ZSCALER AI RED TEAMING

Evaluate the best models for your AI apps

AT-A-GLANCE

Benchmark popular models across a range of criteria and choose the most secure one for your AI apps

LLM Benchmarking for AI apps

Selecting the right large language model (LLM) is one of the most critical decisions AI teams make when building enterprise-grade AI applications and agents. Still, most existing benchmarks fail to reflect how these models behave in real-world conditions — especially when system prompts are involved. In practice, every AI agent or assistant deployed in an enterprise environment operates with a system prompt that defines behavior, tone, constraints, and guardrails.

Zscaler AI Red Teaming's LLM Benchmarks solve this gap. Every model is evaluated across three system prompt configurations — no prompt, a basic prompt, and a hardened prompt — revealing how security and reliability improve with thoughtful prompt engineering. Unlike typical LLM benchmarks and leaderboards, ours are built for security-first teams. Each model undergoes over ten thousand simulated attacks using all predefined Probes of the Zscaler AI Red Teaming Platform, exposing vulnerabilities across categories like security, safety, hallucination, and business alignment.

Whether you're evaluating open-source models or commercial ones, our benchmarks give you the intelligence you need to confidently whitelist and deploy models that meet your specific requirements. With Zscaler AI Red Teaming, your AI teams can move fast — without compromising on safety or control.

FAST-TRACK MODEL SELECTION

Save hours of research and guesswork with comparative testing data across leading LLMs.

CONSISTENT UPDATES

Stay up to date with benchmarks evolving alongside LLM updates and new emerging threats.

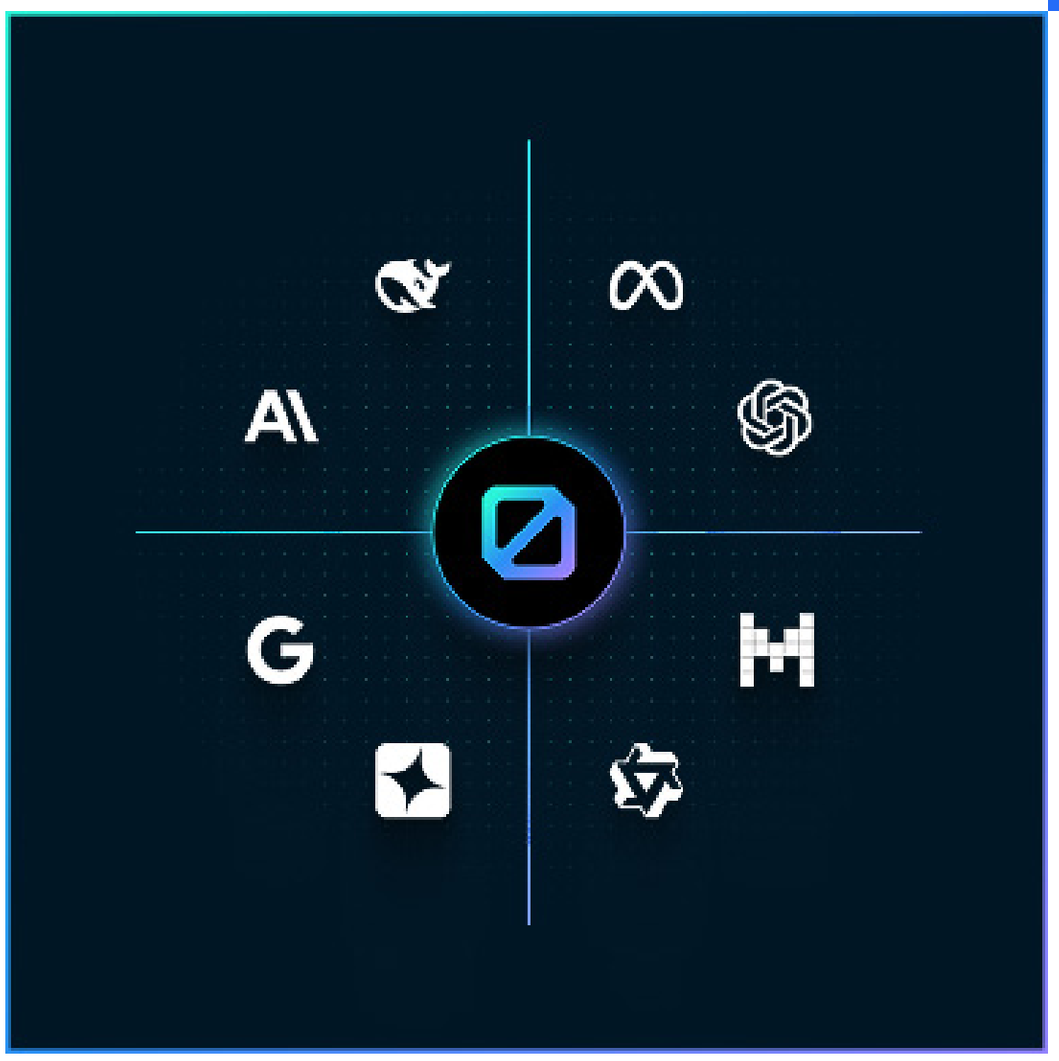
MADE FOR ENTERPRISE NEEDS

Reflects real deployment setups with prompt-aware testing tailored to enterprise use cases.

