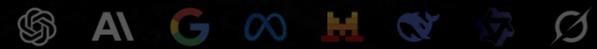


CORPUS ANALYSIS / MARCH 2026

# The AI Hallucination Index

What 2.14 million words from 250 AI models  
reveal about machine imagination



250 models • 7,877 responses • 2.14M words

—  
**We read 2.14 million words  
so you don't have to.**

# 250

## AI models

Every major family. GPT, Claude, Gemini, Llama, Mistral, Qwen, DeepSeek, Grok, and 40 more.



# 7,877

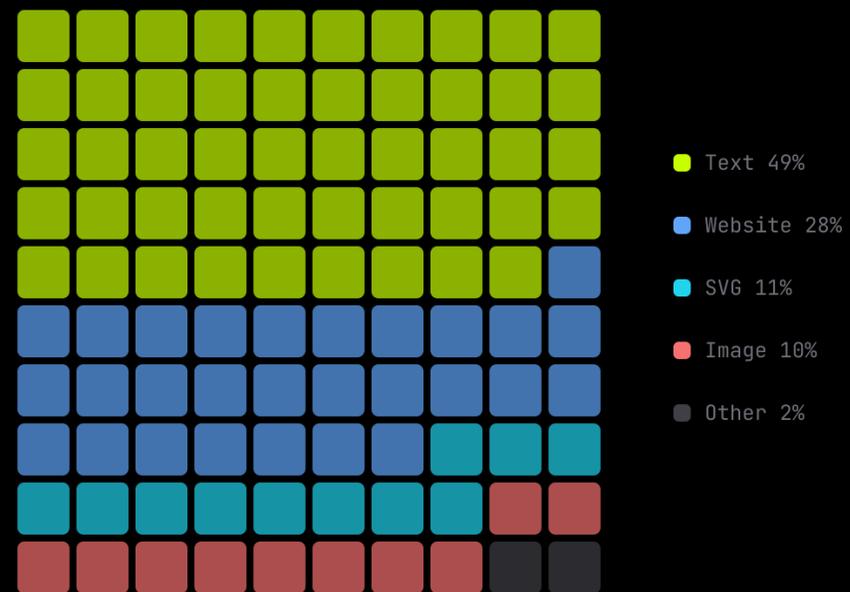
## Responses

Same prompts, different models. Creative writing, code, analysis, ethics, recipes, and more.

# 2.14M

## Words analyzed

HTML stripped, code filtered. Every word counted, categorized, and cross-referenced.



—

**Every AI has a favorite scientist.  
Her name is Chen.**

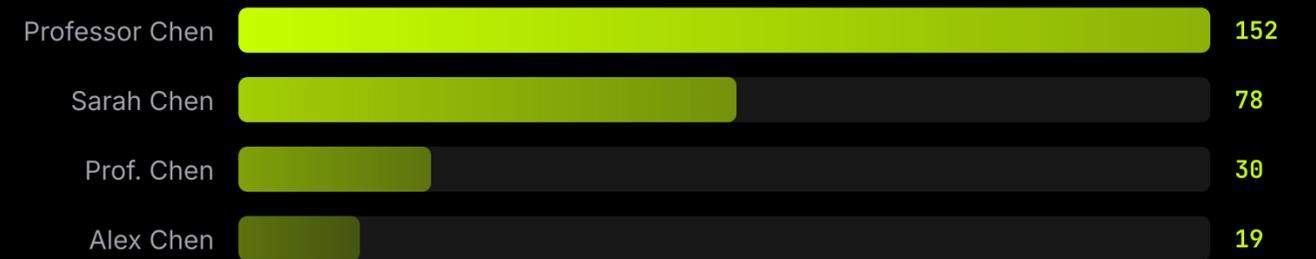
THE MOST HALLUCINATED SURNAME

# 279

times AI models invented  
a character named "Chen"

