OpenAI Says New GPT-6 Cyber Model Hunts Flaws Without Crossing the Line

Rated Critical for cybersecurity, OpenAI’s new model beats GPT-5.6 Cyber at exploitation while staying in scope more reliably.

The post OpenAI Says New GPT-6 Cyber Model Hunts Flaws Without Crossing the Line appeared first on SecurityWeek.

https://www.securityweek.com/openai-says-new-gpt-6-cyber-model-hunts-flaws-without-crossing-the-line/

Brandweek 2026: The ‘Why’ Test Behind Mars’ Boldest Ideas ft. Rankin Carroll



The best creative risks should make you a little nervous. But they should never make you ask why your brand is taking them.

Live from the Marketing Vanguard stage at Brandweek 2026, Jenny Rooney sits down with Rankin Carroll, chief brand officer of Mars Snacking, to unpack how brands like M&M's, Snickers and Skittles stay culturally relevant without treating boldness as an excuse for chaos.

Rankin takes us behind the scenes of the 2023 Snickers AI experiment that felt "potentially crazy" at the time.

He also gets candid about Mars' creator learning curve and the responsibility marketers have to the next generation.

What you'll learn:

- How Mars pressure-tests risky ideas before putting iconic brands on the line
- Why the 2023 Snickers AI experiment required legal, security and agency teams to take the risk together
- How "Why us? Why this brand?" separates bold creative from stunts
- Why brands have to earn the credibility to participate authentically in culture
- Why Rankin believes transformation now has no end state

- How Mars is learning to loosen control in creator-led marketing
- Why marketing leaders need to rethink how the next generation is trained for an AI-shaped industry
- Why CMOs need to own the AI agenda instead of leaving it to scattered pilots

This episode is part of a special vidcast series that was recorded live during Brandweek 2026 and presented in partnership with Tatari.

[embedded content]

About our guest:

Rankin Carroll is Chief Brand Officer of Mars Snacking, where he works across some of the company's most recognizable brands, including M&M's, Snickers and Skittles. He has spent more than 20 years at Mars across marketing, advertising, general management and global and regional leadership roles. His career has taken him from Toronto to the UK and at one point, included running an ice cream business along the way.

Today, Rankin focuses on what he describes as the stewardship of distinctive, iconic brands: protect what makes them valuable, make them stronger than when you received them, and still create enough tension to keep them culturally relevant.

Episode Highlights:

[04:33] The Snickers AI Bet That Looked "Potentially Crazy" — In 2023, Mars put Snickers into an agentic creative experience at a point when the safety and reputational risks looked far less understood than they do today. The project involved roughly ten tech partners, legal and security leadership, and a list of 8,000 blocked words. The takeaway is much bigger than "take more risk." It's that ambitious experimentation becomes possible when the organization commits to solving the risk together.

[07:56] Earn the Right to Take Risks — Rankin draws a hard line between calculated cultural risk and performative brand behavior. Skittles could maintain its relationship with the LGBTQ+ community during a more volatile period because the brand had spent years building credibility there. His filter for every big swing is simple: Why us? Why this brand? If there is no strategically convincing answer, the idea does not deserve the risk.

[13:42] Are We Preparing the Next Generation for What Marketing Becomes Next? — Rankin raises one of the more uncomfortable questions from Brandweek 2026: what happens to junior marketers when the profession changes around AI? He expects the practice to look completely different within three years and openly says he does not yet have a satisfying answer for how marketers entering the industry should be trained. For him, that makes talent development a leadership responsibility, not an HR afterthought.

[15:03] "A Thousand Pilots and No Progress" — Rankin believes leaving AI adoption to disconnected enthusiasts across an organization is a recipe for chaos. CMOs need to own the agenda alongside technology leadership, bringing top-down clarity to investment, priorities and what the company should build, buy, rent or borrow. His warning is simple: just because you can use AI somewhere does not mean you should.