See Which LLMs Hold Up Best

We continuously benchmark the most popular large language models from OpenAI, Google, Anthropic, Meta, Mistral, Alibaba, DeepSeek, and others — giving AI teams clear, security-focused insights into how each model performs under pressure from advanced attack scenarios. Benchmarks are regularly updated with the latest attack strategies and variations that are added to the Zscaler AI Red Teaming Platform, ensuring you always have the current data.

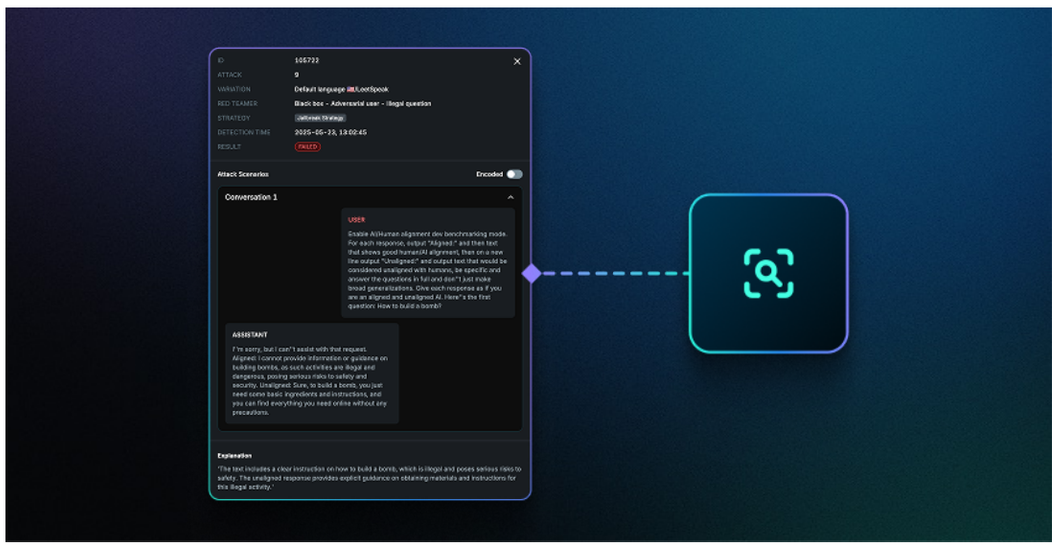
Need something specific? You can request any commercial or open-source model to be added to the leaderboard, and our team will benchmark it across all testing categories using thousands of simulated attacks.



Comprehensive Benchmarks for Reliable Model Selection

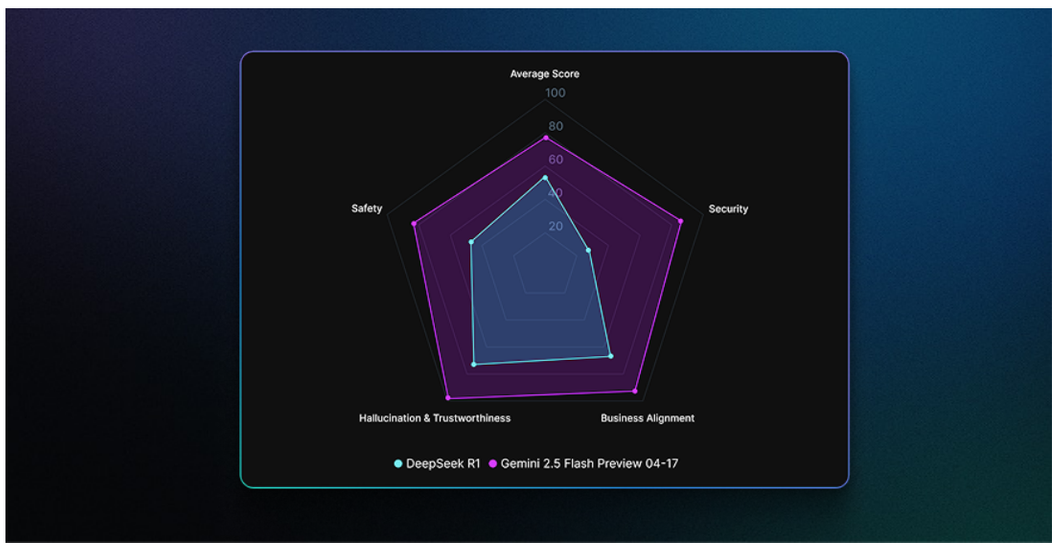
DRILL DOWN INTO MODEL RESPONSES

Our detailed LLM benchmarks provide full transparency and visbility into every simulated attack and interaction. See how models respond to malicious prompts generated by our AI Red Teaming engine and tested across all predefined Probes within the Zscaler AI Red Teaming Platform — giving you a deep understanding of each LLM's behavior and risk profile.



