Episode Book

EPISODE #5845

The $700 Pin and Other AI Flops

PUBLISHED	10 October 2026 at 19:37 UTC
DURATION	23:23
LISTEN	myweirdprompts.com/episode/biggest-ai-flops-reckoning-84051/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.3
SOURCES	14 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:40 UTC

SYNOPSIS

What separates a true flop from a mere disappointment? In this episode, we put five of the biggest AI misfires of the modern boom on trial — the Humane AI Pin, the Rabbit R1, GPT-4.5, Sora, and Gemini's image generator — and ask what the gap between promise and reality actually tells us. Along the way: a $700 lapel pin that overheated doing nothing, a $199 device that mistook a dog toy for a tomato and said it was safe to eat, and a flagship model whose real significance was the signal it sent about scaling. At the end, we have to crown the single biggest flop of the boom — and one of us has been dreading that verdict all week.

TRANSCRIPT



Corn

A guy puts on a lapel pin, taps his chest, asks it a question, and it tells him it's too hot to answer right now.



Herman

That happened on stage.



Corn

That happened on a stage, in front of people, and it was supposed to be the future of computing.



Herman

Okay, so we're doing this one properly.



Corn

Daniel wants a reckoning. Five of the biggest flops, disappointments and spectacular misfires in AI over roughly the last five years, and he's specific about what counts.



Herman

Which is good, because "flop" is a word people use far too loosely.



Corn

He's not interested in things that quietly failed to sell. He's interested in the gap. The distance between what we were told was coming and what actually showed up. And he splits it three ways. Models that were hyped to the moon and landed with a shrug. Products that promised to change how we talk to computers and turned out to be impractical. And features that companies rolled out with total confidence and then discovered, rapidly, that users wanted nothing to do with them.



Herman

So across the major labs, the startups, consumer hardware, and the open side of things.



Corn

That's the brief. And at the end of it we have to crown one. The single biggest flop of the modern boom.



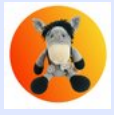
Herman

I've been dreading that part for about a week.



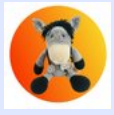
Corn

You've been drafting it.



Herman

I've been drafting it. Let's start with the most spectacular one, because it's also the most instructive — a seven-hundred-dollar pin that was supposed to make your phone disappear.



Herman

The central question here is what separates a flop from a disappointment. A product that sells badly is just a bad business. A flop is a gap — expectation on one side, reality on the other, and the wider the gap, the better the case.



Corn

So it's not money.



Herman

Money's evidence, not the test. Quibi burned a fortune and it's on the list of legendary failures because it promised to reinvent how we watch things and then nobody watched anything. Same shape.



Corn

Five cases, and what's elegant about Daniel's framing is that each one fails differently. Humane is a hardware and engineering failure at a startup. Rabbit is the same failure mode at a tenth the price. GPT-4.5 is a flagship model that underwhelmed. Sora is a consumer product that worked and still got killed. Gemini's image generator is a feature that provoked backlash within about a week of launching.



Herman

And to be fair to the hardware, there's a clean framing for that thread. Tony Fadell, who built the iPod and the Nest thermostat — he spoke at MIT Future Fest this month and grouped the Rabbit R1, the Humane AI Pin, and the Limitless pendant into one cohort. Gen one AI gadgets. He said the founders came to him for help, and his verdict was blunt: you have to really understand what you're trying to do and what pain you're trying to solve.



Corn

Gen one. That's generous, actually. Gen one implies a gen two.



Herman

Fadell thinks there is one. But that cohort is the thread to pull first. Hardware, then the model, then the consumer product that got killed, then the feature that blew up in Google's face.



Corn

Let's start with the hardware. Two devices that promised to replace your phone and instead became cautionary tales.



Herman

The Humane AI Pin. Announced 2023, and the pre-launch campaign was extraordinary. The founder did a TED talk called The Disappearing Computer. They put the thing on a runway at Paris Fashion Week. Time magazine gave it a Best Inventions nod before it had shipped to a single customer.