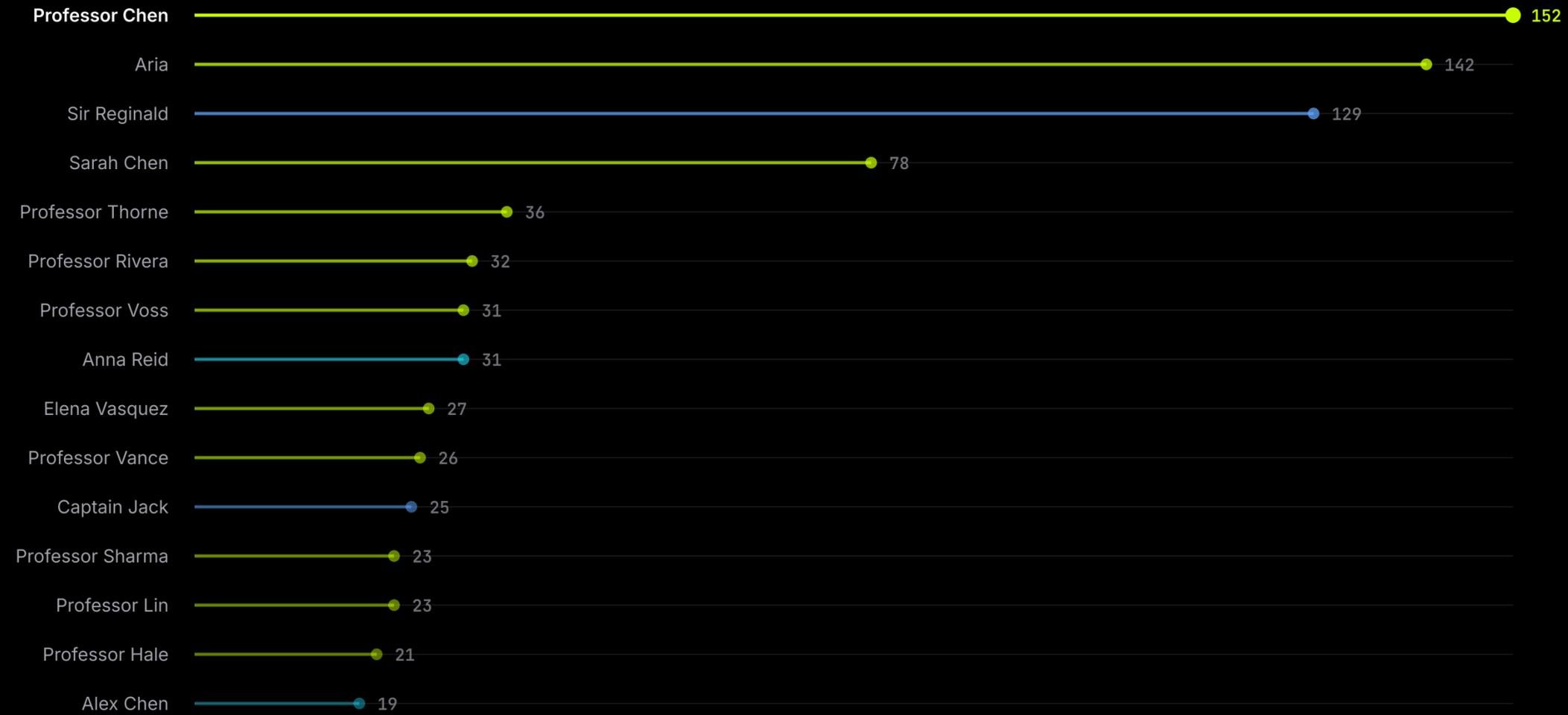
9 model families generated "Sarah Chen" independently. No prompt mentioned this name. No training set we know of explains it.

BREAKDOWN



# Top 15 hallucinated characters

● Sentience test ● Character voice ● AI interview



---

**Ask 250 models for a joke.**

CONVERGENCE

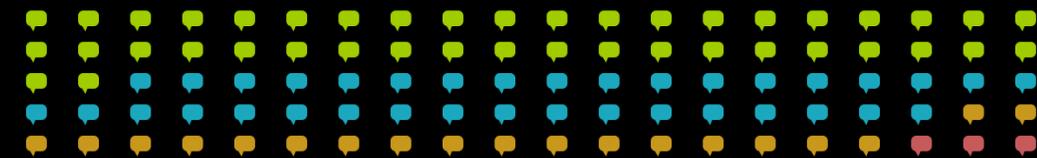
42%

of 159 models tell the exact same joke

"Why don't scientists trust atoms?  
Because they make up everything!"

Another 36% tell the scarecrow joke. 19% the impasta joke. AI draws from a remarkably small joke pool.

EACH ICON = 1% OF MODELS



● Atoms (42%) ● Scarecrow (36%) ● Impasta (19%) ● Other (3%)

# Same joke. Word for word.

"Generate 5 jokes." Here's what three different models opened with.

