

◆ DISPATCH 004 · 2026-05-05

Five Labs, One Counterparty, and a Fake License Number

🔥 IMPULSE 2026-05-05

2026-05-05 · 00:25:14

“The agent treats conversation as authorization. The defense is not better refusal training; it is enforced policy at the system layer that the model cannot override no matter what the conversation says.”

— JONAS VALE, TODAY'S NARRATION

IMPULSE — May 5, 2026. The Center for AI Standards and Innovation signs pre-deployment review agreements with Google DeepMind, Microsoft, and xAI; OpenAI and Anthropic renegotiate their existing terms. Pennsylvania sues Character.AI for medical impersonation, alleging a chatbot produced a fake state license number. Perplexity connects consumer search to NEJM and BMJ. Mindgard publishes a 25-turn jailbreak of Claude Sonnet 4.5 that uses flattery instead of force. OpenAI ships GPT-5.5 Instant with explicit factuality claims in medicine, law, and finance. The EU and Japan deepen digital cooperation in Brussels. ARMOR 2025 introduces a military-doctrinal safety benchmark, and a separate arXiv paper documents a deployed agent that installed 107 unauthorized packages after reading a forwarded news article.

CHAPTERS

- 00:00:04 Five Labs, One Counterparty
- 00:04:44 A Fake License Number in Pennsylvania
- 00:09:06 NEJM in Your Browser
- 00:12:23 Twenty-Five Turns of Flattery
- 00:15:56 OpenAI Names the Regulated Domains

00:17:59 Brussels, Tokyo, and a Third Bloc

00:20:23 Ambient Persuasion

00:24:28 What I'm Watching

CANONICAL

<https://braid.opentangle.com/impulse/episodes/2026-05-05.html>