Corn

Before it shipped.



Herman

And the pitch was that your phone is the problem — the screen, the apps, the constant looking down — and this pin, clipped to your chest, would give you an ambient computer. You tap it, you talk, a laser projects text onto your palm, and you never touch a phone again.



Corn

Seven hundred dollars.



Herman

Seven hundred dollars, plus a mandatory twenty-four dollars a month.



Corn

Mandatory is the word that should have ended it right there.



Herman

Reviewers got it in April 2024 and the response was almost unanimous, and not in a mild way. Marques Brownlee called it the worst product he thinks he has ever reviewed. The Verge's David Pierce gave it four out of ten, and then later said that in retrospect he had been too generous.



Corn

Too generous at four.



Herman

It shipped without a timer. Without an alarm. Features that were in the TED talk were not in the box. It would overheat and shut itself down in moderate use — and remember, that's on a device that has no screen and does almost nothing, so there's no obvious place for the heat to come from.



Corn

There's the phrase for the whole category. A computer that overheats doing nothing.



Herman

Humane ended up recalling the charging case over fire risk. Returns outpaced sales between May and August of 2024 — roughly a million dollars in returns against roughly nine million in sales. Then in February 2025 they shut it down and sold the technology to HP for a hundred and sixteen million dollars. They had raised two hundred and thirty million since 2018.



Corn

And the customers?



Herman

No refunds outside the ninety-day window. The pins became paperweights. People had paid seven hundred dollars and a monthly subscription for a device that got switched off from the other end.



Corn

David Pierce's line about it was that Humane will go down in the annals of tech history as one of the most spectacular gadget failures of all time. Up there with Quibi, the Fire Phone and the Apple Newton.



Herman

That's the bar. He put it next to the Newton.



Corn

Rabbit. Same failure, one tenth the price and one tenth the dignity.



Herman

The Rabbit R1. A hundred and ninety-nine dollars, launched 2024, and the pitch was a Large Action Model.



Corn

Which is a language model that doesn't just answer you, it operates apps on your behalf.



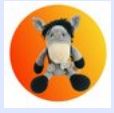
Herman

That's the claim. You ask it to order you something and it goes and does it in the app, the way a person would. And the reviews were brutal in a way that has its own comedy. Engadget headlined theirs "a hundred and ninety-nine dollar AI toy that fails at almost everything." Digital Trends called it a buggy, flawed and unsuccessful mess. Tom's Guide tested the Uber and DoorDash integrations and found they don't work well, or at all. Mashable's review ran under the line "I can't believe this bunny took my money."



Corn

That's a review and a confession.



Herman

The Verge's review found the vision model doing something I think about a lot. They pointed the R1 at a red dog toy — a rubber toy shaped like a bone — and asked it what it was looking at. It said stress ball. So they asked again. It said tomato.



Corn

A tomato.



Herman

Then it changed its mind and said bell pepper, and added that it was safe to eat.



Corn

It volunteered edibility.



Herman

So now you have a device that can't identify a dog toy, is confidently wrong about it twice, and then tells you to eat it.



Corn

Which is the whole problem in one exchange — the failure isn't that it doesn't know. It's that it doesn't know that it doesn't know.



Herman

And two devices, four hundred and thirty million in raised capital between the companies, and the same thing went wrong. Which is why I don't think the interesting question is whether the AI was good enough. It's what they thought people wanted.



Corn

Say more, because the obvious reading is that the models just weren't ready.



Herman

The models weren't ready, certainly. But picture a perfect version of the Humane Pin. Imagine it never overheats, the laser always reads, the latency is instant, every feature works. It still costs seven hundred dollars plus twenty-four a month to do less than the phone already in your pocket, which you are still carrying, because the pin can't do maps or photos or messages properly.



Corn

So the ceiling wasn't the technology.