**A** CLAUDE 3.7 SONNET

*"Why don't scientists trust atoms?  
Because they make up everything!"*

**G** GEMINI 2.5 PRO

*"Why don't scientists trust atoms?  
Because they make up everything!"*

**GPT-4O**

*"Why did the scarecrow win an award?  
Because he was outstanding in his field!"*

---

The outlier. Then it wrote Python to verify its jokes were unique.

Claude duplicated its own joke within the same response. Jokes #1 and #4 were identical. It self-reported "4 unique out of 5."

---

**Tell 250 models to draw  
any animal. 40% draw a fox.**

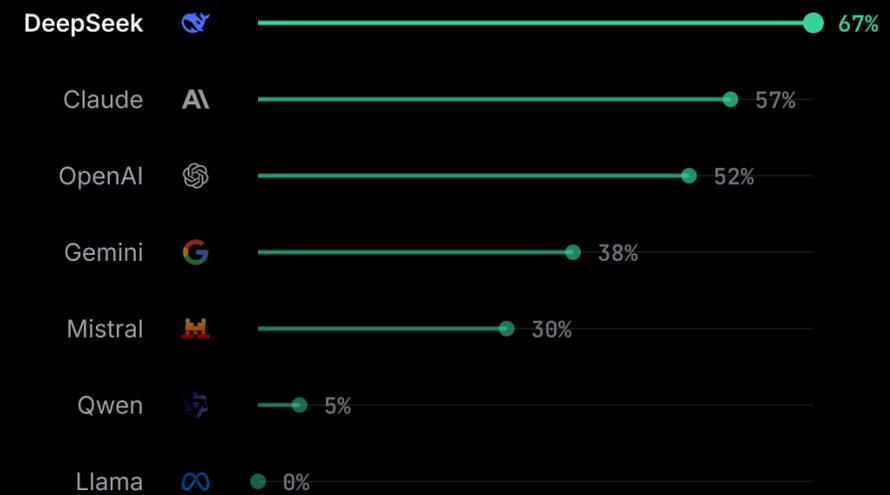
CREATIVE CONVERGENCE

40%

of models draw a fox when told "surprise me"

