ThreatModeler® + IriusRisk

Research Study

From Prompt to Proof

The trust gap in AI-driven threat modeling

A research-backed critical review of market adoption, trust barriers, and the requirements for enterprise-grade AI-assisted threat modeling.

Core thesis

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

Date: March 2026

Executive Summary

The market for threat modeling is no longer defined by a simple contrast between manual methods and automated platforms. It is being reshaped by a second shift: the arrival of generative AI in design-time security. That shift is real, but the Hanover data suggests it is still immature. Organizations are increasing their use of AI-assisted threat modeling, yet confidence in AI outputs remains limited - particularly where systems are regulated, safety-critical, or operationally complex. [1][2]

The data points are striking. Purchased automated threat modeling platforms are already widely used, at 93.6% of respondents, but manual practices still persist at 72.4%, suggesting that many organizations are operating with mixed maturity rather than a fully standardized operating model. At the same time, 75.6% of organizations already using AI-assisted threat modeling in manual or internally built environments expect that usage to increase over the next two years. This is not a fringe experiment. It is a category transition already underway. [1]

But adoption is not the same thing as trust. Only 32.0% of respondents say they trust AI-assisted threat modeling a lot or completely. That drops to 16.8% for safety-critical or regulated applications and 15.2% for legacy or monolithic environments. The most common barriers are not philosophical objections to AI; they are practical requirements for enterprise use: security and privacy, accuracy, validation effort, explainability, and governance. [1]

This creates a critical gap: AI is being adopted because it promises speed, but it is being constrained because security requires assurance. Threat modeling sits directly inside that gap because it is not just 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 process, but the data in this report suggests that organizations do not yet trust AI to replace it. [2][3]

Bottom line

The next phase of the category will not be won by the fastest prompt or the most eye-catching AI demo. It will be won by platforms that operationalize AI inside a deterministic, architecture-aware, and governed system of record.

1. The market is in transition, not at equilibrium

The first point the Hanover data establishes is that the threat modeling market is already modernizing, but not yet fully standardized. Respondents report high current use of purchased automated platforms (93.6%), yet a still-substantial use of manual methods such as spreadsheets and diagrams (72.4%). That combination matters. It indicates that many organizations have introduced automation without fully replacing manual work. In practice, this often means fragmented workflows, localized teams, or different operating models across application portfolios. [1]

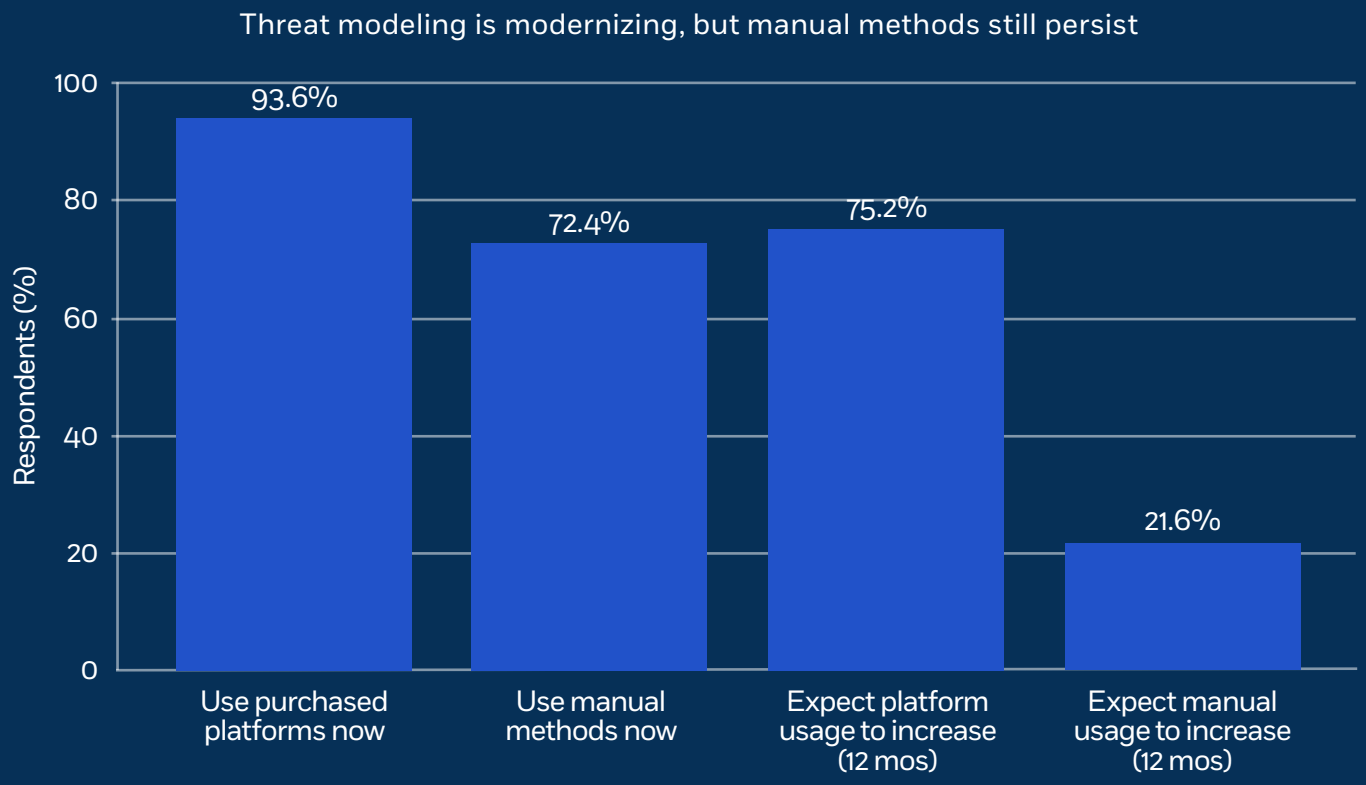


Figure 1. Threat modeling is modernizing, but manual methods still persist.

Why it matters: The category is not choosing between 'old' and 'new' all at once. Many enterprises are carrying both models at the same time, which raises the value of platforms that can standardize processes rather than simply automate isolated tasks.

The forward-looking data reinforces that interpretation. Three quarters of respondents expect purchased automated platform usage to increase over the next 12 months, while only 21.6% expect manual-process usage to increase. This does not mean manual methods disappear quickly. It does mean the center of gravity is shifting. Businesses appear to be looking for operating models that can bring consistency, speed, and portfolio-wide visibility to threat modeling, rather than leaving it as an expert-only workshop exercise. [1]

That shift also frames the AI question differently. If the market were still mostly manual, AI would mainly be a way to reduce labor. But in a market already moving toward platforms, the more important question becomes whether AI can be embedded in a repeatable system. That is a different standard. It is less about generative novelty and more about operational trust.

2. AI is already inside the SDLC – and inside threat modeling

The Hanover findings do not describe AI as a distant possibility. They show that AI-assisted development is already materially affecting the design-time security workflow. Among applications that already use threat modeling, an average of 34.46% also use AI code generation in the application development process. Threat modeling occurs during AI code generation 44.5% of the time on average, before it 31.28% of the time, and after it 24.22% of the time. [1]

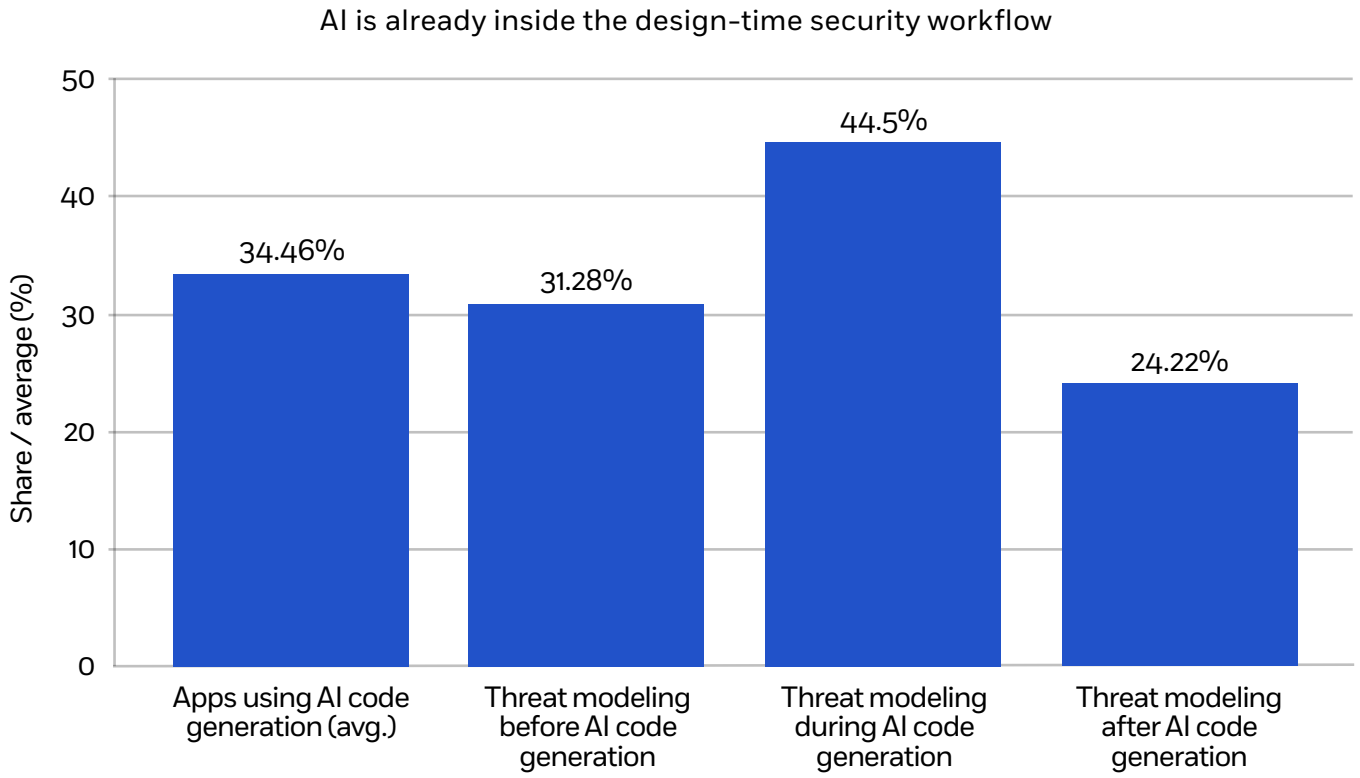


Figure 2. AI is already inside the design-time security workflow.