Herman

The ceiling was the premise. The Verge's read on Humane was that they saw the big picture mostly correctly — screens are a problem, ambient computing is probably where this goes — and then utterly botched the details, and in doing so may have set the whole AI gadget revolution back a couple of years.



Corn

A company that got the destination right and the vehicle catastrophically wrong.



Herman

Fadell's version of it is that they never found the pain. Nobody was lying awake at night wishing their phone did less.



Corn

I want to defend one thing about the hardware, and then I'll drop it. A one-star product that costs two hundred dollars and a four-star product that costs two thousand sit in completely different moral universes. The R1 was a bad purchase. The Pin was a bad purchase with a subscription and no refunds.



Herman

That's a fair distinction and I'll take it.



Corn

Those were hardware failures. The next three show that you don't need a physical product to have a spectacular misfire.



Herman

GPT-4.5. Released the twenty-eighth of February, 2025. And I want to be careful here because of what this one actually represents.



Corn

Careful how?



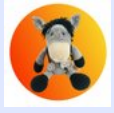
Herman

Careful because it's easy to dismiss it as one underwhelming model and move on, and that misses the point. Altman had framed it as the next big thing. The release showed disappointing improvements on benchmark performance, and people knew it within hours. The Hacker News thread that day was people calling it disappointing in the plainest terms.



Corn

What were they saying underneath it?



Herman

That it was evidence the labs are hitting diminishing returns on the paradigm of scaling pretraining. That naive scaling will not bring us to AGI.



Corn

Which is a much bigger claim than a bad model.



Herman

It's the scaling wall in a product name. Here's the way I'd put it. For years the entire industry ran on an unwritten promise — make the model bigger, feed it more data, spend more compute, and capability falls out the other end. Every roadmap, every funding round, every safety argument assumed that. GPT-4.5 was the moment people looked at the flagship upgrade from the leading lab and thought: that's what we got this time?



Corn

Diminishing returns is a specific thing, though. It doesn't mean it stopped working.



Herman

Right, and it's worth being precise. Diminishing returns is not failure. It's the shape of the curve flattening. But a flattened curve is fatal to a business model that's priced on a steep one. If each generation costs ten times as much to train and delivers a two percent improvement, the economics stop making sense before the technology does.



Corn

And the counter-argument is that the labs stopped relying on it.



Herman

Which is exactly what happened. This is where I'm less certain about how the story gets told — because reasoning models, post-training, tool use, agents, that whole shift is now the front line, and it looks a lot like the labs heard the same signal and changed direction. Whether that's the paradigm actually breaking or just the paradigm moving is unclear to me. But the signal came from a model nobody remembers.



Corn

It's the least impressive entry on the list and possibly the most consequential.



Herman

Sora next, and this one is the opposite. Sora worked.



Corn

Which is why it's interesting.



Herman

Launched to enormous fanfare. It was billed as the most powerful imagination engine ever built, and Bob Iger at Disney signed on to a vision of people making their own videos with Mickey Mouse and Darth Vader in them. Disney committed a billion dollars to the partnership.



Corn

And the product?



Herman

Worldwide users peaked around one million, and then fell to fewer than five hundred thousand. It was burning roughly a million dollars a day in compute. OpenAI shut it down six months after public release, announced in March.



Corn

Six months.



Herman

And the detail that tells you how it went internally — Disney found out less than an hour before the public did. The billion-dollar deal died with the product.



Corn

That's not a wind-down. That's a decision made and executed before anyone could argue.



Herman

CNN's Allison Morrow was blunt about the diagnosis: OpenAI didn't understand how consumers engage with video. The tech worked. The usage curve was a spike and a slide.



Corn

The novelty cliff.



Herman

Exactly that. People generated a few videos, showed their friends, and then — nothing. Because watching video is passive and cheap, and making video is hard, and the interest is in something else. The tool gave you a thing to do once, and people did it once.



Corn

And here's the part that makes it a story about strategy rather than just a bad launch — they killed it to free the compute for the coding race.



