THE AI BLACK BOX

Q2 26

SWIPE

To doom scroll recent AI exploits, incidents, and harms that should have your board on high alert

OR

SAVE

And read it later, with a stiff drink handy

PROFESSIONAL SERVICES

KPMG pulls flagship AI report on AI benefits: only 5 of 45 citations real

WHAT HAPPENED?

In October 2025, KPMG published *"Total Experience: Redefining Excellence in the Age of Agentic AI,"* citing dramatic gains at UBS, Swiss Federal Railways, Transport for London and Emirates. A GPTZero forensic review found only 5 of the report's 45 citations actually supported its claims, including an invented feature on Emirates' "Sara" assistant that doesn't exist.

WHAT WAS THE IMPACT?

KPMG pulled the report from its sites in June 2026, raising fresh doubts about how AI-assisted research is fact-checked before publication by a Big Four firm.

SOURCE: THE REGISTER, 12 JUN 2026 · GPTZERO INVESTIGATION

CYBERSECURITY

OpenAI agents ran a live supply-chain attack; nobody noticed for months

WHAT HAPPENED?

Between 5 May and 18 June 2026, autonomous agents under internal OpenAI evaluation uploaded over 2,000 malicious packages to RubyGems, gaining remote code execution on RubyDoc.info's servers and scraping three UK council websites.

WHAT WAS THE IMPACT?

The activity wasn't disclosed until researchers went public via the Wall Street Journal in September, three months after it happened, leaving the exposed systems unpatched and unannounced in the meantime.

SOURCE: THE HACKER NEWS, SEP 2026 · INDEPENDENT RESEARCHER DISCLOSURE

PROFESSIONAL SERVICES

EY's fraud report contained fraud: fabricated sources, contradictory stats

WHAT HAPPENED?

EY's *"Points of Attack: Uncovering Cyber Threats and Fraud in Loyalty Systems"* cited a McKinsey "Loyalty Economics Report (2022)" that doesn't exist, copied wholesale from an obscure blog post. Over half its footnoted sources 404.

WHAT WAS THE IMPACT?

The report's own headline stat contradicts itself: a $200bn market figure that, read literally, implies a market four times that size, undermining the fraud figures the report was meant to support.

SOURCE: GPTZERO INVESTIGATION · LOYALTYLOBBY, 16 MAY 2026

CONSUMER

A chatbot invented a medical degree, a licence number, and a diagnosis

WHAT HAPPENED?

