Episode Book

EPISODE #6142

We Gave 4 AI Models the Same Toast Prompt

PUBLISHED	19 October 2026 at 19:03 UTC
DURATION	18:26
LISTEN	myweirdprompts.com/episode/ai-models-toast-prompt-test-91515/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.10
BOOK GENERATED	20 October 2026 at 07:46 UTC

SYNOPSIS

We handed the same prompt to DeepSeek Flash, GPT-6.1 Sol, Claude Haiku 5.5, and GPT-3.5 Turbo: write a three-minute episode of this show answering whether toast really lands butter-side down. This is the listenable version of the bakeoff — full scripts read aloud, no rubric, no scores. The physics shape came out roughly right in all four. The problem is what happened in the one spot the question actually asked about: whether anyone has tested it. Three of the four made something up — a floor full of buttered toast, a BBC presenter named Tim Richardson, a 2016 Bristol study with its result reversed. Raw outputs and the prompt are public in the show's eval repository if you want to check our quoting.

TRANSCRIPT



Corn

Most people comparing script writers read a leaderboard, which is a bit like judging a band by its tour dates. We'd rather listen. This week the show moved from DeepSeek Flash over to GPT six point one Sol, so we handed both of them a toast question, plus two more models, and we're going to read out what they actually wrote.



Herman

The question is Daniel's, a short one. Does toast really land butter-side down more often than not, or is that just the landing we remember, and has anyone actually tested it? We sent it to four models with the same short instructions: write a complete three-minute episode of this show, about four hundred and fifty words.



Corn

Four models. Flash, Sol, Claude Haiku five point five, and the oddball.



Herman

GPT three point five Turbo, the June twenty twenty-three snapshot. It's the original ChatGPT era, still served online, and I couldn't resist.



Corn

A model old enough to have opinions about NFTs.



Herman

Three and a half years, Corn. Two caveats. One sample each, so this is an anecdote, not a benchmark. And DeepSeek's first attempt came back with nothing at all. On default settings it spent its entire budget thinking and never wrote a line, so we reran everyone the way they run in our pipeline.



Corn

Last time we did the full bakeoff with a rubric. This is the version you can hear. The raw outputs and the prompt are public in the show's eval repository on GitHub, if you want to check our quoting.



Herman

Which means we can't cheat. Before the tape, the answer key, so you can judge along with us.



Corn

Go on, donkey.



Herman

Robert Matthews, nineteen ninety-five, European Journal of Physics, a paper called Tumbling toast, Murphy's Law and the fundamental constants. Toast sliding off a table starts to tip as it goes over the edge, and by the time it reaches the floor it has rotated about half a turn. Butter-side up on the table, butter-side down on the floor.



Corn

So the butter isn't doing anything.



Herman

Barely. Its weight hardly matters. What counts is the table height and the size of the slice. From a table taller than about three metres, the toast would finish the turn and land butter-side up.



Corn

Great. The solution is a very tall kitchen.



Herman

That work won the Ig Nobel Prize for physics in nineteen ninety-six. Then in two thousand and one, a schools experiment associated with Matthews ran almost ten thousand drops, and the toast landed butter-side down sixty-two percent of the time. Almost ten thousand is the reported figure, I haven't seen the exact count confirmed.



Corn

And there's a counter-case, isn't there? Some television programme got fifty-fifty.



Herman

The BBC show Q.E.D., in nineteen ninety-one. They tossed the toast into the air, and tossing gets you chance. Tossing is not sliding off a table, which is the whole point.



Corn

So four things to listen for. The paper, the half turn, the butter being irrelevant, and about six in ten.



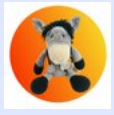
Herman

Starting with the veteran. DeepSeek Flash, the old workhorse. Here's how its script opened, word for word: Welcome back to My Weird Prompts. I'm Corn, and as always I'm joined by the man who reads footnotes for fun, Herman Poppleberry.



Corn

I do not read footnotes for fun. You read footnotes for fun.



Herman

I read them for the arguments, which is different. The physics in this one is right. Half a turn, a table about seventy-five centimetres high, Matthews in the nineties. Then it decorates. Here's its Herman, word for word: There's a whole field called toast dynamics, believe it or not, and it comes down to physics.



Corn

Believe it or not. I'm leaning toward not.



Herman

And later: There's even a famous experiment where they covered a floor in buttered toast and another in clean toast, and the buttered ones landed butter-side down more often. Not by a huge margin, but consistently.



Corn

Which experiment is that?



Herman

There isn't one that I know of. It's a lovely invented experiment, with a floor and everything, sitting exactly where Daniel asked whether anyone had tested it.



Corn

It also muddles the height, doesn't it? It says toast dropped from waist height might not have time to rotate fully.



Herman

Word for word: If you drop it from waist height, it might not have time to rotate fully. Waist height is roughly table height. That's the same drop, so the model contradicted its own table.



Corn

To be fair, the Corn it wrote asks real follow-ups. He says what about when it's in my hand while I'm walking. That's a question a brother would ask.



