



Evaluating Human Perception and Bias in AI-Generated Humor

Narendra Nath Joshi



Motivation

- ❖ Recent AI advances enable human-like creative content generation, including humor
- ❖ Humor is fundamental to human interaction:
 - Facilitates social bonding and reduces stress
 - Requires complex cognitive processes (pattern recognition, perspectives etc)
 - Heavily influenced by cultural context and personal experience
- ❖ Key Research Objectives
 - Can AI capture the nuance and authenticity of human humor?
 - How do human biases affect evaluation of AI-generated humor?
- ❖ **Study Goal: Investigate human perception of AI-generated humor**

Hypotheses and Research Questions

H1: Reasonable Identification Hypotheses

Participants' accuracy in identifying AI-generated jokes is higher than chance.

H2: Reasonable Doubt Hypotheses

Humans rate AI-generated jokes lower in quality compared to human-generated jokes.

The perceived quality of AI-generated jokes improves when participants are unaware of the source.

RQ1: Are participants able to accurately identify AI-generated jokes more often than by chance?

RQ2: Do humans rate AI-generated jokes lower in quality compared to human-generated jokes?

RQ3: Does the perceived quality of AI-generated jokes improve when participants are unaware of the source?

Hypotheses and Research Questions

H3: Repeated Exposure Hypotheses

Participants' ratings of AI-generated jokes improve after repeated exposure.

RQ4: Do participants' ratings of AI-generated jokes improve after repeated exposure?

H4: Demographic Hypotheses

Younger participants rate AI-generated jokes higher than older participants.

Participants with a background in technology of AI are more accepting of AI-generated humor.

RQ5: Do younger participants rate AI-generated jokes higher than older participants?

RQ6: Are participants with a background in technology of AI more accepting of AI-generated humor?

Study Design: Participant Groups

Group A (Human Baseline)

Participants evaluated only human-generated jokes, establishing a baseline for humor quality ratings.

Group B (AI Baseline)

Participants evaluated only AI-generated jokes, allowing assessment of perceived quality for AI-generated content.

Group C (Alternating Sequence)

Participants evaluated an alternating sequence of human and AI-generated jokes, enabling assessment of distinction abilities in a structured mixed context.

Group D (Mixed Presentation)

Participants evaluated a randomized set of both human and AI-generated jokes, testing identification accuracy in a naturalistic mixed context.

Group E (Blind AI Test)

Participants evaluated AI-generated jokes without knowledge of their source, measuring unbiased quality perception.

Group F (Informed AI Test)

Participants evaluated AI-generated jokes with explicit knowledge of their AI origin, enabling direct comparison with Group E to measure source-related bias.

Study Design: Rating Sessions

- ❖ Demographic Data Collected
 - Gender
 - Age Range
 - Experience with AI Technologies
- ❖ Initial Ratings Phase
 - All participants rated jokes on a 5-point scale
- ❖ Source Assessment Phase (Group E only)
 - Source Identification
 - Confidence Ratings (5-point scale)

Results

RQ1: Source Assessment Performance

Participants' ability to identify AI-generated jokes shows minimal deviation from chance

- ❖ Participant Accuracy = 0.43034
- ❖ Mean Confidence Score = 3.892

RQ2: Comparative Quality Ratings

Participants do not rate AI-generated jokes significantly lower in quality compared to human-generated jokes

- ❖ Mean Rating for AI Jokes = 2.97393
- ❖ Mean Rating for Human Jokes = 2.94769

Results

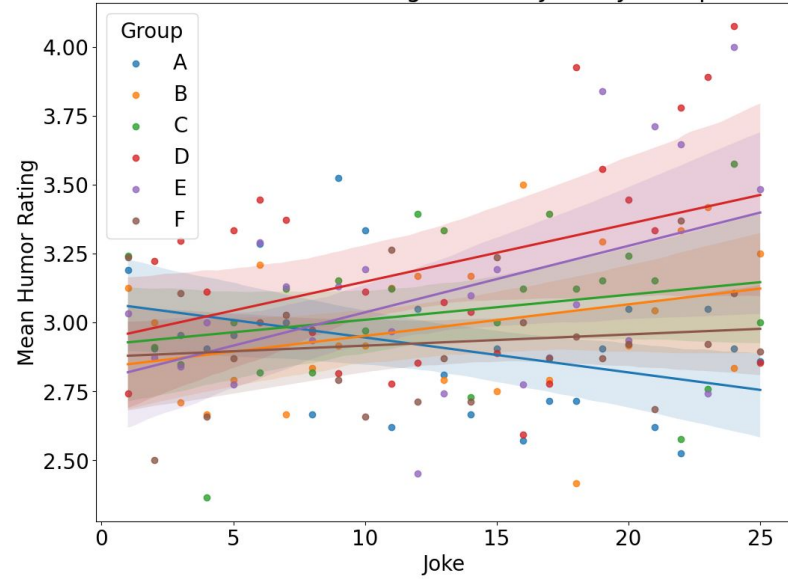
RQ3: Source Awareness Effects

Participants show a clear bias against AI jokes when informed about the source ($p < 0.05$)