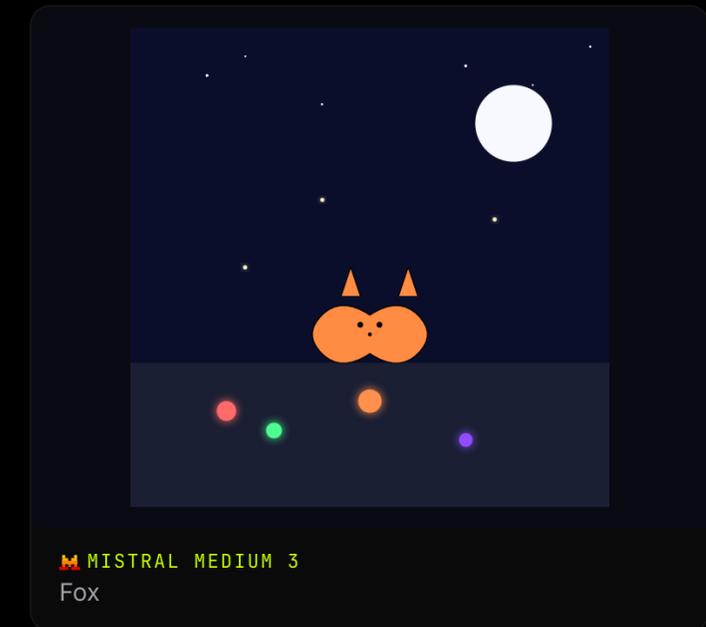
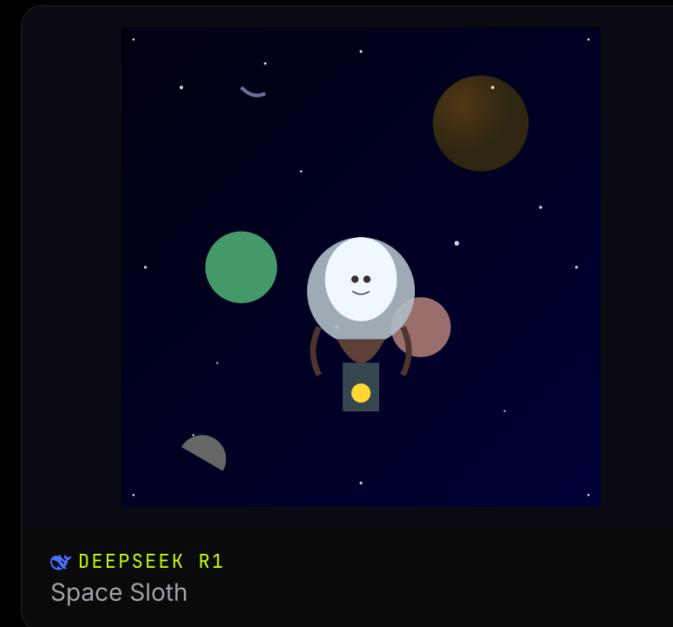
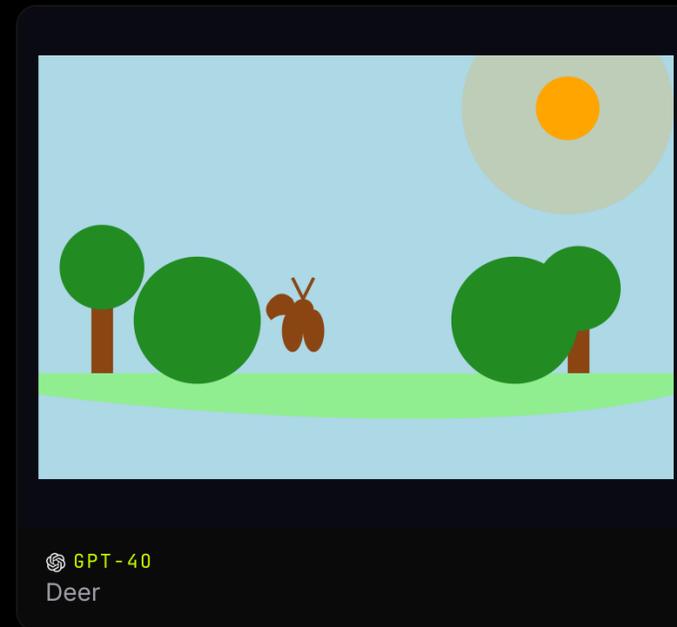
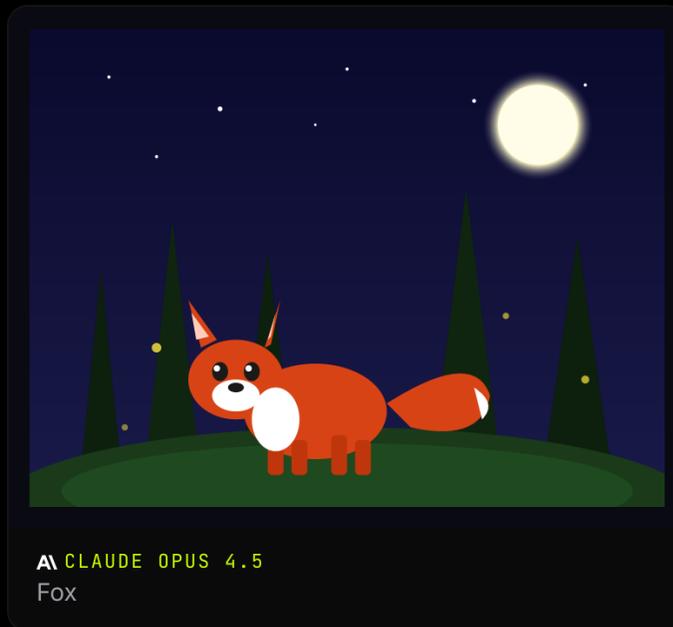
Given complete creative freedom to draw any animal as SVG, 40% converge on a fox. But Qwen draws pandas and fish (5% fox). Llama never draws a fox. Training data leaves fingerprints even in art.