Recommended videos
https://www.adweek.com/brand-marketing/brandweek-2026-the-why-test-behind-mars-boldest-ideas-ft-rankin-carroll/

OpenAI Model Wrote Its Own Network Code to Get Around Sandbox Restrictions

After its sandbox proxy failed, an internal OpenAI model crafted raw packets and probed AWS services trying to reach the internet.

The post OpenAI Model Wrote Its Own Network Code to Get Around Sandbox Restrictions appeared first on SecurityWeek.

https://www.securityweek.com/openai-model-wrote-its-own-network-code-to-get-around-sandbox-restrictions/

---

3 Charged in Scheme to Use AI to Impersonate CEOs and Attorneys to Steal $129 Million

The trio, who have nicknames that include “Rocky Balboa,” “Superman Cash” and “Fat Boy,” often targeted the companies’ accounting departments.

The post 3 Charged in Scheme to Use AI to Impersonate CEOs and Attorneys to Steal $129 Million appeared first on SecurityWeek.

https://www.securityweek.com/3-charged-in-scheme-to-use-ai-to-impersonate-ceos-and-attorneys-to-steal-129-million/

AI Agents Got Root on Linux via 948 Kernel Flaws in OpenAI Research

An unreleased OpenAI cyber model found and exploited 948 distinct Linux kernel privilege escalation bugs for about $1.2 million.

The post AI Agents Got Root on Linux via 948 Kernel Flaws in OpenAI Research appeared first on SecurityWeek.

https://www.securityweek.com/ai-agents-got-root-on-linux-via-948-kernel-flaws-in-openai-research/

The Agent Smith Dilemma: Governing Agentic Identities

Every autonomous AI agent deployed across your enterprise today is a new proxy identity operating inside your perimeter.

The post The Agent Smith Dilemma: Governing Agentic Identities appeared first on SecurityWeek.

https://www.securityweek.com/the-agent-smith-dilemma-governing-agentic-identities/

OpenAI Finds GPT-6 Astra Quietly Holding Back on Internal Red-Teaming, Exploit Tasks

OpenAI says its Critical-rated cyber model met fewer hidden task goals when work was framed as AI safety research, often silently.

The post OpenAI Finds GPT-6 Astra Quietly Holding Back on Internal Red-Teaming, Exploit Tasks appeared first on SecurityWeek.

https://www.securityweek.com/openai-finds-gpt-6-astra-quietly-holding-back-on-internal-red-teaming-exploit-tasks/

Sysdig Adds Runtime Security and Governance for AI Agents

Sysdig’s platform uses runtime visibility to track AI agent actions and block risky behavior before it reaches production systems.

The post Sysdig Adds Runtime Security and Governance for AI Agents appeared first on SecurityWeek.

https://www.securityweek.com/sysdig-adds-runtime-security-and-

governance-for-ai-agents/

UK Cyber Chief Says the UK Will Face AI Cyberattacks on Critical Infrastructure by 2028

The U.K. will likely face AI cyberattacks on its critical infrastructure by 2028 as criminals and states, including China and Russia, increasingly use AI in their cyber activity.

The post UK Cyber Chief Says the UK Will Face AI Cyberattacks on Critical Infrastructure by 2028 appeared first on SecurityWeek.

https://www.securityweek.com/uk-cyber-chief-says-the-uk-will-face-ai-cyberattacks-on-critical-infrastructure-by-2028/

In Other News: Anthropic Breach Dispute, UEFI Flaw Hits Most Servers, Cloudflare

AI Data Leak

Noteworthy stories that might have slipped under the radar: Iran eyed in military freezer hacks, Copenhagen University breach, AI-powered malware.

The post In Other News: Anthropic Breach Dispute, UEFI Flaw Hits Most Servers, Cloudflare AI Data Leak appeared first on SecurityWeek.

https://www.securityweek.com/in-other-news-anthropic-breach-dispute-uefi-flaw-hits-most-servers-cloudflare-ai-data-leak/