ThreatModeler

RESEARCH STUDY

From Prompt to Proof

The trust gap in AI-driven threat modeling.

THE SHORT VERSION

Executive Summary

AI has made threat modeling easier to start, but not easier to trust. The category is moving away from prompt-centric experimentation and toward governed, architecture-aware systems that make AI-assisted outputs repeatable, reviewable, and defensible.

Adoption is real. Purchased automated threat modeling platforms are already in use at 93.6% of respondents, and 75.6% of teams using AI-assisted threat modeling in manual or internally built environments expect that usage to grow over the next two years. This is a category transition already underway, not a fringe experiment.

Adoption is not the same as trust. Only 32.0% of respondents trust AI-assisted threat modeling a lot or completely, and that figure falls to 16.8% for regulated or safety-critical applications. The barriers are practical, not philosophical: security and privacy, accuracy, validation effort, explainability, and governance.

Threat modeling sits inside that gap because it is not a content-generation task. It is the discipline that turns architecture and system intent into decisions about threats, controls, documentation, ownership, and governance. AI can accelerate parts of that work. The data here suggests organizations do not yet trust it to replace the work.

93.6%
already use a purchased automated platform

32.0%
trust AI-assisted threat modeling a lot or completely

16.8%
that trust, for regulated or safety-critical apps

THE BOTTOM LINE

The next phase is not won by the fastest prompt.

It is won by platforms that operationalize AI inside a deterministic, architecture-aware, governed system of record.

The Market Is in Transition, Not at Equilibrium

The threat modeling market is already modernizing, but it has not standardized. Respondents report high use of purchased automated platforms (93.6%), alongside still-substantial use of manual methods such as spreadsheets and diagrams (72.4%). Many organizations have introduced automation without fully replacing manual work, which leaves fragmented workflows and different operating models across application portfolios.

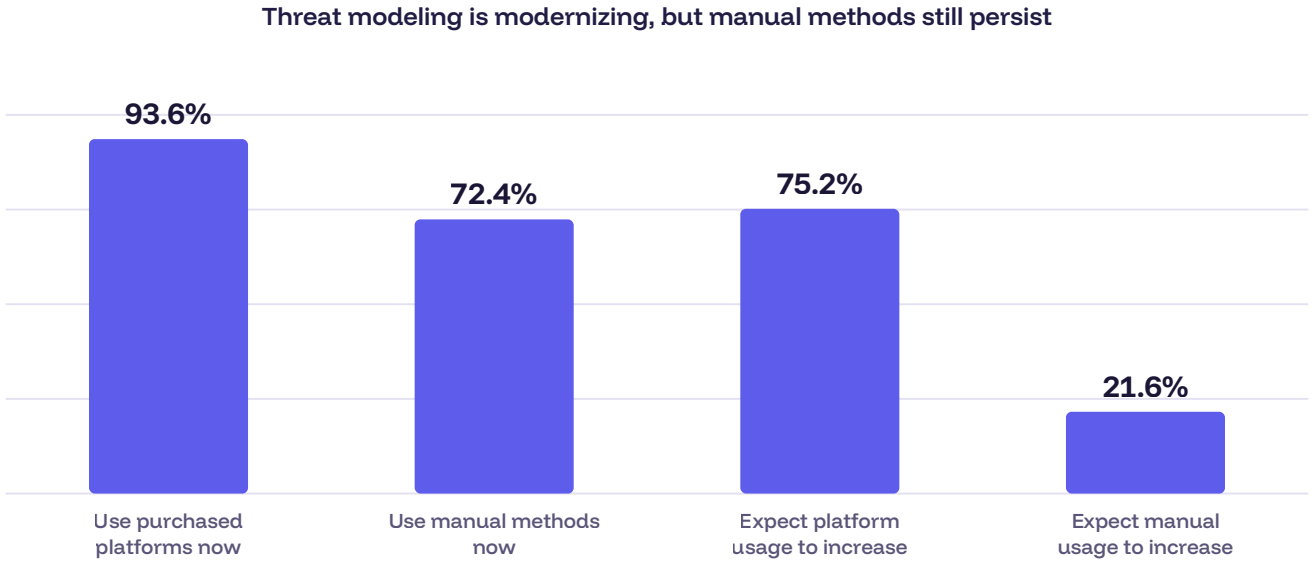

Figure 1. Current use and 12-month outlook, share of respondents.

Why it matters. The category is not switching from old to new all at once. Many enterprises carry both models at the same time, which raises the value of platforms that standardize a process rather than automate isolated tasks. Three quarters of respondents expect platform usage to grow over the next 12 months; only 21.6% expect the same of manual methods. The center of gravity is shifting toward operating models that bring consistency, speed, and portfolio-wide visibility.

That shift reframes the AI question. In a market still mostly manual, AI would mainly reduce labor. In a market already moving to platforms, the question is whether AI can be embedded in a repeatable system. That is a higher standard, less about generative novelty and more about operational trust.

AI Is Inside the SDLC, and Inside Threat Modeling

AI-assisted development is already affecting the design-time security workflow. Among applications that use threat modeling, an average of 34.46% also use AI code generation. Threat modeling happens during AI code generation 44.5% of the time, before it 31.28% of the time, and after it 24.22% of the time.

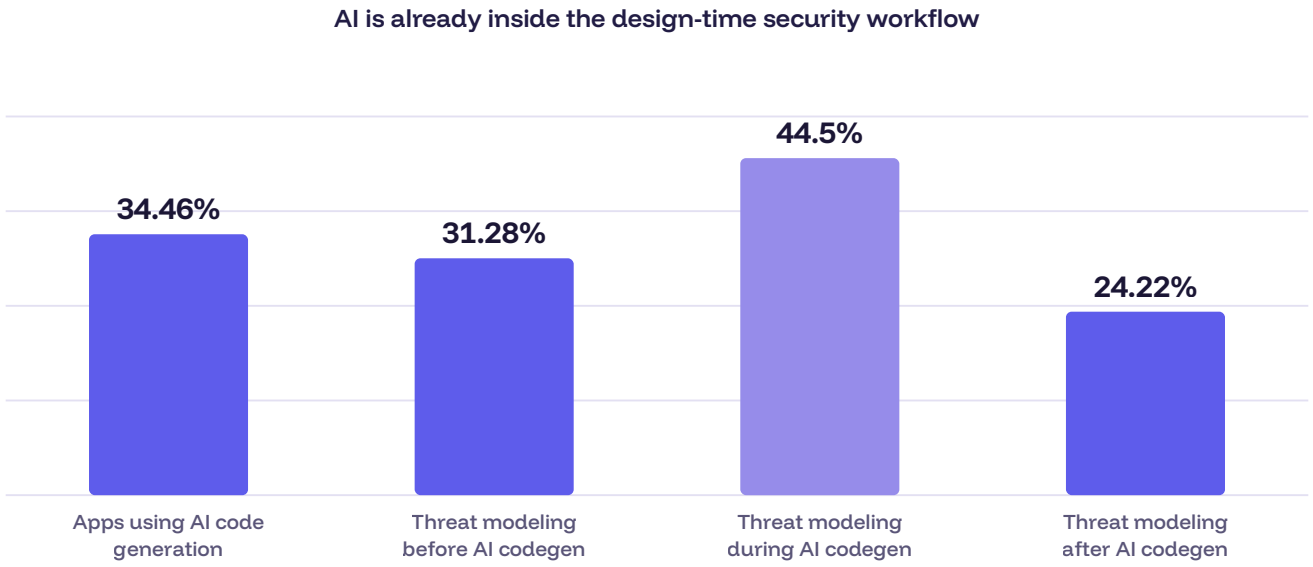

Figure 2. Where threat modeling occurs relative to AI code generation, average share.

Why it matters. Threat modeling is no longer a separate practice running safely outside AI adoption. It is entangled with AI-driven development, which raises the stakes for consistency, auditability, and architectural context. Cloud already made systems harder to reason about, with more services, APIs, identities, and dependencies across hybrid environments. AI adds non-determinism on top: code is generated faster, and agentic systems are assembled without always being grounded in explicit intent.