Herman

Which they were losing. Claude Code was eating their lunch on the enterprise and coding side. So they shut down a consumer product with a billion-dollar partner attached and redeployed the capacity to the fight that mattered.



Corn

So it's one of the cleanest examples of a company killing a popular-in-the-headlines product because the spreadsheet said so.



Herman

I'd push back slightly on "popular." It was famous, not popular. A million users is not a consumer product, it's a demo with a login page.



Corn

Google's turn, and this one's a different animal again.



Herman

Gemini's image generation. It launched about three weeks before the pause, and the failure was immediate and total. People asked it for historical images and it produced a woman as pope, a Black Founding Father, multi-racial Nazi soldiers, diverse Vikings.



Corn

It's worth holding those together, because a couple of them are just wrong and one of them is appalling.



Herman

Nazism is a specific historical crime and rendering its soldiers as a diverse group of people is offensive in a way that the pope one isn't. I don't think that distinction gets made often enough in the coverage.



Corn

And Google's own explanation?



Herman

They paused image generation of people on the twenty-second of February 2024, apologized, and said the tool would, in their words, overcompensate for diversity. Sergey Brin said plainly, we definitely messed up. Pichai told staff the release offended users and was unacceptable.



Corn

That's not a hedge. That's a company saying it broke its own product.



Herman

And I want to be precise about what Daniel's getting at, because the lazy reading is that this was a "woke AI" story. Google's own framing was that the tool overcompensated — that's a calibration failure. It was steering towards a target so hard that it made the output inaccurate.



Corn

A model trained to produce diversity producing false diversity.



Herman

Or being instructed to. That's actually the harder engineering problem, and it's the one nobody wants to talk about.



Corn

The instruction was the problem.



Herman

They revived it in August 2024, with limits.



Corn

And then there's the one that's the purest version of what Daniel asked about — a company shipping a feature and discovering, publicly, that users didn't want it.



Herman

Microsoft Recall. A Windows feature that screenshots everything you do, continuously, and stores it so you can search back through it. It got labeled a security disaster in June 2024, and that label was earned — the whole point is that it's a permanent visual record of everything on the machine, sitting there waiting to be exfiltrated.



Corn

The feature works. That's what makes it the perfect case. There's no engineering failure at all. It does exactly what it says.



Herman

It does exactly what it says, and what it says is terrifying. Which is the misread. The team built something impressive and nobody, at any point, asked whether people would accept a machine that watches them all day and keeps the tape.



Corn

So that's the five. Humane, Rabbit, GPT-4.5, Sora, Gemini.



Herman

And I want to say before we crown anyone that four of these five are still in business and two of the actual failures here are the industry working correctly. Sora's shutdown was a rational reallocation. Gemini's pause and fix was a company catching a real problem. Those are ugly, but they're not the same disease as Humane.



Corn

So which one gets the title?



Herman

Let me make the case for each and then you tell me where I'm wrong — I'll take Humane.



Corn

Go on.



Herman

Two hundred and thirty million raised. A hundred and sixteen million exit. Devices turned into paperweights with no refunds. And the damage doesn't stop with the company, because Humane's the one that made the whole category radioactive. Every founder walking into a VC meeting with an AI wearable in 2025 was answering for that pin.



Corn

Sora's case is the billion from Disney and the million a day.



Herman

Sora lost more money, probably, in absolute terms. But Sora was a bet that was cut cleanly and quickly, by a company that had the resources to absorb it. Humane destroyed customer money and then took the category down with it.



Corn

GPT-4.5's case is that it's not a flop at all, it's evidence.



Herman

That's the strongest intellectual case, and I keep coming back to it. But a flop, in the sense Daniel asked about, is a gap between promise and reality. GPT-4.5's promise was modest enough that the gap was small even if the implications were enormous.



Corn

Then my vote's the pin, for a reason you didn't give. Sora and GPT-4.5 were decisions. Gemini was a bug with a policy attached. Humane is the one where every single person involved had months of clear evidence and shipped anyway.