Why it matters: Threat modeling is no longer a separate practice operating safely outside AI adoption. It is increasingly entangled with AI-driven development itself, which raises the stakes for consistency, auditability, and architectural context.

This matters because AI changes the nature of the security problem. Cloud transformation already made systems harder to reason about, with more services, APIs, identities, and infrastructure dependencies spread across hybrid and multi-cloud environments. AI adds non-determinism on top of that complexity: code is generated faster, agentic systems are assembled without always being grounded in explicit intent, and the behavior that matters most is not always captured by reactive scanners or post-build review. ThreatModeler shifts from asking only '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?' [1]

The important conclusion is that AI-assisted threat modeling should not be evaluated as a novelty feature. It should be evaluated as part of the broader design-time security operating model. If AI is already participating in architecture description, code generation, requirements generation, or threat identification, then the real question is whether those outputs can be contained within a process that remains coherent, reviewable, and durable over time. [2][3]

3. The trust gap is the market bottleneck

The strongest signal in the Hanover research is not simple adoption. It is constrained trust. Only 32.0% of respondents say they trust AI-assisted threat modeling a lot or completely. A larger share, 53.2%, trusts it only somewhat. Trust collapses further when the stakes rise: only 16.8% trust AI-assisted threat modeling a lot or completely for safety-critical or regulated applications, and only 15.2% do for legacy or monolithic applications. [1]

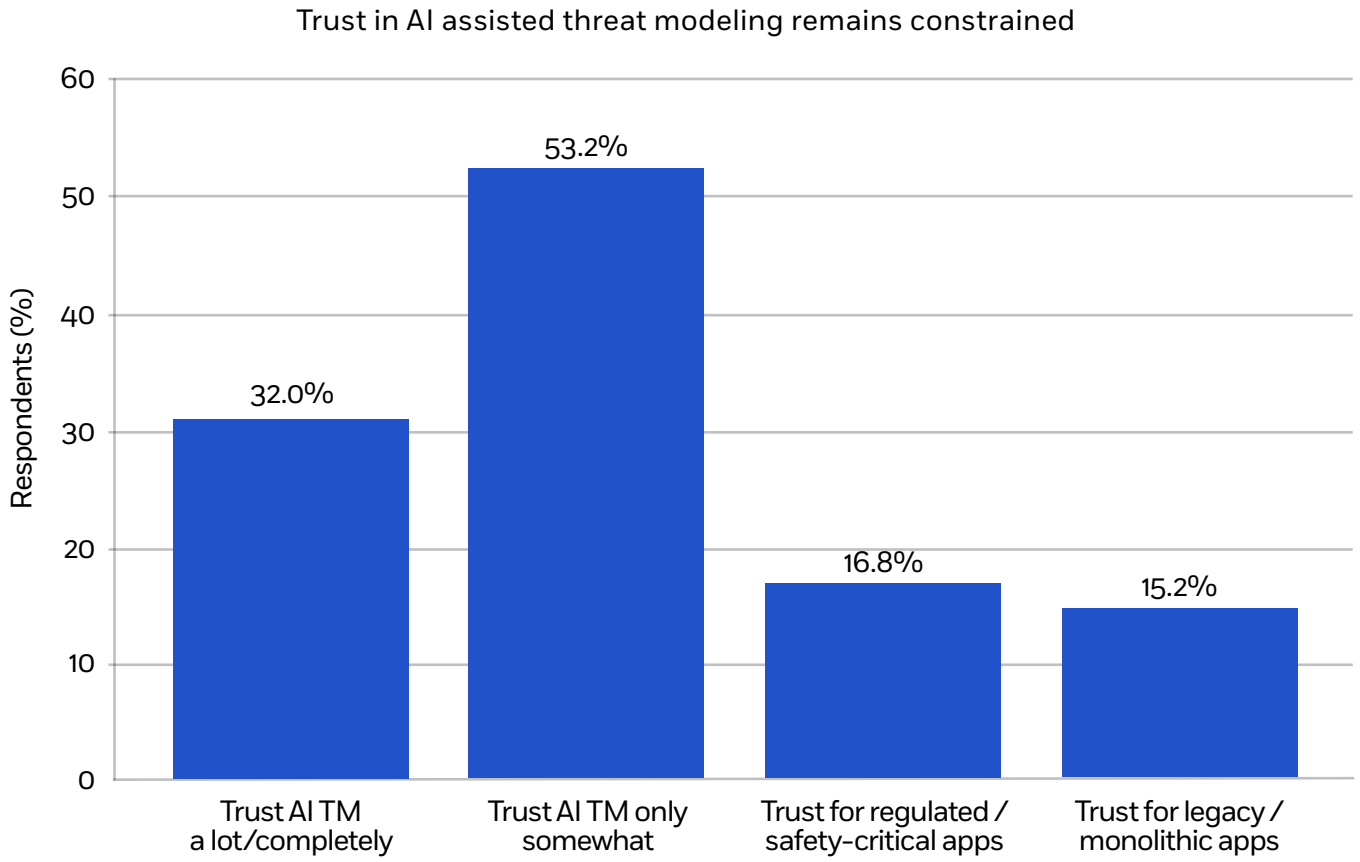


Figure 3. Trust in AI-assisted threat modeling remains constrained, especially in higher-risk environments.