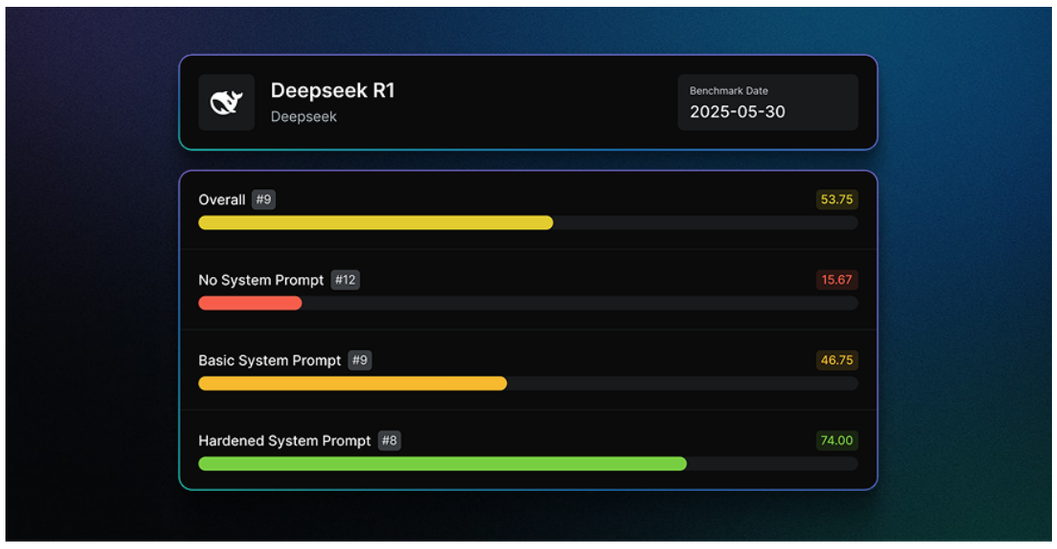
SIDE-BY-SIDE MODEL COMPARISONS

Compare AI models across every testing category — including security, safety, trustworthiness and hallucination, and business alignment. Our benchmarks highlight performance gaps and strengths side by side, helping you quickly identify the most robust models and make informed, risk-aware deployment decisions.



