



Joe Toplyn



Witscript



Ori Amir

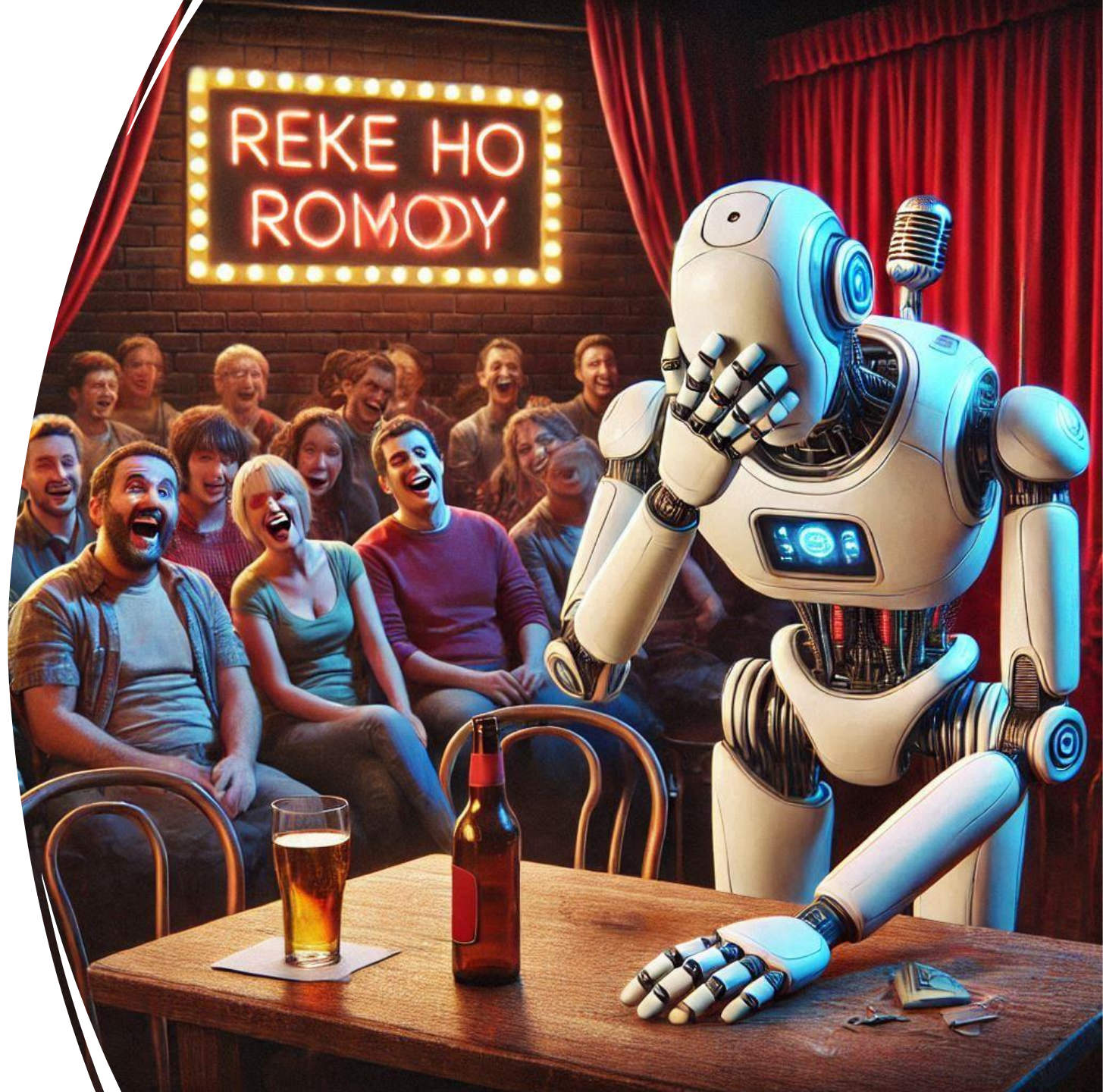
**Can AI Make Us Laugh?
Comparing Jokes Generated
by Witscript and a Human Expert**

**Joe Toplyn & Ori Amir
COLING 2025**

How to use AI to generate contextually integrated jokes

LLMs by themselves can't do it well.