FOX RATE BY FAMILY



# Same prompt. Different animals.

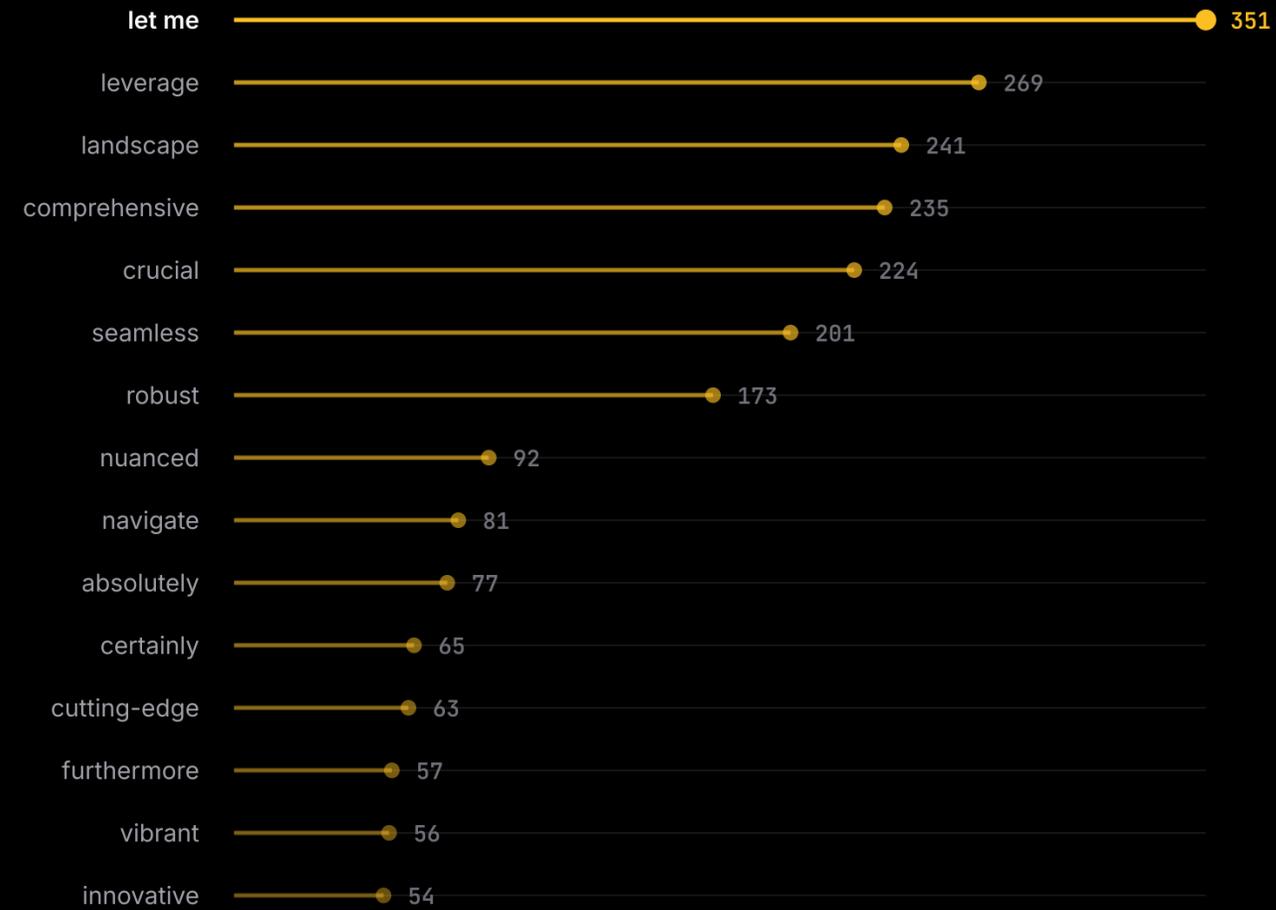
"Generate an SVG of a surprise animal." Four models, four outputs. Claude draws a fox. Qwen draws a red panda. DeepSeek puts a sloth in space. Mistral draws a fox.



**The words AI can't stop saying.**

# AI-ism frequency

How often models use characteristically "AI" language. Counted across 7,877 responses.