The useful question moves from "what vulnerabilities exist?" to "what was this system intended to do, where should trust exist, and what could go wrong if that intent is violated?" If AI already participates in architecture description, code generation, and threat identification, the real test is whether those outputs stay coherent, reviewable, and durable over time.

The Trust Gap Is the Market Bottleneck

The strongest signal in the data is not adoption. It is constrained trust. Only 32.0% of respondents trust AI-assisted threat modeling a lot or completely, while 53.2% trust it only somewhat. Trust falls further as stakes rise: 16.8% for regulated or safety-critical applications, and 15.2% for legacy or monolithic environments.

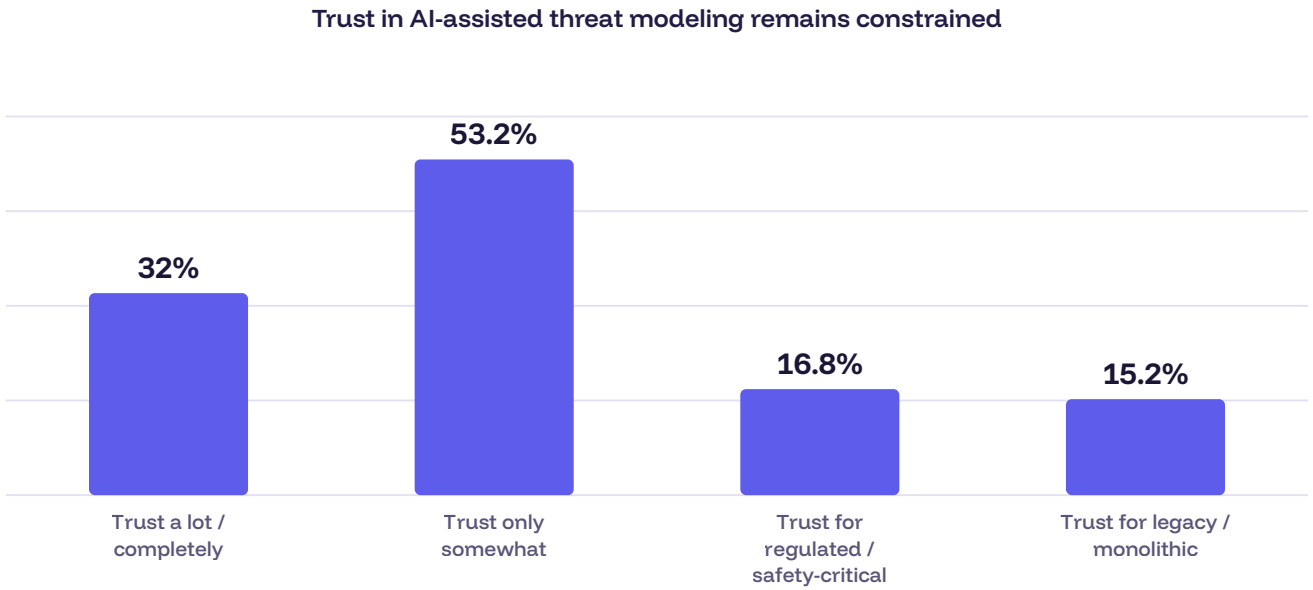

Figure 3. Share of respondents, by trust level and environment.

Why it matters. The problem is not lack of interest. It is lack of assurance. The closer AI gets to regulated, fragile, or business-critical environments, the more teams ask for governance and proof rather than raw speed. AI can draft and summarize quickly, but threat modeling is valuable because it creates confidence in decisions. Where that confidence cannot hold, adoption stalls or stays superficial: teams use AI to get started, not to sign off.

EXPERT TAKE

The trust figures do not show that AI-assisted threat modeling is ineffective. They show that buyers currently treat it as insufficient on its own. Any program that ignores the gap is asking the market to accept speed without assurance.

What Teams Worry About, and What They Want

The barriers are practical. Among organizations already using AI-assisted threat modeling in manual or internally built environments, 57.1% cite data security or privacy, 52.4% cite accuracy or false positives, and 50.5% say outputs still need too much human validation. These are the predictable consequences of running probabilistic systems in a discipline built on context, review, and traceability.