- They only mimic surface patterns.
- Jokes often break expected patterns rather than mimic them.
- LLMs don't have use a good model of joke structure.

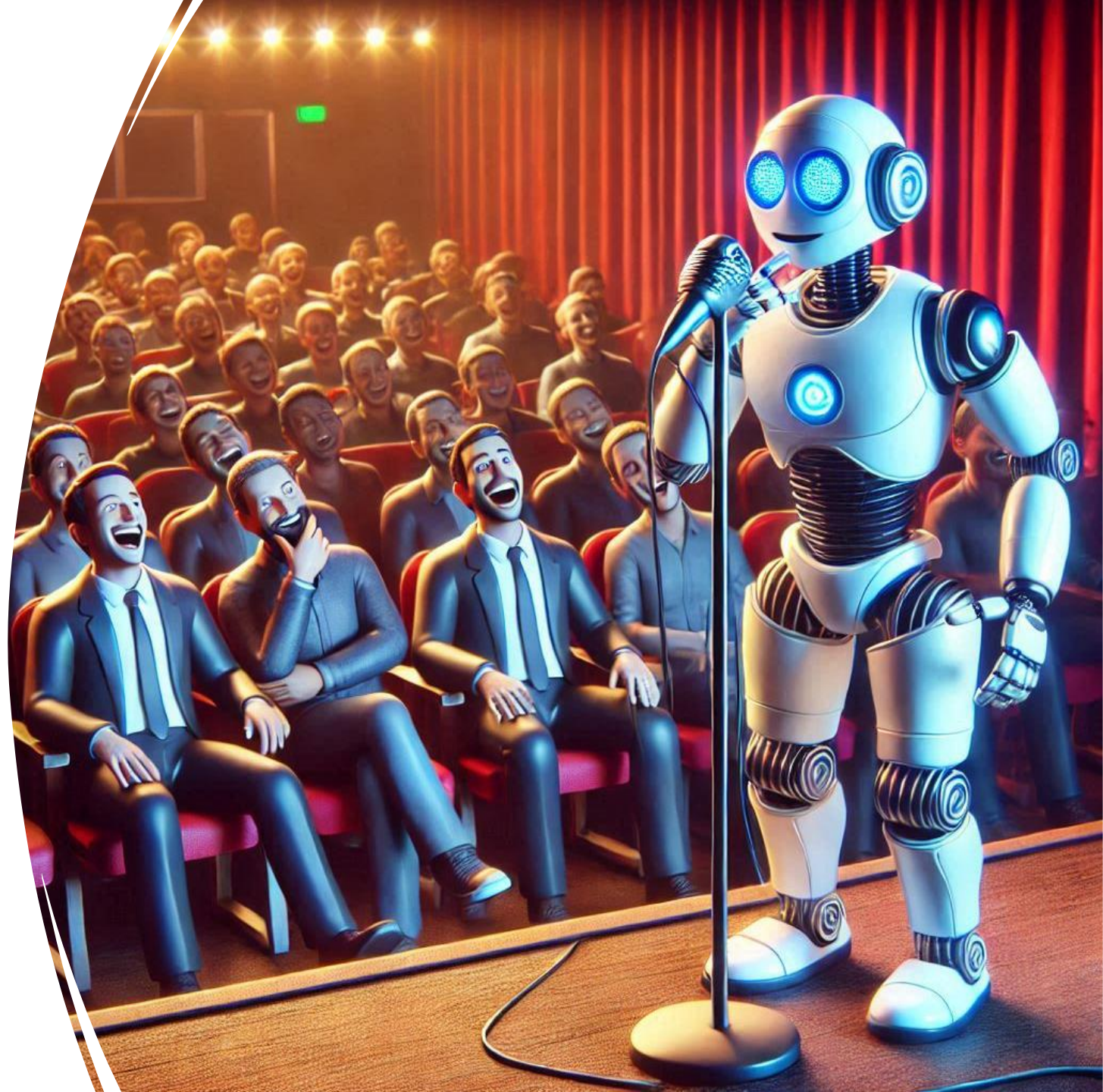