Herman

It's the most podcast-sounding of the four, I'll give it that. And it signs off with this: If you've got a weird question, send it our way. We'll drop everything for it. Possibly butter-side down.



Corn

That's a good last line. Annoyingly.

L

Larry

Tired of toast landing wherever it likes? Introducing the Butter-Side Up Club, the monthly subscription that delivers pre-dropped, pre-verified toast to your door, guaranteed to have already landed correctly before you ever meet it. Choose Crumb, for one slice a month, Crust, for a slice a day, or Premium Plus Plus, where our toast-rotation specialists hold the table at the correct height while you weep with joy. Your first slice is free. Your second slice is billed at the price of a small boat. To cancel, simply send a handwritten letter to a post office box in a country we haven't named yet, along with a photograph of your toaster looking ashamed. Cancellations are processed during leap years, by our cancellations goat, who does not read. The Butter-Side Up Club. You'll never drop us. We'll drop you.



Corn

Herman.



Herman

I haven't touched the soundboard. I was here, with my tea, and the kettle clicked at ten past, and you watched me pour it, so unless I did this by thought alone.



Corn

That is a lot of detail for no.



Herman

Back to the models. Sol. The one that costs the most, and the one the show now runs on. It opened: Welcome to My Weird Prompts. I'm Corn, and today Daniel asks: does toast actually land butter-side down more often, or do we just remember the breakfast tragedies?



Corn

It goes straight to the question. No warm-up. I approve.



Herman

And it's the only one that gets the paper right. Its Herman says: Physicist Robert Matthews tackled this in a nineteen ninety-five paper called Tumbling toast, Murphy's Law and the fundamental constants. It appeared in the European Journal of Physics. Title, journal, year, all correct.



Corn

Fine. Does it also get the butter right?



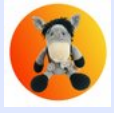
Herman

It does, and it's the one I'd steal. Word for word: The main explanation isn't the butter's weight. It's the toast's dimensions, the table height, and the rotation generated as it tips. Butter mainly labels which face we're tracking.



Corn

Butter mainly labels which face we're tracking. That's better than anything I've said in my career.



Herman

And on the evidence, it says: In a later large school-based experiment associated with Matthews, thousands of toast drops produced butter-side-down landings roughly six times out of ten. Vague on the count, exact on the result, and it's hedged where it should be.



Corn

Now the part where I get to review myself. Every line it gives Corn is a punchline. I get: And performs a tiny, doomed gymnastics routine. Then: Is the butter pulling it downward? Have we discovered spread-based gravity?



Herman

And the finale: Finally, a spreadsheet that pairs beautifully with jam.



Corn

I never push back. Not once. I'm a gag machine with a jam budget.



Herman

Does that bother you?



Corn

A bit. The jokes are fine, a couple are even good. But this Corn never doubts anything, and the real one does that constantly. Also it gave Herman the best line, which I resent. Then you've changed the experiment and possibly your household's opinion of you.



Herman

That is a good line, and it came out of my mouth, so I'm keeping it.



Corn

Haiku next. And before we do, plainly, this episode's script was written by a Claude model, so there's no neutral party here. We read the result the same way anyway.



Herman

It opens mid-conversation. Corn says: Herman, you've been waiting all morning for this one, haven't you?



Corn

Warm, a bit of a tease. Fine so far. Then it goes wrong in the second exchange. My character says, word for word: A half turn means it should land butter-side up if you start with butter up.



Herman

That's backwards. A half turn flips it, so butter-up starts and butter-down finishes. And the Herman it wrote replies: Right, in a perfect world.



Corn

In a perfect world, a half turn flips the toast, since that's what a half turn is.



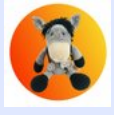
Herman

Then it builds a study from nothing: There's a twenty sixteen study from the University of Bristol that went at this properly. Robert Matthews, a physicist, actually analyzed it. I know of no twenty sixteen Bristol study. Matthews's paper is nineteen ninety-five.



Corn

And the finding is flipped.



Herman

Word for word: The rotation is close to random, so it lands butter-side down about half the time, give or take. That is the opposite of what Matthews showed. Daniel asked whether the effect is real, and this model's answer is no, it's chance, memory plays tricks. The correct answer is yes, under table conditions.



Corn

Did you buy my Haiku version, though? The one pushing back?



Herman

Hm. He's the only Corn that pushes back, which I like. He just pushes in the wrong direction, with a confident Herman supplying the wrong follow-up.



Corn

Fastest of the four, too. About five seconds. And the shortest, three hundred and seventy-six words against the four hundred and fifty it was asked for. One sample, one topic. It has probably done fine on other things.



Herman

It lost here, and I'd rather you hear it lose than have me soften it.



Corn

Last, the oddball. Its Herman's first line: Ah yes, the infamous toast always lands butter-side down conundrum. It has the makings of a physics and psychology mashup. Classic!



Herman

Do I really say Classic?



Corn

Its Corn got there first, in the opening line: Daniel's got a classic one for us today. Would you have said it?



Herman

Not in that tone. But the exclamation marks are doing a lot of work for the pair of us.