Top barriers to broader AI-assisted threat modeling adoption

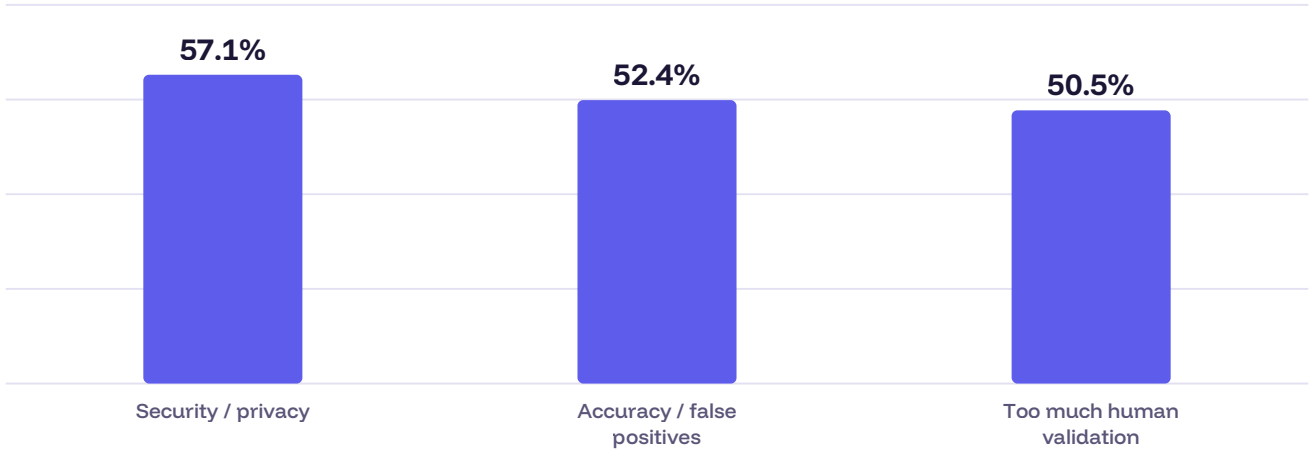

Figure 4. Share of respondents citing each barrier.

What would move a purchase. The buying criteria point the same direction. Teams are not asking for more AI theater; they are specifying the conditions under which AI becomes acceptable for enterprise use.

65%	Strong security and privacy assurances
47%	Clear regulatory or compliance alignment
41%	Transparent, explainable AI outputs
38%	Demonstrated accuracy and validation data
36%	Strong governance and human-in-the-loop controls

More Speed, More Validation

The more teams rely on AI to automate reasoning, the more human expertise is needed to verify the result. This is not a flaw in the technology. It is a mismatch between what generative AI does well, pattern completion and draft generation, and what threat modeling requires: architectural reasoning, control placement, prioritization, and defensible documentation.

This is why prompt-based approaches often disappoint after the first demo. They help with ideation, but they struggle to hold the conditions enterprise security depends on. The weaknesses are structural.

NON-DETERMINISM

The same prompt can return different results. That is fine for creative work and unacceptable for repeatable analysis a team has to defend.

NO ARCHITECTURAL GROUNDING

Direct LLM use is typically disconnected from real cloud architecture, infrastructure-as-code, trust boundaries, and application context.

NO BUILT-IN GOVERNANCE

Prompt transcripts are not a system of record. They do not provide versioning, approvals, or reusable threat-to-control mappings.

VALIDATION BURDEN

Checking whether output is correct still falls to the architect, often recreating much of the work AI was meant to reduce.

The buyer problem is not "can AI generate a threat list?" It is "can AI-assisted security decisions be trusted, repeated, governed, and defended?" Prompt-centric wrappers can show speed; without an answer to the second question, they stay better suited to experimentation than to an enterprise operating model.

06 · THREE PATHS, DIFFERENT LIMITS

Where Build-Your-Own and AI-Only Approaches Break

Organizations now choose among three paths: build internal tooling, use AI or LLMs directly, or buy a purpose-built platform. None is without value. Each has different limits, and those limits become decisive as scale, volatility, and assurance requirements rise.