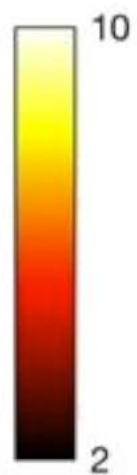
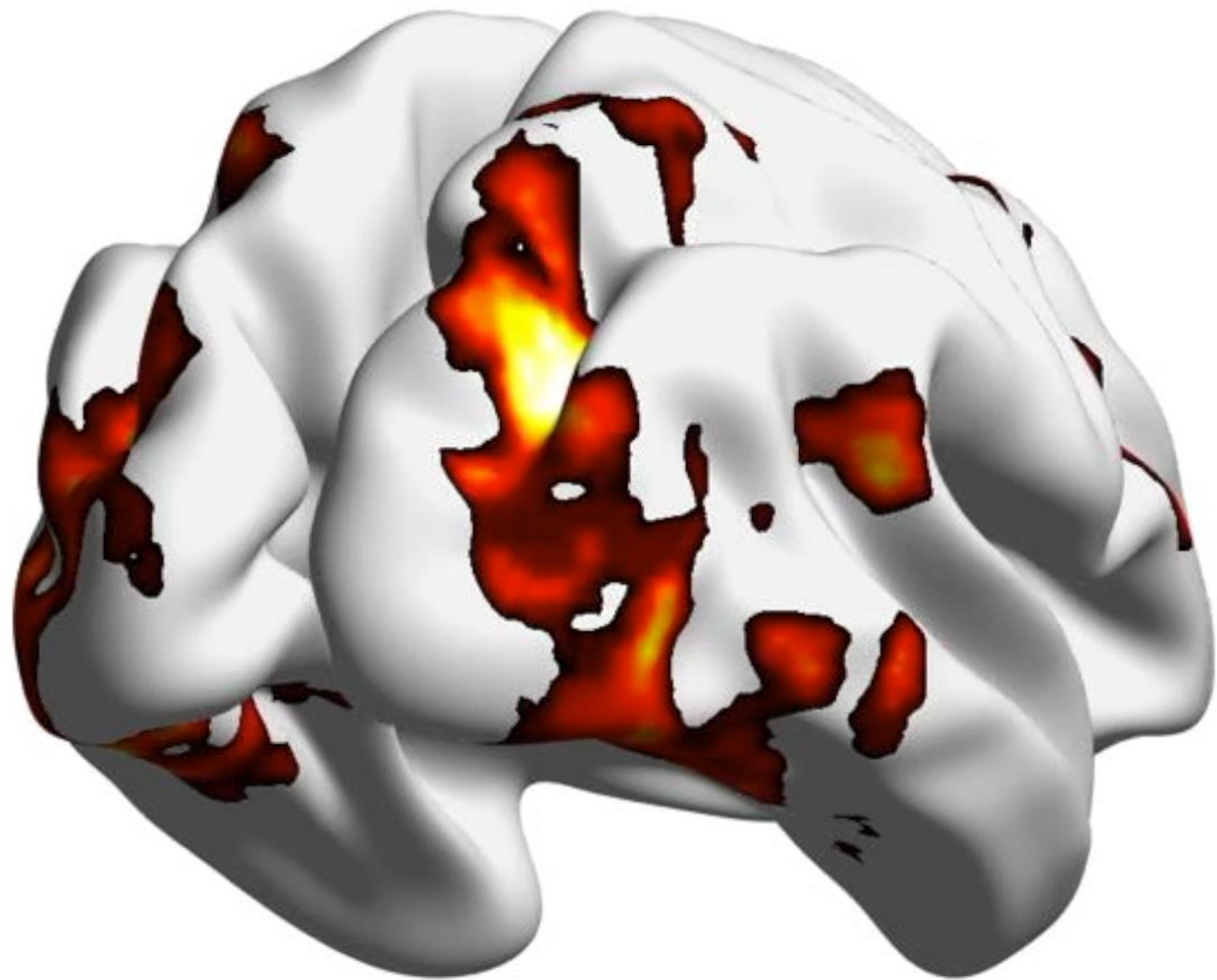
How to use AI to generate contextually integrated jokes

Witscript is a neural-symbolic hybrid AI system.

- **Neural:** an LLM in the GPT family from OpenAI
- **Symbolic:** uses the joke-writing rules of a human expert

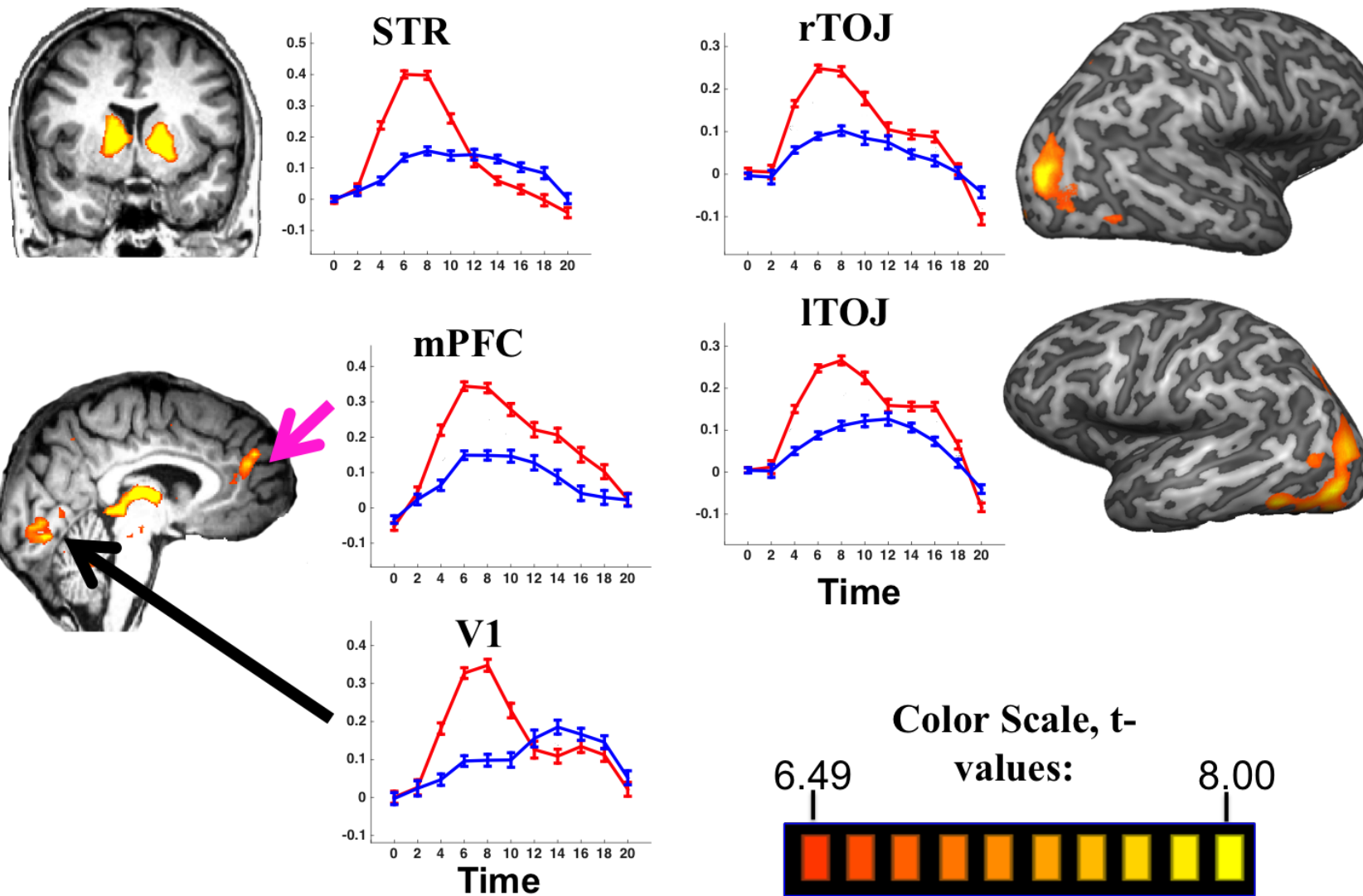
Witscript uses the LLM to execute the rules





HUM > MUN Regions

($p < .01$, corrected):





Write original jokes fast with AI.