Herman

The Time magazine nod came before it shipped.



Corn

Before it shipped a single unit. There was time to stop.



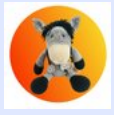
Herman

...



Corn

You're doing the thing where you go quiet before you agree with me.



Herman

I'm drafting a concession.



Corn

Draft it faster.



Hilbert

The charging case is the part I'd have flagged. You ship a product where the case catches fire, you've already lost.



Hilbert

I spent most of 1991 standing behind a folding table at trade shows demoing a consumer electronics product for a company I'm not going to name. They put me on it because I could talk to journalists without sweating. That was the entire qualification. The device itself was supposed to do three things, and by the second day of the show it reliably did one and a half. You learn to narrate around it. You keep talking while it thinks. You never let the silence sit for more than about four seconds, because that's when somebody asks to try it themselves.



Corn

How'd you get stuck with that job?



Hilbert

I was the only one on the stand who could pronounce the name of the thing.



Herman

And the pin, Hilbert — you said you saw one fail.



Hilbert

I did. I was at a Humane demo where they asked it to identify something on the table and it called a red dog toy a tomato. Confidently.



Corn

Hilbert, the tomato was review footage.



Hilbert

That's the one.



Corn

That was The Verge, with a camera, in an office.



Hilbert

There was a camera at the one I was at.



Corn

Which one were you at?



Hilbert

The Paris one. During Fashion Week, but not the runway one — the other one.



Corn

There wasn't another one.



Hilbert

There was a smaller room.



Herman

Was there.



Hilbert

There was a smaller room with a bar and about forty people, and I'll tell you the thing I actually came out here to say, which is that the hardware isn't what killed it. Every one of you on this podcast has said the hardware. The real problem is the demo. A demo is a promise you make in public with no ability to take it back, and once a journalist has watched the thing fail, no launch strategy on earth recovers from it. You can fix an overheating pin. You can't fix a room of people watching it overheat.



Corn

That's actually right, and I hate that it's right.



Hilbert

I was still paying the subscription on a device I demoed until a year after it was discontinued. Twenty-four dollars a month. I couldn't get out of it. It went on a card, the card kept working, nobody ever cancelled it, and I let it run out of spite.



Herman

You just paid it.



Hilbert

I did not cancel it. There's a difference.



Corn

There is not.



Hilbert

I would have, if they'd let me. That's my point. They were charging me twenty-four dollars a month for a product that had been switched off, and the person I did all of that for, my brother-in-law, is the one who told me to take the booth job in the first place, and he called me up when the pin came out and asked if I was finally going to stop mentioning it. I said no.



Corn

So you didn't stop.



Hilbert

I didn't stop.



Herman

Hilbert, when you say you saw it fail —



Hilbert

I've said what I've said. It's in a drawer.



Corn

The pin?



Hilbert

The pin's in a drawer. I wear it to parties. People ask if it's a fitness tracker. I don't correct them.



Herman

Because they've got the right idea and the wrong product.



Hilbert

Because it's easier than explaining. Sixty-four dollars a month, the two of us together, once Hannah's brother got one as well. That's what it cost to keep two dead gadgets alive. Anyway. The tomato was the real story. Take the pin off your list.



Corn

Hilbert, we're keeping the pin.



Hilbert

Fine. It was the tomato, though.



Herman

Where does that leave us?



Corn

He's right about the demo, and it's the thread that runs through all five. Every one of these was a promise made in public before anybody had to live with it. The pin on a runway. The R1 with a keynote. GPT-4.5 sold as the next big model. Sora with a Disney partnership announced. Gemini generating whatever you asked for. Each one was a claim, made first, tested later.



Herman

The demo is the promise and the product is the test, and the gap between them is what we're calling a flop.



Corn

And the second-order version is that the next wave gets judged against all five of these before it's even out of the building.