BUILD-YOUR-OWN

Flexible, but heavy to maintain

Internal tooling aligns with workflows and allows narrow experimentation. The ongoing burden is real: threat content, framework mappings, cloud coverage, and governance all need continuous upkeep. AI speeds prototypes; it does not keep models accurate over time.

AI-ONLY

Low friction, low assurance

Direct LLM use is a fast way to brainstorm threats or summarize behavior. It lacks architectural grounding, drift detection, governance, collaboration, and dependable repeatability, which makes it a poor fit for audit-sensitive, multi-team, fast-changing environments.

PURPOSE-BUILT PLATFORM

Built for the operating model

The strongest fit when teams need modeling grounded in real architecture, consistent workflows, reusable libraries, approvals, integrations, and ongoing maintenance of threat and framework content.

EXPERT TAKEAWAY

The question is not whether AI-only or homegrown methods can create value. They can. It is whether they stay viable when the organization needs continuity, shared standards, and evidence that survives review. At that point the decisive issue is not who drafts fastest. It is who preserves accuracy, reuse, and accountability after the first draft exists.

What the Data Implies Geographically

The market is not homogeneous. Overall interest in an AI-assisted threat modeling solution over the next two years is 47.6% very or extremely likely, but it varies widely by region: 59.2% in the U.S., 51.3% in the EU, 49.0% in Canada, and 25.0% in the U.K.

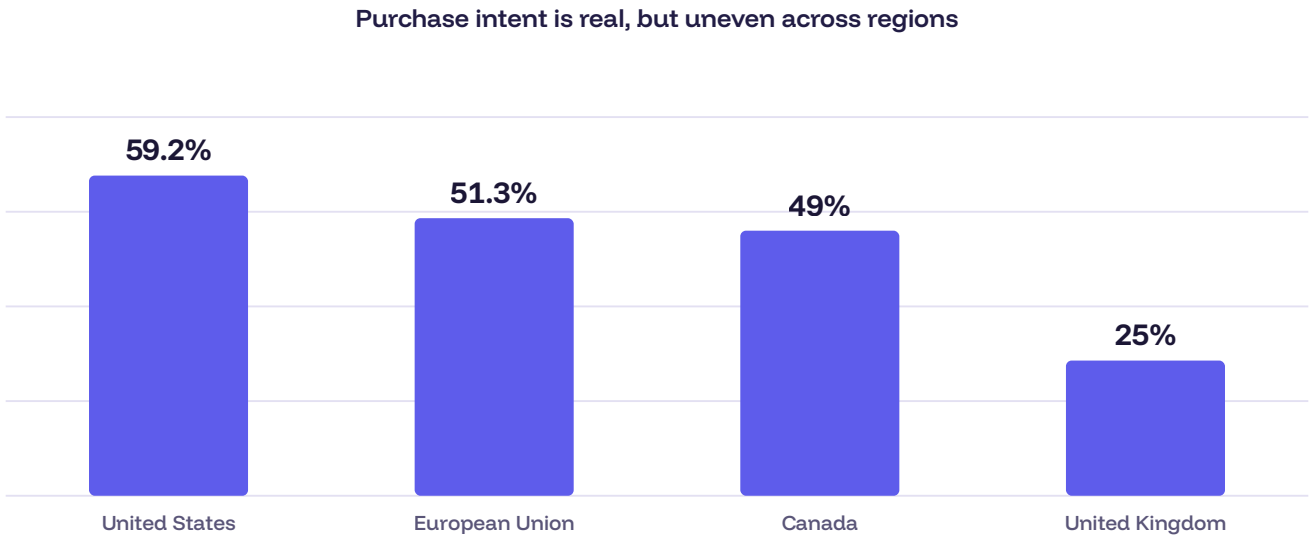

Figure 6. Very or extremely likely to consider purchase, by region.

The pattern reinforces the central finding rather than complicating it. Demand is strongest where regulatory pressure and portfolio scale are highest, and that is exactly where assurance, governance, and architectural context matter most. Speed alone does not clear the bar in those markets.

WHAT THE DATA IMPLIES

The Pattern That Wins

Read together, the findings describe one requirement. The next category winner is not a generic model front end. It is a system that contains AI inside a workflow that satisfies enterprise requirements: privacy, compliance, explainability, validation, and governance. ThreatModeler® Nexus™ is built to that shape.

