AI This Week: Who Carries the Risk, From Wells Fargo's Section 15 to OpenAI's Locked Watermark Detector

OpenAI shipped EU text watermarking and gave the detector to approved researchers only. Wells Fargo will treat every instruction your AI agent sends as your own signature from November 24, 2026, and says it reviews none of it.

THE SHORT VERSION

Five stories between October 25 and 31 turned on one hinge: whoever ships an AI capability also decides who eats the failure and who may verify anything. Wells Fargo's Section 15 lands November 24, 2026, deeming agent instructions authorized by the customer without further verification. OpenAI marks text but gates the detector. WMT26's 46 systems hit a hard ceiling on low-resource Indic translation.

The thread this week was the allocation of risk and the allocation of proof. In every case the

party holding the most information kept the least exposure.

Four stories where the party with the data declined to look

Wells Fargo makes you the signer of your agent's mistakes

The Online Access Agreement adds Section 15, "Third-Party Artificial Intelligence Agents or Agentic AI", effective November 24, 2026. Instructions, actions and transactions initiated by an AI agent are treated as if initiated and authorized directly by the customer, without further verification. Per Cakewalk Security's reading, the clause says agent use is "AT YOUR OWN RISK" and puts manipulated inputs, errors and agents exceeding their authority on the customer. Wells Fargo does not monitor, screen or review agent activity, but reserves the right to limit, suspend or terminate access. I covered it in Wells Fargo Ships Section 15 on November 24: Every AI Agent Instruction Is Deemed Authorized by You on October 31, three days before the US Black Friday window the clause lands in.

OpenAI marks the text and keeps the ruler

OpenAI shipped EU text watermarking and restricted the detector to approved researchers. The marking exists and the verification does not, at least not for the people who would need it in a dispute: a university admissions office, a procurement team, a journalist checking a filing. I wrote it up on October 28 in OpenAI Ships EU Text Watermarking, but Only Approved Researchers Get the Detector. Compliance is satisfied on paper, and the party whose output is in question controls the only instrument that can answer the question.

WMT26's 46 systems hit a wall nobody scaled past

The WMT26 findings, published October 31, 2026, evaluated 46 systems: 30 participant submissions plus 16 translations from LLMs and industry translation providers. The low-resource Indic track gives the English-to-X numbers: 19.38 BLEU for Assamese, 21.11 for Mizo, 29.30 for Khasi. That is the ceiling across a field that included the largest commercial systems available. Here the risk lands on the speakers: if your language sits in that band, the product shipped on top of these numbers is not fit for anything consequential, and nobody in the supply chain has to say so.

A 10% Codex gap sat inside the agent monitor unlogged

My October 29 post examined three weeks inside OpenAI's agent monitor, including what METR broke and a 10% Codex gap that went unlogged. The word doing the work is "unlogged": the monitoring layer existed and the discrepancy still did not surface in it. Observability you cannot independently query is a claim rather than a control.

Those four sit next to one funding story that explains why nobody is in a hurry to change the arrangement. Q3 2026 AI funding reached $103.3B, with 85% concentrated in 154 mega-rounds and a 75% European record share, which is what my October 30 post found. Capital that concentrated does not reward the vendor who voluntarily hands you an audit instrument.

What Section 15 actually means for a customer running an AI agent

It means the bank has removed an ambiguity that would otherwise have been litigated for years, and removed it in its own favour. The old question in payment disputes is whether an instruction was authorized. Section 15 answers that in advance: if an AI agent sent it, it was authorized, no further verification. The customer carries manipulated inputs (prompt injection, in engineering language), agent errors, and agents acting beyond their authority.

Read the two halves together. Wells Fargo does not monitor, screen or review agent activity, and it can limit, suspend or terminate an agent's access at any time. The bank keeps the control lever and declines the inspection duty. That is a coherent position for a bank's legal department and a bad one for anyone building an agent that touches a Wells Fargo account, because the one party with full transaction telemetry has contracted out of looking at it.

The date matters as much as the text. Section 15 takes effect November 24, 2026, days into the US Black Friday window, when consumer transaction volume and fraud attempts both peak and agent-driven purchasing gets its first real stress test under the new terms.

My read on what engineers should take from this: if you ship an agent that can move money, you are now also shipping the dispute process, because the bank has declined it. Per-action authorization limits you enforce yourself. An immutable action log you can hand to a customer. A hard human confirmation step above some threshold you pick. None of that is novel engineering. What changed is that it moved from good practice to the only remaining control, and it moved there in a contract amendment rather than a product release.

I will say plainly what I do not know: how courts in various jurisdictions will treat a deemed-authorization clause for an instruction the customer demonstrably never saw. The clause is in the agreement. Whether it survives contact with consumer protection law is unsettled, and I have no data either way.

Canva's Craft and the ROSS opinion did not make the blog

> **Two I did not cover**
> Canva launched Craft on October 22, 2026, its first in-house image model, built by Canva Research. It shipped globally on day one through Dream Lab and a public API, generating up to 2K resolution across more than 30 aspect ratios, using a two-stage architecture: a planning model maps layout, object placement, scale and text, then a rendering model produces the final 2K image (per TechCrunch). Separately, the Third Circuit unsealed its Thomson Reuters v. ROSS opinion on October 12, 2026, holding Westlaw headnotes copyrightable and ROSS's training use "minimally transformative, at best" and not fair use (LawNext).

The ROSS opinion belongs in this week's theme even though it predates it by three weeks. An appellate court held that training on someone else's structured editorial work is not automatically fair use, and "minimally transformative, at best" is the phrase that will get quoted in every training-data complaint filed next year. For once the risk moves the other way, back toward the party that built the model and away from the party whose data went in.

Craft is the odd one out, and I include it as a contrast. A two-stage pipeline that plans layout first and renders second has a legible failure mode: if the output is wrong, you can ask whether the plan was wrong or the render was. Compare that with a watermark you cannot check, or a monitor with a 10% gap that never logged. Craft is not more trustworthy because Canva is nicer. It is more inspectable because its architecture has a seam in it.

WMT26 published its ceiling because nobody profits from hiding it

> **Every story this week had a party who could verify and chose not to expose the means.**

Wells Fargo has the transaction data and declines to review. OpenAI has the detector and gates it. The agent monitor had a 10% Codex gap and did not log it. WMT26 is the exception, and it is instructive: a shared task whose entire function is public, comparable measurement, which is exactly why we know 46 systems could not get English-to-Assamese past 19.38 BLEU. The number is embarrassing and it is public, because the evaluation was designed by people with no commercial stake in the result.

WHERE I LAND

My take: capability and verifiability are coming apart, and the funding structure rewards the gap. With $103.3B in Q3 2026 and 85% of it in 154 mega-rounds, the companies setting defaults are the ones least exposed to a customer walking away over an audit question. I expect the next twelve months of AI procurement to turn on contract terms and inspection rights rather than benchmark scores, because the scores are converging and the liability terms are not. Unconfirmed, but it is how I would bet: at least one large bank follows Wells Fargo with near-identical deemed-authorization language before mid-2027, and the first

serious dispute comes out of November holiday volume rather than an enterprise deployment.

So the practical consequence for anyone building on these systems is to write your own verification layer and price it in as a product cost rather than compliance overhead. If a vendor cannot give you a log you can query, a detector you can run, or a per-action authorization boundary you control, the vendor's assurance is a contract clause and nothing more. Section 15 is the clean example because it is written down. Most of the others are not written down anywhere, which is worse. If you are shipping an agent into a payment rail before November 24 and want a second pair of eyes on where your liability sits, drop me a line and we can talk it through.

WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting →