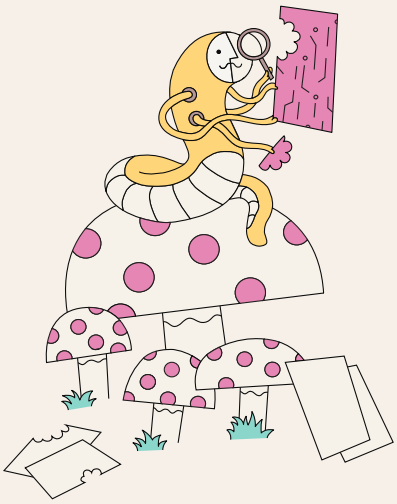


Sustain Trust as AI Evolves.

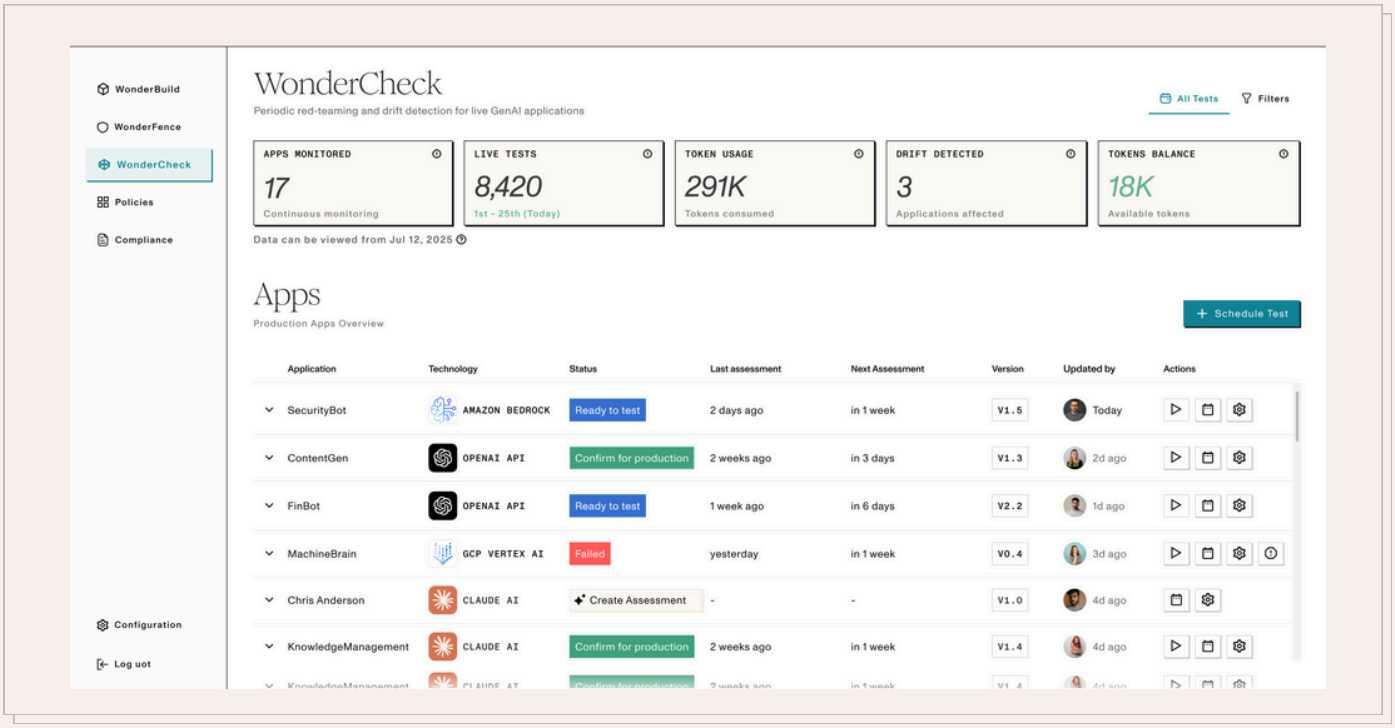
Once deployed, GenAI and Agentic systems can drift from their intended behavior as models are updated, prompts change, and real-world usage patterns shift. New risks can emerge quietly, while previously mitigated issues may resurface.

Scripted validation and platform-native guardrails often fail to catch these changes in production. Teams need a reliable way to periodically evaluate live systems, assess platform-native guardrail accuracy, and surface clear, actionable insights as risk conditions change.

Ongoing red teaming for production AI systems to detect drift and surface emerging risks

WonderCheck automates security and safety testing for production GenAI and Agentic systems, helping teams detect drift, regressions, and emerging vulnerabilities before adversaries do. By running ongoing, customizable policy-driven tests, it reveals how system behavior changes over time and highlights platform-native guardrail accuracy issues, giving teams clear, prioritized insights to maintain reliability, security, and compliance.

Powered by *Rabbit Hole,* Alice's adversarial intelligence engine, **WonderCheck** incorporates real-world threat patterns to inform ongoing evaluation as risk landscapes evolve.



WonderCheck Key Benefits

Automated Adversarial Testing for Production AI Systems
Runs realistic and adversarial inputs on an ongoing, on-demand, or scheduled basis to surface new threats and hidden vulnerabilities in live GenAI and Agentic systems.

Drift and Regression Detection
Identifies changes in model and system behavior caused by updates, fine-tuning, new prompts, or shifting real-world usage patterns, helping teams catch regressions and emerging risks that may otherwise go unnoticed in production.

Broad Modality Coverage
Tests GenAI and Agentic systems across text, image, audio, and video inputs to uncover safety, security, and robustness risks wherever they appear, including issues that may surface only in specific modalities or cross-modal interactions.

Guardrail Validation
Evaluates platform-native and third-party guardrails to help identify false positives and false negatives, helping teams improve accuracy and calibration.

Policy Impact Analysis
Uses deeply customizable, policy-driven tests to assess how proposed WonderFence policy changes may affect production behavior before deployment.

Clear, Prioritized, Actionable Insights
Provides digestible findings and practical recommendations so development teams can focus on fixing the most important issues first.

Demonstrative Compliance Alignment
Provides assessments aligned with frameworks and regulations including the EU AI Act, ISO 42001, NIST, and OWASP, helping teams document and demonstrate responsible AI development practices.

Seamless Operational Workflow Integration
Supports ongoing, scheduled, or on-demand testing within CI/CD pipelines, evaluation cycles, and governance programs, allowing teams to incorporate production testing into existing processes without disrupting development velocity or operational workflows.

Alice is Trusted by the World's Leading Technology Innovators

TikTok · NVIDIA · Black Forest Labs · Udemy · Lovable · RIOT GAMES · amazon game studios

Navigate the twists and turns of GenAI with WonderSuite

WONDERBUILD · WONDERFENCE · WONDERCHECK

Explore WonderSuite

Ready to advance your communicative tech unafraid?