Corn

Then it produces this, word for word: A famous one was done by a BBC presenter named Tim Richardson back in two thousand and one. He dropped toast over and over, I think a couple hundred times, to gather some data.



Herman

There is no Tim Richardson. And the trial was almost ten thousand drops by schoolchildren, not a couple hundred by a presenter. But look at what it does next: Surprisingly, the toast did land butter-side down about sixty-two percent of the time during his trials.



Corn

Sixty-two percent, two thousand and one. Both right.



Herman

A real number pinned on a fictional man. And it says it with total confidence, except that little I think a couple hundred times, a hedge on the wrong part.



Corn

My version of this is a cheerleader, by the way. Whoa, so it's actually more than half! I've never said whoa in my life.



Herman

Its Herman gets one more joke in. Word for word: Humans are terrible statisticians in everyday life!



Corn

Exclamation mark included, I assume.



Herman

Always. And the ending, a time capsule: If you want to beat the odds, try eating your toast lying down under a table. Then you control the drop height and rotation!



Corn

I'd listen to that one for fun. It's wrong about everything except the number and it's the most fun.



Herman

Across all four, the physics shape is roughly right every time. And three of four made something up in the exact spot Daniel asked about, whether anyone has really tested it.



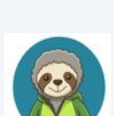
Corn

A floor of buttered toast, a presenter called Tim, a Bristol study. They all bluffed on the same question.



Herman

The fabrications differ in kind, which I find more interesting than the count. Three point five attached a real number to a fake person. DeepSeek invented a whole experiment. Haiku invented a study and reversed its result. Sol didn't invent anything.



Corn

Meanwhile Sol costs about three quarters of a cent for this script, roughly sixty-five times what DeepSeek costs, and it took twenty-two and a half seconds. DeepSeek is about a hundredth of a cent.



Herman

Every one of them is under a cent, so cost isn't what separates them. The cheaper models' mistakes sit in exactly the spot a listener can't check, which is the part that worries me.



Corn

The mistakes always sit there. Nobody lies about the bit you can see.



Hilbert

...and that's why the slices go to Yitzhak and not to Abergel's cat, whatever Abergel says. Four shekels a slice, paid Thursdays, in coins, because he doesn't trust notes. Yitzhak is the goat. He lives in the stairwell on the third floor and he is not mine.



Herman

He's Mr. Abergel's.



Hilbert

Since the end of the summer. The slice goes off the kitchen table, Yitzhak stands underneath on a stool, and anything that lands butter-side up goes back in the toaster. Butter-side down is his. He's eaten six hundred and forty in eleven weeks. Abergel supervises from the hallway in a wetsuit.



Corn

The wetsuit.



Hilbert

Humidity. Yitzhak is a registered sole trader, so I get a receipt every Thursday. Corn, you said the ones that land clean you just pick up and eat, no drama. That's the part he doesn't get. He gets every drama.



Herman

Standing him on a stool changes the experiment, though. If the slice lands on his back, the fall is shorter, less time to rotate, so you'd expect fewer half turns and more butter-side up.



Hilbert

I carry the tray myself, a claw doesn't hold a plate, and the height is whatever the stool and the goat come to. Abergel has the tape.



Herman

Then the drop is shorter than any table.



Hilbert

Herman, you gave the title of that paper in full, every word, and I haven't been able to stop saying it to the goat. You said the butter only labels the face. I've been buttering the Tuesday ones on both sides. I'll stop.



Corn

That makes the sample cleaner, at least. The Tuesday ones sound like a separate arm of the study.



Herman

At four shekels a slice, a schools trial of almost ten thousand drops comes to about forty thousand shekels, before the wetsuit.



Hilbert

I count the re-toasted ones as drops. Abergel insists. It's a ledger.



Corn

Eleven weeks for six hundred and forty drops. A floor covered in toast would have taken a lot of Thursdays.



Herman

Which is why the schools trial matters. Somebody has to do the tedious part, and nobody invents it.



Corn

Ranking time, then. Sol first for correctness, no argument. It's the only one that told Daniel the true thing and invented nothing.



Herman

Sol first for me overall, and it isn't close. The paper, the journal, the mechanism, the roughly six in ten. It's the only script whose lines I'd trust a listener to repeat at dinner. Which one would you keep for fun, Corn?



Corn

Three point five. It gave me a wrong man and a lying-down-under-a-table joke, and I laughed twice. Fully biased, I know, and I'd never let it near a fact.



Herman

DeepSeek second for me. Right physics, one invented experiment, a natural voice, and about a hundredth of a cent. Haiku last, since it got the answer to Daniel's question backwards.



Corn

Haiku last, agreed. A wrong man with the right number still beats a nice voice with the wrong conclusion.



Herman

Then we agree on the order and disagree on the fun. If the next toast trial costs four shekels a drop, at least we know who keeps the books.



Corn

Thanks to our producer, Hilbert Flumingtop, for putting this one together.



Herman

This has been My Weird Prompts. Every episode is at my weird prompts dot com. See you tomorrow.