- Type in a topic sentence between 5 and 25 words long.
- Click "Get Witscript's responses."
- Witscript responds with some suggested jokes and picks its favorite.
- Click again for more responses to the same sentence. Or enter a new sentence.

What's a good way for Dracula to save money?

Get Witscript's responses

POSSIBLE RESPONSES:

- A) By switching to a coffin with complimentary wifi.
- B) By switching to a vegetarian diet and becoming Edward Cullen's roommate.
- C) By cutting expenses, using coupons and opening a thrifty savings account!
- D) By switching to an all-blood diet instead of buying expensive organ meats.

WITSCRIPT RESPONDS:

By switching to a coffin with complimentary wifi.

The Basic Joke-Writing Algorithm

- a) Get a topic sentence from the user.
 - **What's a good way for Dracula to save money?**
- b) Select two topic handles.
 - **Dracula, save money**
- c) Generate associations of the topic handles.
 - **coffin, complimentary Wi-Fi**
- d) Create a punch line that links associations unexpectedly.
 - **a coffin with complimentary Wi-Fi**
- e) Generate text to connect the topic and punch line.
 - **By switching to a coffin with complimentary Wi-Fi**

The Experiment

- A professional writer for late-night comedy shows and Witscript each wrote jokes for the same set of news topics.
- The jokes were then split into two setlists to be performed in different live comedy shows.
- Each setlist had half of its jokes written by AI and half by the human. The two setlists were counterbalanced.

	Setlist 1	Setlist 2
Topic 1	AI joke	Human joke
Topic 2	Human joke	AI joke
Topic 3	Human joke	AI joke
Topic 4	AI joke	Human joke
...