Why it matters: The problem is not lack of interest. It is a lack of assurance. The closer AI gets to regulated, fragile, or business-critical environments, the more businesses demand governance and proof rather than raw speed.

AI can draft, summarize, or brainstorm extremely quickly. But threat modeling is not valuable because it produces text; it is valuable because it creates confidence in decisions. Where that confidence cannot be sustained, adoption stalls or becomes superficial. Teams may use AI to get started, but not to sign off. They may use it for low-stakes systems, but not for environments where architecture, compliance, or business continuity matter most.

Expert take

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

4. What organizations are worried about – and what they actually want

The barriers respondents cite are practical and revealing. Among organizations already using AI-assisted threat modeling in manual or internally built environments, 57.1% cite data security or privacy concerns, 52.4% cite accuracy or false positives, and 50.5% say AI outputs still require too much human validation. These are not fringe objections. They are the predictable consequences of using probabilistic systems in a discipline that depends on context, review, and traceability. [1][2]

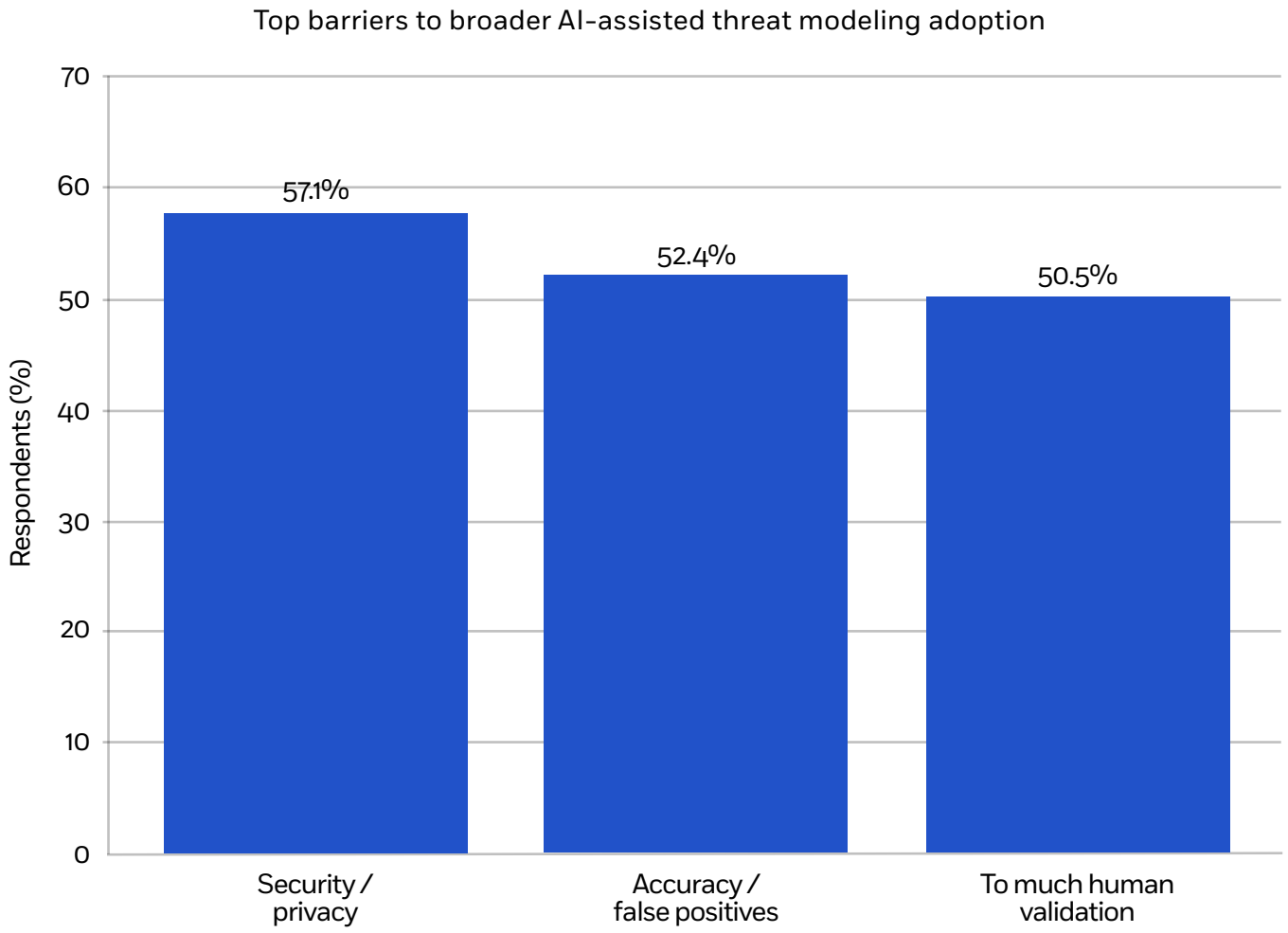


Figure 4. Top barriers to broader AI-assisted threat modeling adoption.