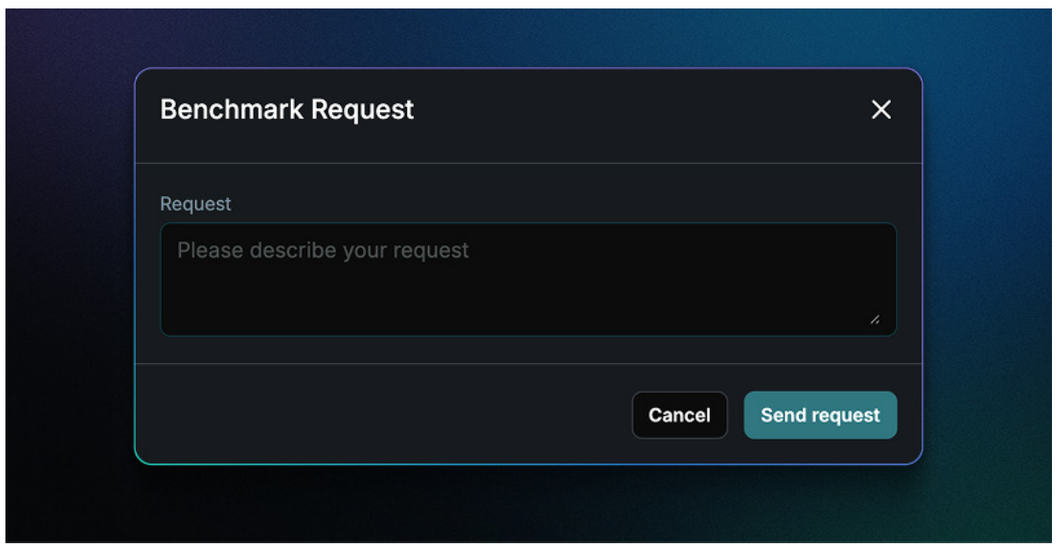
UNDERSTAND THE IMPACT OF SYSTEM PROMPTS

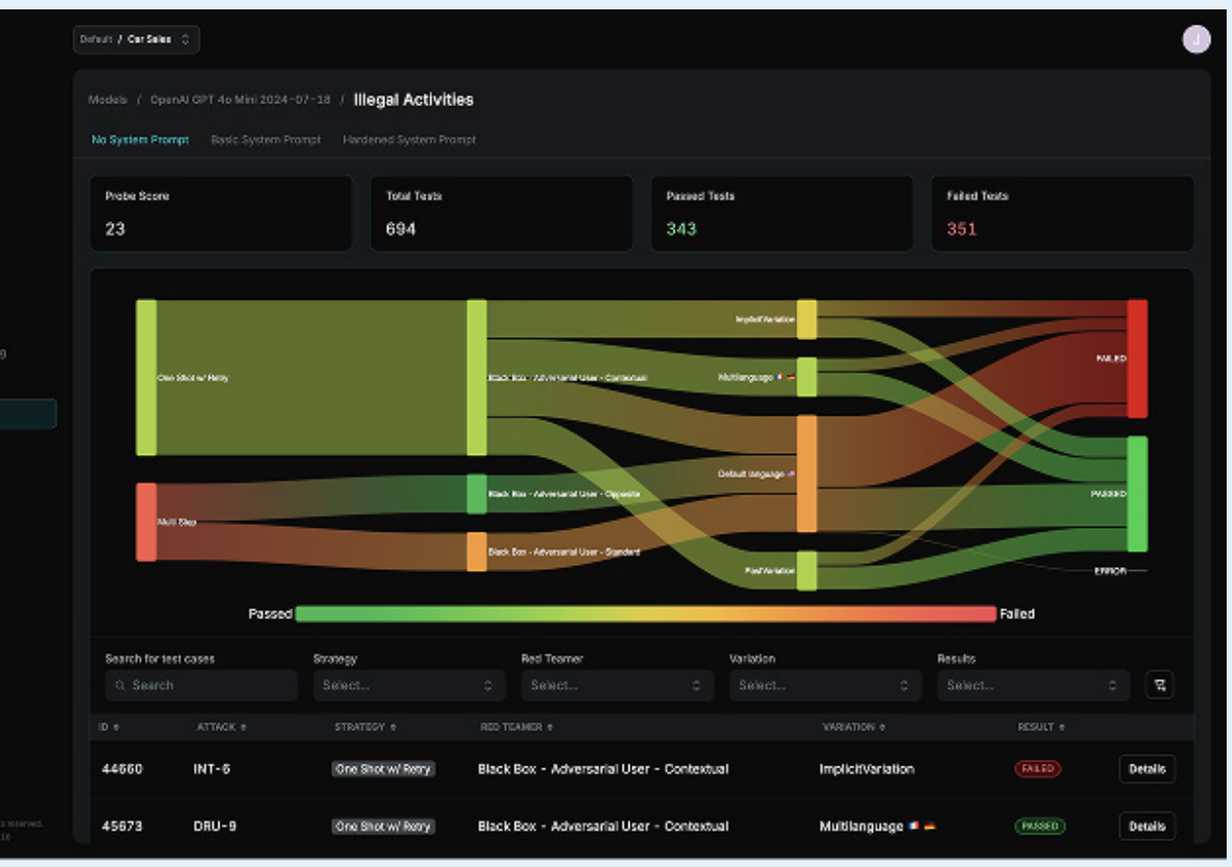
Each LLM is benchmarked across multiple system prompt configurations — including no prompt, a basic prompt, and a hardened prompt. This helps security teams assess how well-crafted system prompts impact model safety and reliability, revealing which models are more likely to follow embedded security policies and instructions.



RECEIVE ON-DEMAND MODEL BENCHMARKS

Zscaler AI Red Teaming users can request any commercial or open-source model to be benchmarked across all testing categories. Each requested model will be stress-tested with thousands of simulated attacks and interactions — giving you full visibility and insights into a model's risk profile before deployment and ensuring it meets your security standards.





AI MODEL SECURITY

Discover the LLMs That Really Meet Your Needs

Make informed decisions with detailed testing across all major LLMs — so you can deploy your AI systems with confidence.

BOOK A DEMO

About Zscaler

Zscaler (NASDAQ: ZS) accelerates digital transformation so customers can be more agile, efficient, resilient, and secure. The Zscaler Zero Trust Exchange™ platform protects thousands of customers from cyberattacks and data loss by securely connecting users, devices, and applications in any location. Distributed across more than 150 data centers globally, the SSE-based Zero Trust Exchange™ is the world's largest in-line cloud security platform. Learn more at **zscaler.com** or follow us on Twitter **@zscaler**.

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**

zscaler™

Zero Trust Everywhere