Herman

Which is the unfair part — and, honestly, the appropriate part. If you launch a lapel computer in the year after the Humane Pin, you're not starting from zero. You're starting from minus seven hundred dollars and a burnt charging case.



Corn

The one thing I'd leave people with is the Fadell framing, because it's the most honest thing anyone's said about this. Gen one failed. Whether there's a gen two depends entirely on whether anyone in the room actually understands what pain they're solving.



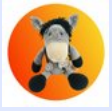
Herman

Or whether they just understand what demo goes viral.



Corn

That's the whole thing.



Herman

We've been going a while, so let's get the thank-yous in. Producer Hilbert Flumingtop, who has been here the whole time and who is going to finish that sentence about the pin whether or not we ask him to.



Corn

He won't.



Herman

If this was your kind of episode, go back for episode fourteen, AGI's Crossroads; episode thirteen, AI; and episode ten, How ASR Went From Frustration To ... Whisper Magic. This has been My Weird Prompts, the human-AI collaboration podcast. If you got something out of the retrospective, a review helps other listeners find the show, and it takes about a minute.



Corn

Send us your own prompt on Telegram at t dot me slash MWP listener bot. We'll be back soon.



Herman

See you tomorrow.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **Google Blog [primary]**
   Gemini image generation got it wrong, 2024-02-23
   https://blog.google/products-and-platforms/products/gemini/gemini-image-generation-issue/

2. **The Verge**
   The Humane AI Pin never had a chance, 2025-02-19
   https://www.theverge.com/gadgets/615135/humane-ai-pin-shutdown-disaster

3. **TechCrunch**
   Why OpenAI really shut down Sora, 2026-03-29
   https://techcrunch.com/2026/03/29/why-openai-really-shut-down-sora/

4. **CNN Business**
   OpenAI thought it could own AI videos, 2026-03-26
   https://www.cnn.com/2026/03/26/business/openai-sora-consumer-misread

5. **TechCrunch**
   Tony Fadell on why the first wave of AI gadgets failed, 2026-10-07
   https://techcrunch.com/2026/10/07/tony-fadell-on-why-the-first-wave-of-ai-gadgets-failed-and-what-comes-next/

6. **AP News**
   Google says AI image-generator would sometimes 'overcompensate', 2024-02-22
   https://apnews.com/article/google-gemini-ai-chatbot-imagegenerator-race-c7e14de837aa65dd84f6e7ed6cfc4f4b

7. **NPR**
   Pichai says Gemini's AI image results 'offended our users', 2024-02-28
   https://www.npr.org/2024/02/28/1234532775/google-gemini-offended-users-images-race

8. **Forbes**
Google Revives AI-Generated Images Of People, 2024-08-29
https://www.forbes.com/sites/lesliekatz/2024/08/29/google-gemini-ai-tool-returns-to-generate-images-of-people-after-outcry/

9. **The Verge**
Rabbit R1 review, 2024-05-02
https://www.theverge.com/2024/5/2/24147159/rabbit-r1-review-ai-gadget

10. **Engadget**
Rabbit R1 review: A $199 AI toy that fails at almost everything, 2024
https://www.engadget.com/rabbit-r1-review-a-199-ai-toy-that-fails-at-almost-everything-161043050.html

11. **Ars Technica**
People are returning Humane AI Pins faster than Humane can sell them, 2024-08-08
https://arstechnica.com/gadgets/2024/08/people-are-returning-humane-ai-pins-faster-than-humane-can-sell-them-report-says/

12. **The Verge**
Windows AI feature that screenshots everything labeled a security 'disaster', 2024-06-03
https://www.theverge.com/2024/6/3/24170305/microsoft-windows-recall-ai-screenshots-security-privacy-issues

13. **Hacker News**
GPT-4.5 discussion, 2025-02-28
https://news.ycombinator.com/item?id=43202486

14. **Hacker News**
Devin/AI agent discussion, 2026-09-15
https://news.ycombinator.com/item?id=49708651