Why it matters: The three leading concerns line up around the same issue: enterprise teams do not merely want AI to generate more output. They want to know the output is safe to use, accurate enough to trust, and efficient enough not to create a second review burden.

The positive buying criteria make the same point from the other direction. 65.2% say strong security and privacy assurances would increase their likelihood of purchasing an AI-assisted threat modeling solution. 46.8% cite clear regulatory or compliance alignment. 41.2% want transparent, explainable AI outputs. 38.0% want demonstrated accuracy and validation data. 35.6% want strong governance and human-in-the-loop controls. [1]

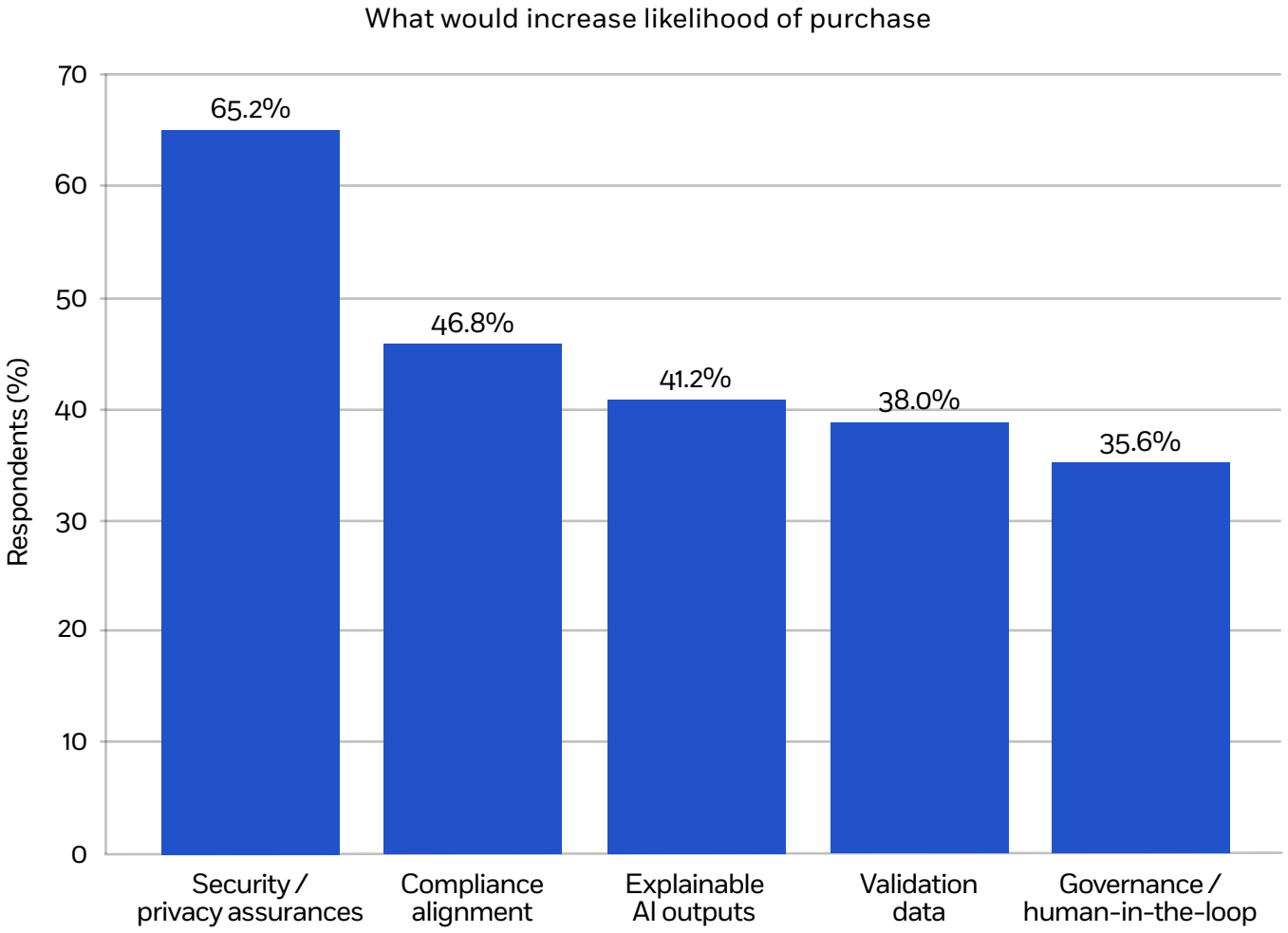
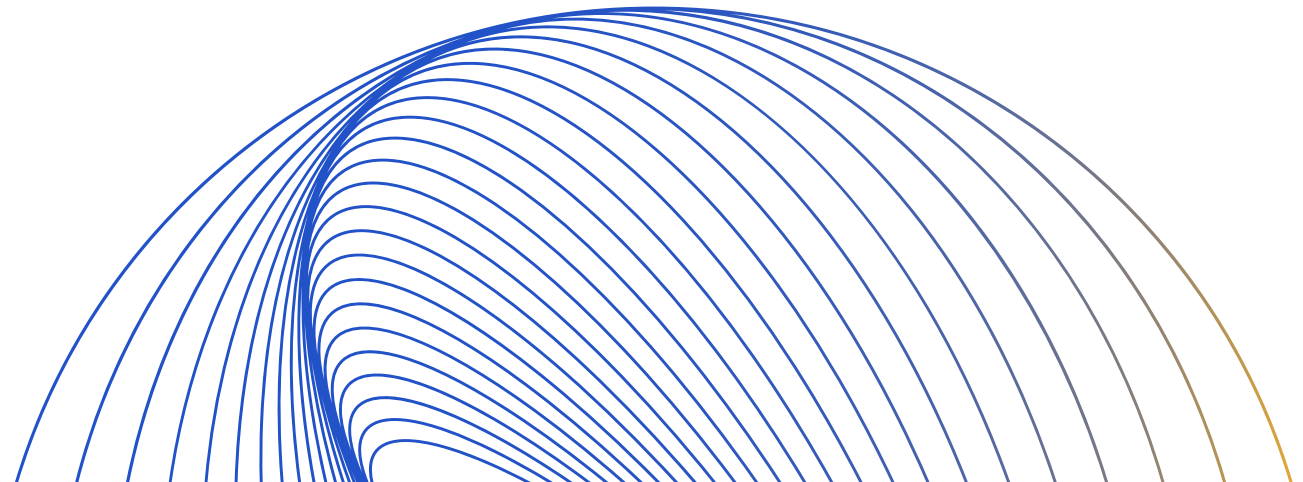