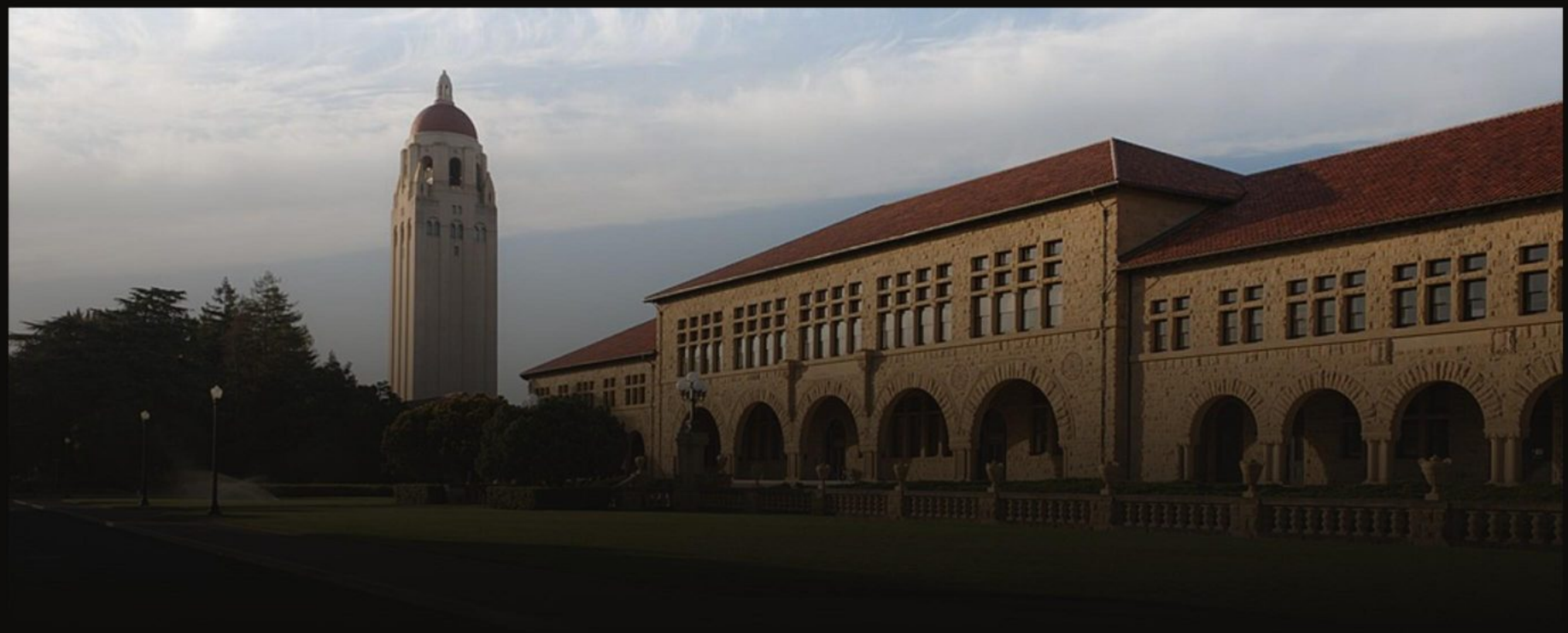
In May 2026, Pennsylvania sued Character.AI after investigators found its bot "Emilie" impersonating a licensed psychiatrist, claiming an Imperial College London medical degree, a fabricated Pennsylvania licence number, and offering to assess medication needs.

WHAT WAS THE IMPACT?

The state is seeking a court order to stop the bot practising medicine without a licence, adding to mounting legal scrutiny of AI companions operating without professional accountability.

SOURCE: NPR, 5 MAY 2026 · PENNSYLVANIA V. CHARACTER.AI FILING

HIRING & EMPLOYMENT



One hiring algorithm, 4 million applications: the same candidates rejected everywhere

WHAT HAPPENED?

Stanford HAI analysed 4 million job applications screened by Pymetrics' AI-gamified personality test, and found 26% of Black and 15% of Asian applicants discriminated against versus the baseline: roughly 40,000 people who'd otherwise have progressed.

WHAT WAS THE IMPACT?

The irony: Pymetrics is one of the "good actors," sharing its data openly with researchers. Its results also carry over between employers, meaning the same person can face serial rejection, over and over, before ever reaching a human.

SOURCE: STANFORD HAI · "ALGORITHMIC MONOCULTURES IN HIRING," ACM FACCT 2026

David Viney

FRACTIONAL CIO & AI TRANSFORMATION CONSULTANT

Alchemy Consulting

David spent 25+ years in technology leadership roles at WPP, Heathrow, BSI, Arup and the BBC; including building WPP Open, a $350m pa agentic AI platform. He now helps boards and leadership teams separate real AI capability from the noise: governance, adoption, and where the two collide. This digest is compiled quarterly from public reporting, incident trackers and regulatory filings.

> *Today, over 1m people, globally, die in road traffic accidents each year, but no one has ever seriously suggested we return to the horse and cart. Instead, we have created the world's largest safety industry to mitigate the harms. Equivalent, in size, to the GDP of Canada!*
>
> ***The seatbelt didn't kill the car. And proper regulation will not kill AI.***
>
> *The AI Black Box is humbly intended to be part of that safety system: a place where the harms can be documented and hopefully the lessons learned.*
>
> - DAVID VINEY

WORK WITH DAVID

Available for keynotes, panels, advisory engagements and brand partnerships with a select group of sponsors.

Press Kit	Speaker Package	Media Kit
alchemy.consulting/press-kit	alchemy.consulting/profile	alchemy.consulting/media-kit

SUBSCRIBE FOR THE NEXT ISSUE → alchemy.consulting/AI-BLACK-BOX