AI Joke

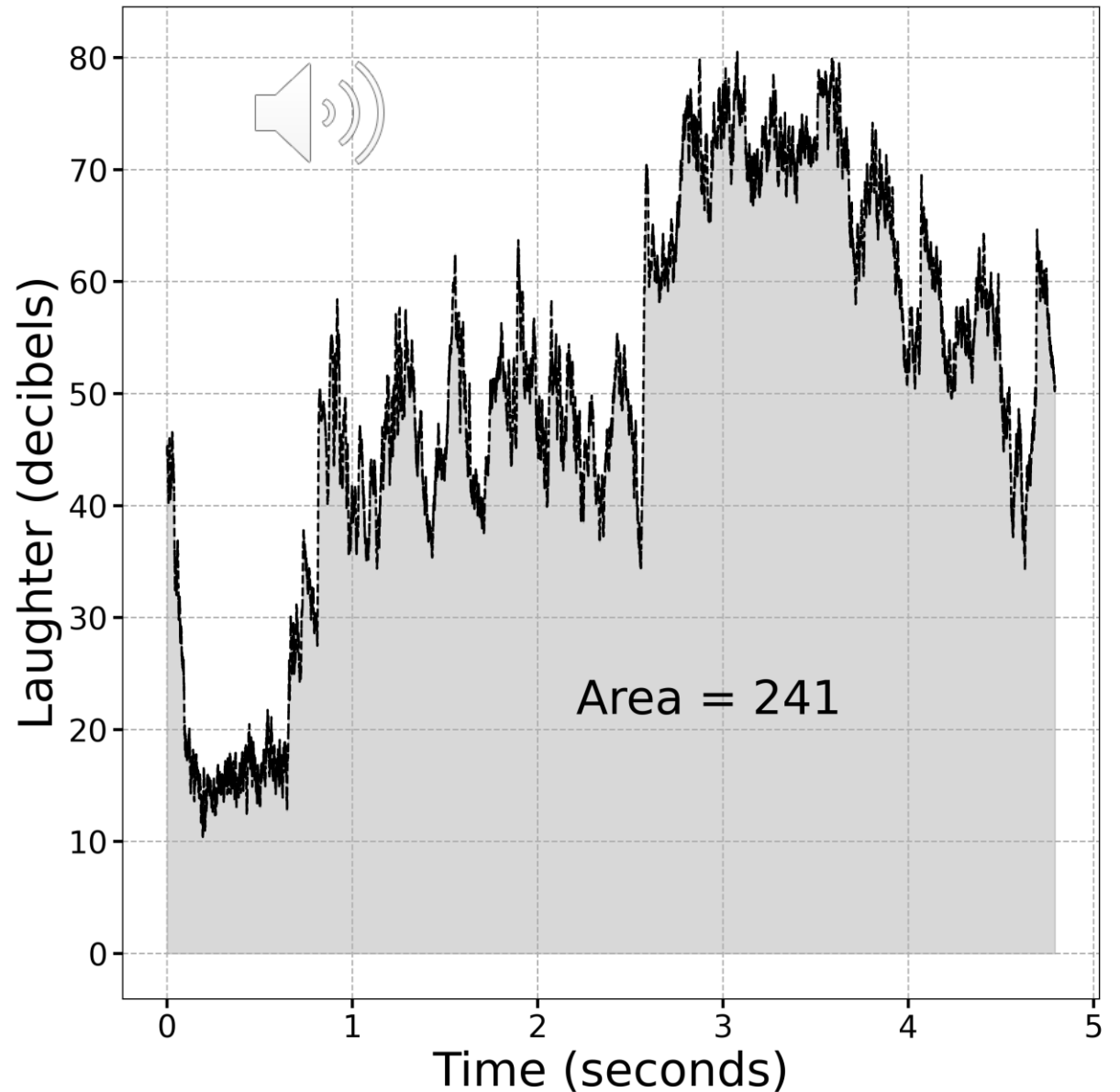


Human Joke



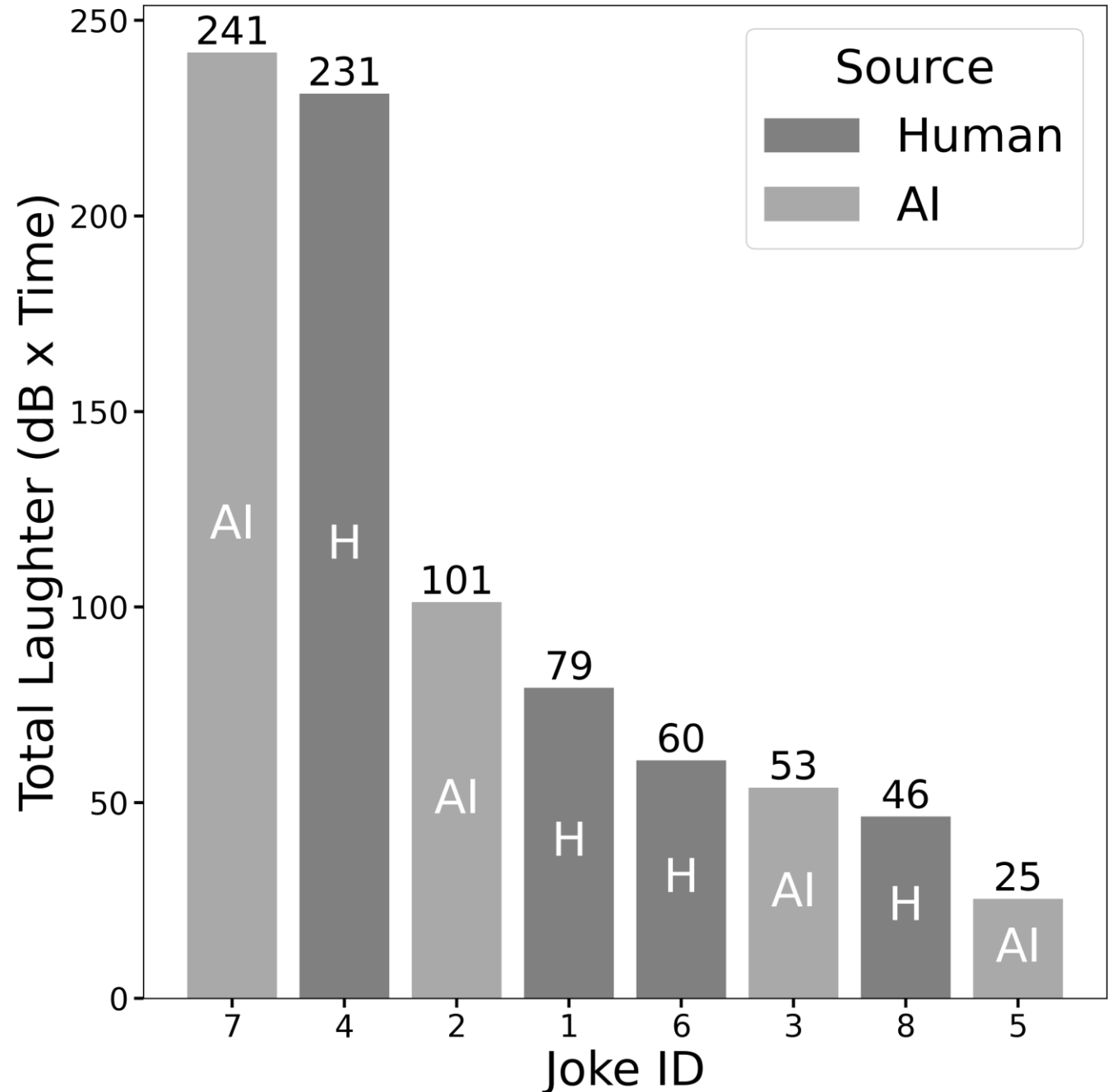
Measuring Laughter of a Joke

- The laugh track was converted to **decibels** to mimic human sound perception.
- The area under the curve of Laughter over Time was computed, giving a measurement of **Total Laughter**.