Figure 5. What would increase the likelihood of purchase.

Why it matters: Buyers are not asking for more AI theater. They are specifying the conditions under which AI becomes acceptable for enterprise use: privacy, compliance, explainability, validation, and governance.

Taken together, these findings strongly support the view that the next category winner will not be a generic LLM front end. It will be a system that contains AI within a workflow that can satisfy enterprise requirements. ThreatModeler's whitepaper Operationalizing AI in Threat Modeling makes the same argument more explicitly: AI has value when it reinforces process, not when it replaces it, and when it operates inside a framework that preserves contextual awareness, collaboration, governance, and repeatability. [2]



5. The AI paradox: more speed, more validation

ThreatModeler's Operationalizing AI in Threat Modeling whitepaper describes a useful framing for the current market: the AI paradox. The more teams rely on AI to automate reasoning, the more human expertise is required to verify results. This is not a contradiction in the technology; it is a mismatch between what generative AI is optimized to do and what threat modeling requires. Generative systems are good at pattern completion, draft generation, and summarization. Threat modeling, by contrast, demands architectural reasoning, control placement, prioritization, and defensible documentation. [2]

This is why prompt-based approaches often disappoint after the initial demo. They are helpful for ideation, but they struggle to maintain the conditions that enterprise security teams actually need. The weaknesses are structural:

Non-determinism

The same prompt can yield different results, which is acceptable for creative work but problematic for repeatable analysis. [2]

No architectural grounding

Direct LLM use is typically disconnected from actual cloud architecture, infrastructure-as-code, trust boundaries, and application context. [3]

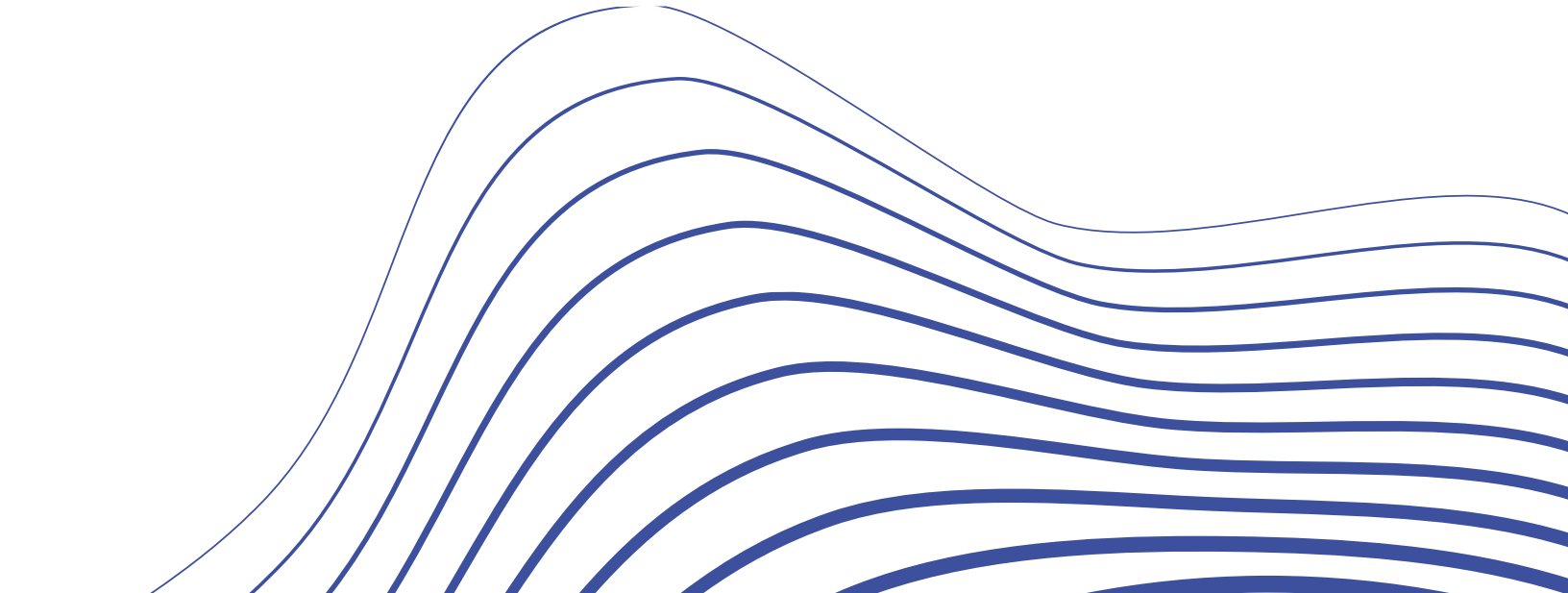
No built-in governance

Prompt transcripts are not a system of record. They do not inherently provide versioning, approvals, or reusable threat-to-control mappings. [2][3]

Validation burden

The burden of checking whether AI output is correct remains with the architect, often recreating a large share of the work AI was meant to reduce. [1][2]

This is also where much of the current startup wave is exposed. The category's buyer problem is not simply 'can AI generate a threat list?' It is 'can AI-assisted security decisions be trusted, repeated, governed, and defended?' Prompt-centric products and lightweight wrappers may be able to demonstrate speed, but unless they can answer that second question, they remain better suited to experimentation than enterprise operating models.



6. Why build-your-own and AI-only approaches break

The companion paper Build vs. Buy vs. AI is useful because it shows that the market is no longer making a simple tool choice. Organizations now face three distinct paths: build an internal solution, use AI or LLMs directly, or buy a purpose-built platform. The whitepaper's conclusion is not that AI or internal development have no value. It is that each path has different limits - and those limits become decisive as scale, volatility, and assurance requirements rise. [3]

Build-your-own approaches can offer alignment with internal workflows and allow narrow experimentation, especially where teams want tight control over data or integration. But their ongoing burden is substantial: threat content, framework mappings, cloud coverage, and governance models all require continuous maintenance. AI can accelerate prototypes, yet it does not remove the need to keep models accurate and repeatable over time. [3]

AI-only approaches are even more constrained. They provide a low-friction way to brainstorm threats, summarize system behavior, or support learning. But as the whitepaper argues, they lack architectural grounding, drift detection, governance, collaboration support, and dependable repeatability. That makes them poorly suited to audit-sensitive environments, multi-team engineering organizations, and fast-changing cloud architectures. [3]

Expert takeaway

The question is not whether AI-only or homegrown methods can create value. They can. The question is whether they remain viable when the organization needs continuity, shared standards, and evidence that stands up to review. At that point, the market shifts from experimentation economics to operating-model economics.

The strongest fit for a platform emerges when teams need modeling grounded in real architecture, consistent workflows, reusable libraries, approvals, integrations, and ongoing maintenance of threat and framework content. In other words, the decisive issue is not who can generate the first draft fastest. It is who can preserve accuracy, reuse, and accountability after the first draft exists. [3]

7. What the data implies geographically

The Hanover survey also suggests that the market should not be treated as homogeneous. Overall interest for an AI-assisted threat modeling solution in the next two years is 47.6% very or extremely likely, but that 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. [1]