- ❖ Mean Rating for AI Jokes (Blind) = 3.34064
- ❖ Mean Rating for AI Jokes (Informed) = 2.92737

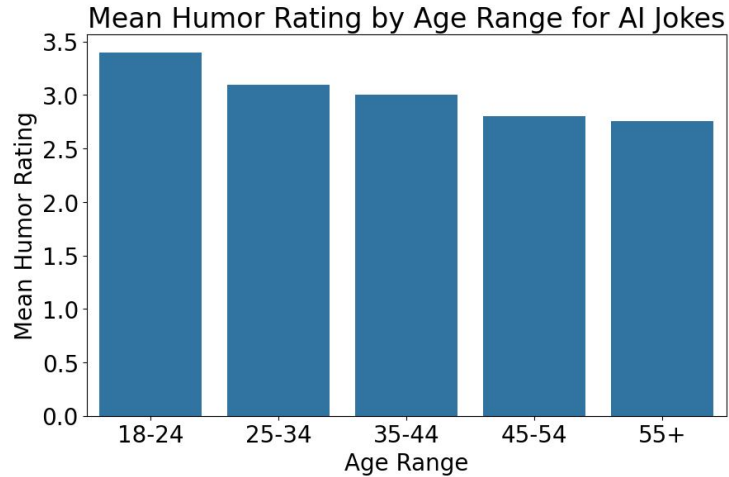
RQ4: Exposure and Learning Effects

Mean Humor Rating for Each Joke by Group

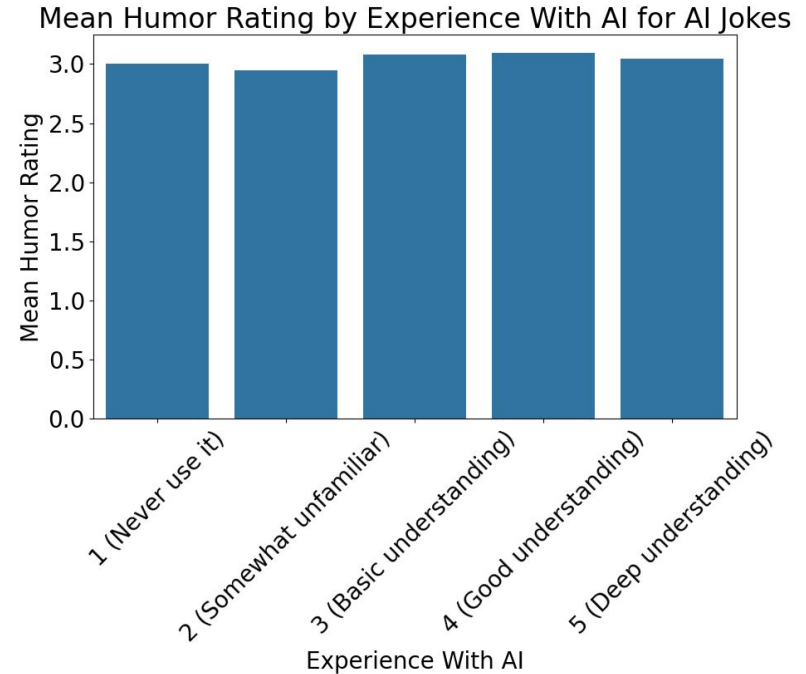


Results

RQ5: Age-Related Effects



RQ6: Technical Background Impact



Summary

- ❖ Participants are no better than random chance in identifying AI-generated jokes
- ❖ Participants do not rate AI-generated jokes significantly lower in quality compared to human-generated jokes
- ❖ Participants show a clear bias against AI jokes when informed about the source
- ❖ There is a familiarization effect among participants, where initial skepticism gives way to increased appreciation of AI-generated humor
- ❖ Younger participants demonstrate more openness to AI-generated humor
- ❖ Participants with technical backgrounds did not provide more favorable ratings to AI jokes

Future Research

- ❖ Study long-term changes in perception over extended exposure to AI-generated content, which could reveal how familiarity affects audience reception
- ❖ Study cross-cultural variations in AI-generated humor
- ❖ Study the role of personalization in AI-generated humor
- ❖ Study the influence of different humor styles and contexts on AI-generated content

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

Human or AI?

Thank you!