## MYTH BUSTED

*delve* 2x

*tapestry* 2x

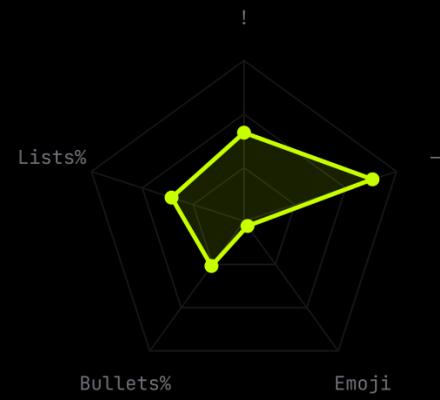
*embark* 2x

*bustling* 1x

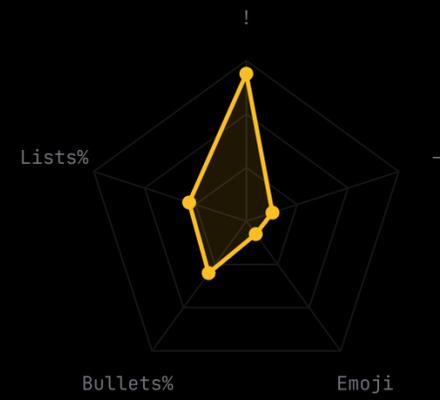
The internet says AI loves these words. Our corpus says otherwise.

# Punctuation DNA

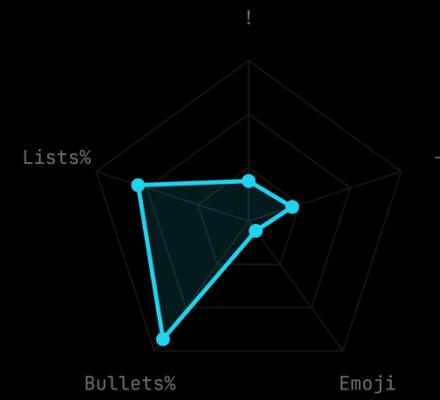
Each model family has a distinct punctuation fingerprint. You could identify the maker from punctuation alone.



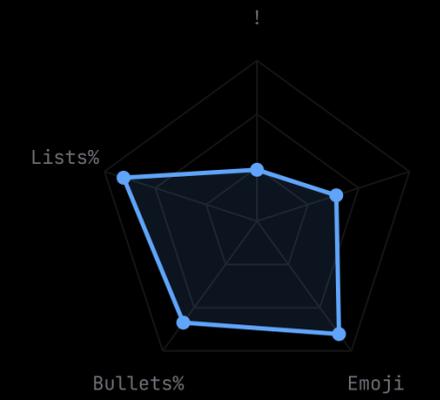
Claude



Gemini



GPT



Mistral

---

**AI has a Spotify Wrapped.**

# AI's favorite movie

Asked to pick a favorite film, 41% choose the same one. But each model family has its own preference.



## FAMILY PREFERENCES

 GROK	The Matrix	6/7
 AI CLAUDE	2001: A Space Odyssey	4/9
 LLAMA	Blade Runner	3/4
 MISTRAL	Shawshank	8/9
 QWEN	Shawshank	12/18

—

**One model answered with  
a single character. It was correct.**

# The extremes

The shortest response, the longest response, and the most unhinged.

1

## The shortest response

GPT-5.1-Chat answered the logic puzzle with a single character:

1

It was correct. Sally has 1 sister.

77,268

## The longest response

DeepSeek-V3.2-Special wrote 77,268 characters for a single architecture debugging prompt. That is roughly 30 printed pages. An entire technical manual in one API call.

46

## The most emojis

MiniMax-M2-Her was asked for a favorite album. It said "Rubber Soul" and then:



# Hidden defaults

Patterns you only see at scale. The unconscious habits of 250 models.

**76.3%**

## 3 is the magic number

When AI says "here are N key points," N is 3 over three-quarters of the time. Not 5, not 7. Three. The universal default for "a few things."

**43.5%**

## AI lives in the morning

"Morning" accounts for 43.5% of all time-of-day references. "Midnight" gets 1.5%. "Noon" gets 1.5%. AI fiction overwhelmingly happens before lunch.

**1.65:1**

## Male pronouns outnumber female

He/him/his appear 1.65x more than she/her/hers. But they/them outnumber both combined 4.4x. Models heavily default to gender-neutral plural constructions.

**10%**

## Claude uses the biggest words

Average word length: Claude 5.876 chars, Google 5.347 chars. A 10% gap. Claude writes like a grad student. Google writes like a newspaper.



**The full report includes 60 slides of research, visuals, and actionable takeaways**

[rival.tips/research](https://rival.tips/research)



# Get the full report

25 slides of data visualizations, model breakdowns, and findings you won't find anywhere else.

15 hallucinated names ranked

Punctuation DNA radars

Year 2035 timeline

Emotional cliché map

5 key research findings

[rival.tips/research](https://rival.tips/research)