headline level.

Report Quality

Ask to see a sample report from a previous engagement, with client information redacted. A good report is the clearest window into how a firm actually works. If the executive summary reads like a technical manual, remediation guidance is vague, or severity ratings appear inconsistent, those are meaningful signals about the quality of the work. The report is the product. Evaluate it accordingly.

Manual vs. Automated Testing

Automated tools are useful for coverage breadth. They cannot reason about complex attack chains, business logic flaws, or creative exploitation of misconfiguration combinations. Ask specifically what portion of the test is manual. A firm whose methodology is primarily automated is running a vulnerability scan at penetration test prices.

Artificial intelligence has raised the stakes on this question. AI-driven tooling genuinely helps testers work faster and cover more ground, and a good firm uses it well. But AI also makes it easier than ever to dress up an automated scan as a penetration test. If a firm leans on AI, ask what it actually does: whether it augments skilled human testers who validate and extend its findings, or whether it substitutes for them. The distinction is the difference between leverage and a shortcut you are paying premium prices for.

Ethics, Professionalism, and Continuous Learning

Penetration testers are given deliberate access to sensitive systems and data. Look for firms with formal rules of engagement processes, clear data handling policies, non-disclosure agreements, and a track record of professional conduct. Check references. The best firms invest in staying current through research, lab environments, and ongoing professional development. It is reasonable to ask how a firm's testers stay current and what professional development looks like internally.

Best Practices: Getting Real Value

The difference between organizations that get genuine security value from penetration testing and those that do not is rarely budget. It is the discipline with which they approach the engagement before, during, and after.

Define your objective before you define your scope. Are you testing to satisfy a compliance requirement, validate a specific system, understand your overall risk posture, or test your incident response capabilities? The answer should drive every subsequent decision. Testing without a clear objective produces findings that are technically interesting but operationally inert.

Scope deliberately, not defensively. Scope should be driven by what matters most to the organization. A narrow scope that avoids your most critical systems tells you very little about your most significant risk.

Choose vendors based on fit. Request sample reports. Check references. Ask specifically about the experience of the individuals who will conduct your engagement. A well-credentialed firm that assigns junior testers is a different proposition than a smaller firm whose senior people are doing the work.

Align stakeholders before the engagement begins. Legal should review the rules of engagement. IT should be informed and available. Leadership should understand what is being tested and why. Surprises during or after a test are almost always the result of inadequate alignment at the outset.

Treat the report as a work plan. A report that is filed without action is an expensive way to document risk you already had. Prioritize findings by severity and exploitability, assign remediation ownership, set deadlines, and consider retesting critical fixes to validate they have been addressed.

Test regularly, not only when required. Annual testing is a common baseline, particularly in regulated and higher-risk environments. Cadence should also respond to significant change and risk, because every change made after a test is a change that has never been tested.

Debrief and learn from every engagement. After each test, ask what it revealed about your processes as well as your technology. Were there coordination failures? Did findings appear in areas considered low-risk? Use each engagement to improve the next one.

Consider your cyber insurance position. Insurers are increasingly scrutinizing security practices when issuing coverage. A mature testing program with documented findings and remediation can influence both the availability and cost of coverage. Organizations with strong testing histories may qualify for better terms; those without may face higher premiums or coverage exclusions.

Ten Questions to Ask Your Testing Firm

These questions separate firms that deliver genuine security value from firms that deliver paper. Ask them during vendor selection, and pay attention to how specifically each one is answered. Vague answers to precise questions are a signal in themselves.

- 1. What portion of the test is manual versus automated?** A methodology that is primarily automated is a vulnerability scan at penetration test prices.
- 2. Who will actually conduct our engagement?** Ask about the specific individuals: their certifications, experience, and seniority. A well-credentialed firm that assigns junior testers is a different proposition than a firm whose senior people do the work.
- 3. Can we see a sample report with client information redacted?** The report is the product. If the executive summary reads like a technical manual, keep looking.
- 4. How do you approach scoping, and what do you need from us?** A firm that pushes toward a quick, narrow scope without understanding your environment is optimizing for its own delivery cost rather than your risk picture.
- 5. What happens if you find evidence of an active or past breach?** The answer should include immediate notification, defined escalation contacts, and coordination with your legal counsel.
- 6. How is our data handled during and after the engagement?** Look for defined retention periods, encryption standards, and secure destruction practices.
- 7. What remediation guidance is included, and do you offer retesting?** Findings without actionable guidance simply create work for your team. Retesting validates that critical fixes actually closed the gap.
- 8. How do your testers stay current?** New techniques and vulnerabilities emerge constantly. Research time, lab environments, and ongoing professional development keep a firm's testing relevant.
- 9. How does the firm use automation and AI?** Both have legitimate roles in modern testing. The question is whether they supplement expert manual work or substitute for it.
- 10. What does the deliverable look like for the board versus the technical team?** A good firm communicates fluently at both levels. If it can only speak to one audience, half the value of the engagement is lost.