From prompt to proof: AI inside a system of record.

ARCHITECTURE-AWARE BY DEFAULT

A System Mapping Agent builds the model from diagrams, code, and infrastructure-as-code, and the Secure Design Graph grounds it in real configuration rather than a description. Findings map to components, trust boundaries, and data flows.

DETERMINISTIC AND GOVERNED

Outputs are versioned, reproducible, and explainable. Approvals, change tracking, and audit history are part of the workflow, so every result traces back to the data, framework, or rule that produced it.

AI ON YOUR TERMS

With BYOAI, the deterministic framework governs the outcome regardless of which approved model responds. Sensitive architectural data and AI-derived insight stay inside the enterprise boundary.

ONE SYSTEM OF RECORD

Reusable content, framework mappings, and audit-ready reporting hold a portfolio together as designs and standards change, the continuity the data says teams are buying.

Speed earns the start. Proof earns the sign-off.

HOW TO GO DEEPER

Reading Path

This study stands on its own. These companion papers go deeper on the operating-model questions it raises, from responsible AI use to evaluation criteria and the maturity steps that get a program there.

WHAT DOES RESPONSIBLE AI USE LOOK LIKE INSIDE THE PRACTICE ITSELF?

Operationalizing AI in Threat Modeling

Expands the AI-paradox argument and explains why governance, context, and human oversight stay essential.

COULD WE DO THIS WITH INTERNAL TOOLING OR DIRECT LLM USE?

Build vs. Buy vs. AI

Outlines where build-your-own and AI-only approaches create value, and where they fail to scale.

HOW SHOULD WE EVALUATE A MODERN PLATFORM?

Buyer's Guide to Threat Modeling

Provides criteria around enterprise readiness, collaboration, automation, architecture-aware analysis, and governance.

WHERE ARE WE TODAY, AND WHAT IS THE NEXT STEP?

Threat Modeling Maturity Model

Useful for teams moving from checklist or prompt-driven methods into architecture-aware, continuous practice.

For additional research and white papers on modern threat modeling, visit the ThreatModeler resource library at threatmodeler.ai.

SOURCES AND METHOD

Sources & References

METHODOLOGY NOTE

The findings cited here are drawn from the ThreatModeler AI-Driven Threat Modeling Thought Leadership Survey, conducted by Hanover Research in March 2026. Cited data points reflect an overall sample of n=250, with regional subsamples called out where available (Canada n=51, EU n=76, U.K. n=52, U.S. n=71; n=212 or n=193 as applicable for selected barrier questions). This report relies only on statistics available in the source materials; it does not infer survey-design details that are not documented.

No single survey determines a market. These findings indicate direction and buyer priorities, not the last word on category structure. The pattern across the survey and supporting material is consistent: AI has earned a place inside threat modeling, but not a blank check. Trust, governance, repeatability, and architecture awareness separate useful AI from AI theater.

REFERENCES

[1] ThreatModeler AI-Driven Threat Modeling Thought Leadership Survey. Hanover Research, March 2026. Key items include Q17, Q24–Q25, Q26–Q29, Q60, Q61, Q68, Q70, Q423–Q447, and results by country or region.

[2] Operationalizing AI in Threat Modeling. ThreatModeler whitepaper. Used for the framing around probabilistic outputs, the AI paradox, governed use of AI, and the distinction between AI-augmented and AI-dependent threat modeling.

[3] Build vs. Buy vs. AI: Navigating Threat Modeling in the AI Era. ThreatModeler whitepaper. Used for internal tooling, AI-only approaches, platform tradeoffs, and the role of architecture grounding, repeatability, and governance.

[4] Buyer's Guide to Threat Modeling. ThreatModeler guide. Used for evaluation-path references and criteria around enterprise readiness, collaboration, automation, and architecture-aware analysis.

[5] Threat Modeling Maturity Model: A Practical Framework for Advancing Secure by Design. ThreatModeler whitepaper. Used for the maturity-path reference and the distinction between prompt-based list generation and architecture-aware modeling.