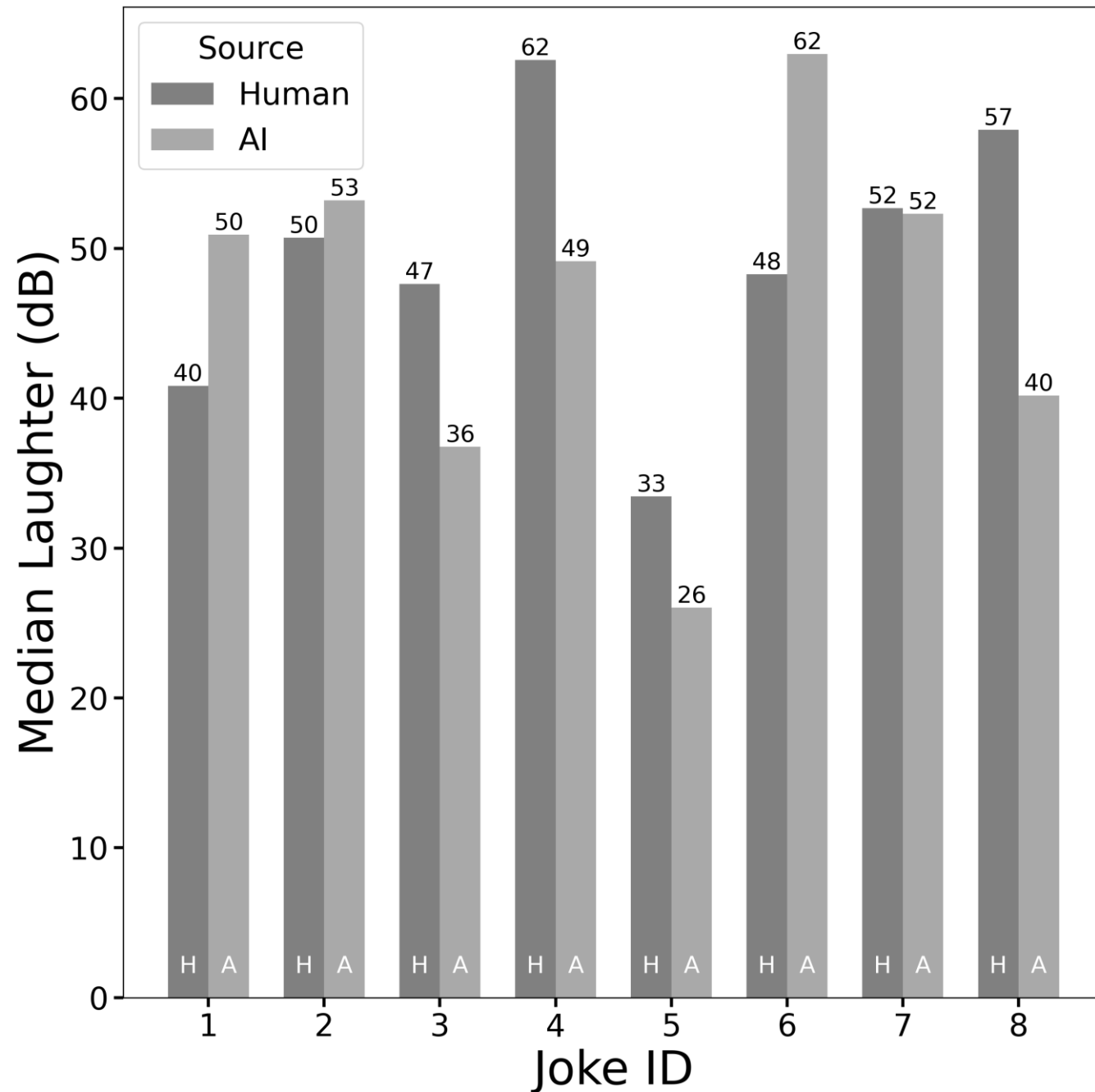
Comparing AI to Human: Single Show

- In a live comedy show with an audience of $N = \sim 35$, a comedian performed human and AI jokes in pseudo-random order.
- Jokes were ranked for funniness based on their Total Laughter (in decibel-seconds).
- The audience's reaction to the AI jokes was indistinguishable from the professional human writer's jokes.



Comparing AI to Human: Two Shows

- Two shows with counterbalanced jokes by AI and Human were performed.
- **Total Laughter** had to be **normalized** to account for audience size and mood differences.
- **Median Laughter** in decibels (see graph) was identical for both shows and did not require normalization.
- **No significant differences** were detected between the AI and Human jokes **for either measure**.



Conclusion

This study provides naturalistic, real-world evidence that when it comes to generating comedy/talk show monologue-style humor, **an AI system can perform at the level of a professional human comedy writer.**

BUT... how valid are the results?



Standard of Proof

Historically, the standard for demonstrating that AI had reached a certain milestone against human performance involved only a few data points.



In 2011, Watson (AI) competed **only once** against two human champions on *Jeopardy!*, and won.



Kasparov played only **six games** with Deep Blue (AI) in 1997, scoring **2.5-3.5**.

Strengths and Limitations

Naturalistic/Real-World Study

Strengths

- The dependent variable is **genuine laughter**.
- Higher **external validity** compared to most studies that employ ratings on a numerical scale.

Limitations

- Laughter can be affected by **factors other than the joke** (e.g., gestures, environment). But these factors should average out.
- Rating on a numerical scale can capture **aspects of amusement** that are not expressed in laughter, and laughter is affected by factors other than amusement.

Live Audience

Strength

Captures feedback from a large number of participants (N = ~55 across shows)

Limitations

- Audience members are not independent; laughter is contagious.
- Some audience members laugh louder than others.

What does this all mean for comedy writing?

Witscript [Write](#) [FAQ](#) [Account](#)

Write original jokes fast with AI.

- Type in a topic sentence between 5 and 25 words long.
- Click "Get Witscript's responses."
- Witscript responds with some suggested jokes and picks its favorite.
- Click again for more responses to the same sentence. Or enter a new sentence.

In the future, robots might be doing standup comedy.

[Get Witscript's responses](#)

POSSIBLE RESPONSES:

- A) But they'll have to update their software for better timing and delivery.
- B) But let's be real, they'll never have the charm of Will Smith in 'I, Robot'.
- C) They'll be programmed to make you laugh, not take over the world.
- D) I can't wait to see a robot try to heckle an audience member and accidentally fry them with a laser beam.

WITSRIPT RESPONDS:

I can't wait to see a robot try to heckle an audience member and accidentally fry them with a laser beam.

www.witscript.com