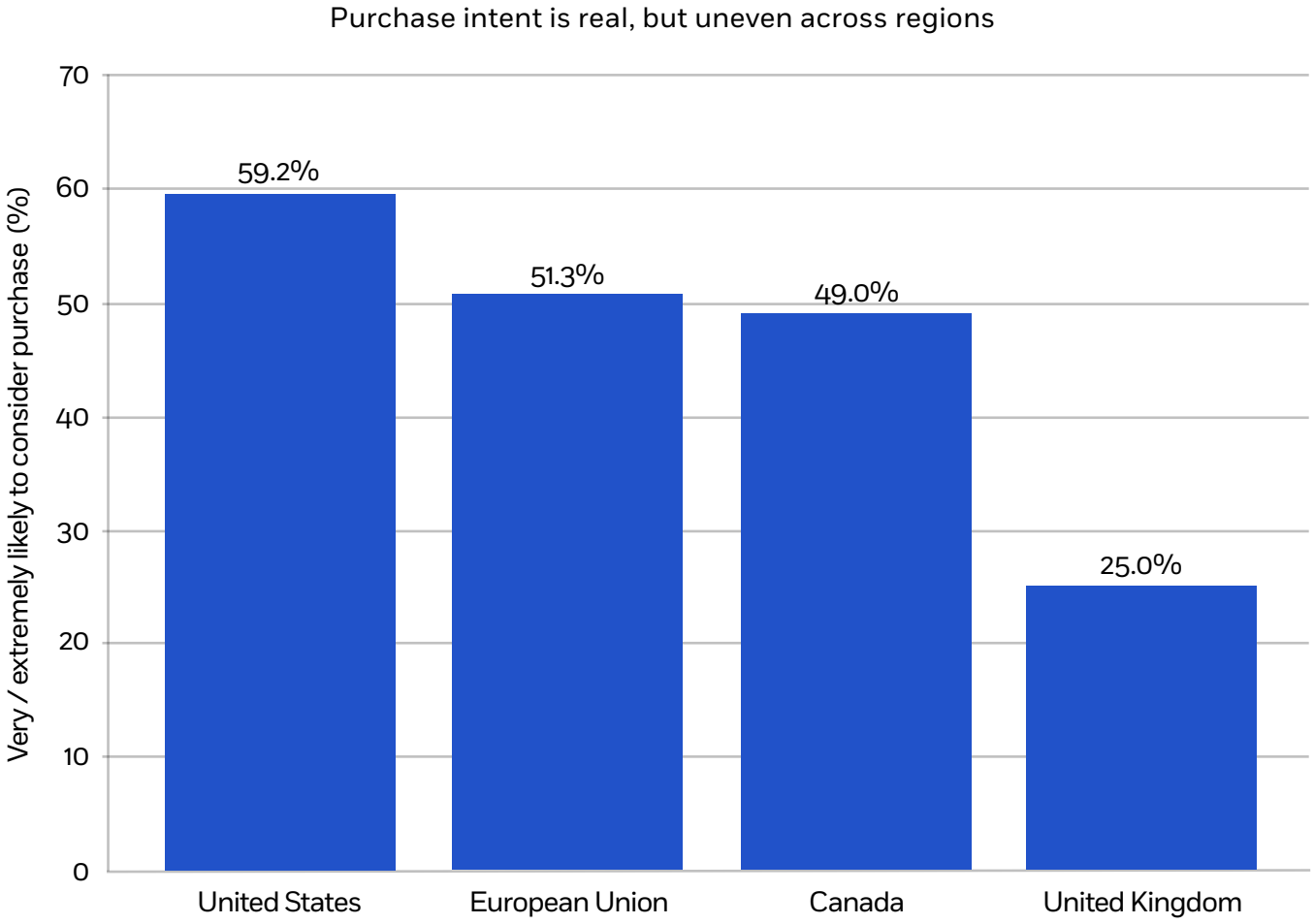


Figure 6. Purchase intent is real, but uneven across regions.

8. Reading path: how to go deeper

This report is designed to stand on its own, but there are also operational and strategic pieces available in addition, to assist business currently tackling this transformational challenge.

Operationalizing AI in Threat Modeling

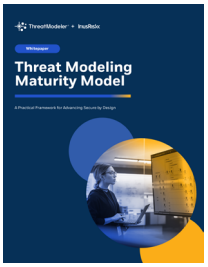
Read this next if the question is, 'What does responsible AI use look like inside the practice itself?' It expands the AI paradox argument and explains why governance, context, and human oversight remain essential. [2]

Build vs. Buy vs. AI

Read this next if the question is, 'Could we do this with internal tooling or direct LLM use?' It outlines where build-your-own and AI-only approaches create value and where they typically fail to scale. [3]

Buyer's Guide to Threat Modeling

Read this next if the question is, 'How should we evaluate a modern platform?' It provides criteria around enterprise readiness, collaboration, automation, architecture-aware analysis, and governance. [4]



Threat Modeling Maturity Model

Read this next if the question is, 'Where are we today, and what is the next maturity step?' It is especially useful for teams trying to move from checklist-driven or prompt-driven methods into architecture-aware and continuous practices. [5]

Methodology note

The Hanover research findings cited here are drawn from the ThreatModeler AI-Driven Threat Modeling Thought Leadership Survey conducted in March 2026. The cited data points reflect an overall sample size of n=250, with specific questions and regional subsamples called out where available. This report relies only on the statistics available in the source materials provided; it does not infer details about survey design that are not explicitly documented.

A final caution is warranted. No single survey determines a market. The Hanover findings should be interpreted as evidence of direction and buyer priorities, not as the last word on category structure. Even so, the pattern across this survey and the supporting public-facing content is consistent: AI has earned a place inside threat modeling, but not yet a blank check. Trust, governance, repeatability, and architecture awareness are what separate useful AI from unsafe AI theater.

Sources and references

[1] ThreatModeler AI-Driven Threat Modeling Thought Leadership Survey, Hanover Research, March 2026. Key cited survey items include Q17, Q24–Q25, Q26–Q29, Q60, Q61, Q68, Q70, Q423–Q447, and Results by Country or Region. Survey base sizes cited in the source material: n=250 overall; Canada n=51; EU n=76; UK n=52; US n=71; n=212 or n=193 as applicable for selected barrier questions.

[2] Operationalizing AI in Threat Modeling: Transforming generative insights into governed, repeatable security outcomes. 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 the analysis of internal tooling, AI-only approaches, platform tradeoffs, and the role of architecture grounding, repeatability, and governance.

[4] Buyer's Guide to Threat Modeling 2026. ThreatModeler guide. Used for evaluation-path references and criteria around enterprise readiness, collaboration, process 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.