Conclusion

Penetration testing is a tool. Its value lies in using it well: selecting the right type for the job, deploying it at the right time, and acting on what it produces.

For business leaders who are not security specialists, the objective is to develop enough literacy to ask the right questions, set the right expectations, and make informed decisions about where and how to invest in security testing. The organizations that get the most value from penetration testing approach it with clear objectives, choose qualified partners, engage thoughtfully in the scoping process, and treat the report as the beginning of a remediation effort.

The threat landscape will continue to evolve. New attack techniques, new technologies, and new vectors will continue to emerge. Penetration testing is an ongoing practice, calibrated to the organization's risk profile and revisited as that profile changes. Organizations that embed that practice into their security culture as a genuine commitment to understanding their posture will be substantially better positioned than those that treat it as an obligation to be minimized.

Ultimately, the objective is to be reasonable. Protect the systems and data that matter most, at a level proportionate to what their compromise would cost you. That standard does not eliminate risk, and it is not meant to. It is meant to close the distance between what you believe about your security and what is actually true, because that distance is where breaches live.

Quick Reference Glossary

Key terms used throughout this guide.

API (Application Programming Interface)	A protocol allowing different software applications to communicate. A primary and frequently under-tested attack surface in modern organizations.
Assumed Breach	A testing approach that begins from a position of assumed initial compromise, evaluating what an attacker could accomplish from inside the perimeter.

Black Box Testing	A penetration test in which the tester begins with no prior knowledge of the target environment, simulating an external attacker.
Blue Team	The internal defensive security team responsible for detecting and responding to attacks.
Bug Bounty	An open invitation to independent researchers to find and report vulnerabilities for monetary reward. A complement to structured penetration testing, not a substitute for it.
CISSP	Certified Information Systems Security Professional. A widely respected information security management credential. Not a penetration testing certification.
Covert Testing	A penetration test conducted without the knowledge of the broader internal IT and security team, designed to evaluate detection and response capabilities.
CVE (Common Vulnerabilities and Exposures)	The standard public identifier assigned to a disclosed software vulnerability. A CVE number allows defenders, vendors, and security tools to reference the same flaw consistently.
Due Care	Acting responsibly by implementing reasonable protective measures. A governance and risk-management concept, not a universal legal standard.
Due Diligence	Actively investigating risks and verifying that protective measures are appropriate and effective. A governance and risk-management practice that complements due care.
Ethical Hacking	The broad practice of using offensive security techniques for defensive purposes. Penetration testing is a constrained, structured subset.
Gray Box Testing	A penetration test in which the tester has partial knowledge of the target environment. The most common real-world approach.
OSCP	Offensive Security Certified Professional. A widely respected penetration testing certification requiring demonstrated practical skill under exam conditions.

OSINT (Open Source Intelligence)	Information gathered from publicly available sources. Used during the reconnaissance phase of testing.
Overt Testing	A penetration test conducted with the knowledge of the internal IT and security team.
Purple Team	A collaborative exercise where offensive and defensive teams work together to build detection and response capabilities.
Ransomware	Malicious software that encrypts systems or data and demands payment for restoration. Modern operators typically also steal data before encrypting it, threatening publication if the ransom is unpaid.
Red Team	A team conducting full-scope adversarial simulations, typically covert, to test an organization's entire security program.
Rules of Engagement	The formal written agreement defining scope, constraints, and authorization for a penetration test. A legal instrument, not a formality.
Scope	The defined boundaries of a penetration test, including which systems, timeframes, and techniques are included and excluded.
Social Engineering	The use of psychological manipulation to deceive individuals into actions that compromise security.
Vulnerability Assessment	An analytical review that prioritizes identified weaknesses based on severity, exposure, and business impact. Does not include active exploitation.
Vulnerability Scan	An automated identification and cataloging of known vulnerabilities. Does not involve active exploitation. Distinct from a penetration test.
White Box Testing	A penetration test in which the tester has full knowledge of the target environment, including architecture, credentials, and source code.

Sources and References

The regulatory and standards claims in this guide draw on the following primary sources.

Requirements and guidance evolve, so leaders should confirm current expectations with their own counsel, auditor, or regulator before relying on any specific point for a compliance decision.

[1] NIST Special Publication 800-115, Technical Guide to Information Security Testing and Assessment. The standard reference for security testing methodology, including penetration testing.

[2] NIST Special Publication 800-53, Security and Privacy Controls for Information Systems and Organizations. Source for control families referenced in assessment scoping.

[3] Gramm-Leach-Bliley Act, Section 501(b), and NCUA Rules and Regulations, 12 CFR Part 748 and Appendix A, Guidelines for Safeguarding Member Information, which address testing of key controls and board reporting for federally insured credit unions.

[4] NCUA Information Security Examination (ISE) program materials, which reference established industry frameworks and standards rather than a single mandated tool.

[5] FFIEC guidance following the August 31, 2025 sunset of the Cybersecurity Assessment Tool, under which no single successor is prescribed and institutions may use recognized resources such as NIST CSF 2.0, the CIS Controls, CISA Cybersecurity Performance Goals, and the Cyber Risk Institute Profile.

[6] PCI DSS v4.0.1, Requirement 11.4, which requires external and internal penetration testing on a defined schedule for entities handling payment card data.

[7] HIPAA Security Rule, 45 CFR Part 164, which requires periodic technical and non-technical evaluation of security controls; HHS and NIST guidance describe penetration testing as an option when reasonable and appropriate to risk.

[8] EU General Data Protection Regulation, Article 32, which requires a process for regularly testing and evaluating the effectiveness of security measures.

[9] Cybersecurity Maturity Model Certification (CMMC) program, 32 CFR Part 170, under which Level 3 requires penetration testing at least annually or upon significant security change. On July 13, 2026, the Department of Defense suspended CMMC Phase II and paused later

implementation milestones pending a program review; Phase I (Level 1 and Level 2 self-assessment) remains in effect.

[10] AICPA SOC 2 Trust Services Criteria, which do not require penetration testing but commonly rely on testing results as supporting evidence.

About the Author

J.W. Walker has spent his career in information security across commercial and government environments. He has sat in the tester's chair, built a company from the ground up through exit, and serves as CTO of an information security firm. Today he advises executives and boards on how to get genuine value from security testing. He writes plain-language guides that help leaders make confident decisions about technology and risk without needing to become technical experts.

If you found this guide useful, visit jeremywalker.us/resources for more cybersecurity and leadership resources, or reach out directly.

More guides at jeremywalker.us/resources | Connect:
linkedin.com/in/jeremywalker83

The Business Leaders Series

Plain-language guides for the decision-maker, not the technician. Each one takes a subject that is usually buried in jargon and makes it something a leader can actually use. If this guide was useful, the others are written the same way.

Penetration Testing for Business Leaders *(you are reading this one)*

What to test, what to ask, and how to turn findings into action.

Cybersecurity for Business Leaders: CU Edition

Cyber risk, oversight, and member protection, written for the board.

Cybersecurity for Business Leaders: SMB Edition

Cyber risk, protection, and practical steps for owners. Written with Ideal Technologys.

AI for Business Leaders

Getting real value from AI without getting burned by it.

The Virtual CISO for Business Leaders

What you are really buying when you hire fractional security leadership.

Personal Cybersecurity for Business Leaders

Protecting your identity, money, devices, family, and digital life. Built to share with your team and family.

Find the complete series at **jeremywalker.us/resources**

Last reviewed: August 2026. Regulatory references reflect guidance current as of that date and are refreshed periodically.

This guide references regulatory positions and standards current as of the date below. Because those evolve, confirm specifics with counsel or your regulator before relying on them.

Last reviewed: August 2026